

UFRRJ

INSTITUTO DE FLORESTAS

**PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS
AMBIENTAIS E FLORESTAIS**

TESE

**Mapeamento digital de solos e predição de atributos
utilizando *Machine Learning* e lógica *fuzzy* na bacia do
Ribeirão Inhaúma, Iconha/ES**

Patrícia Morais da Matta Campbell

2017



**UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
INSTITUTO DE FLORESTAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS AMBIENTAIS E
FLORESTAIS**

**MAPEAMENTO DIGITAL DE SOLOS E PREDIÇÃO DE
ATRIBUTOS UTILIZANDO *MACHINE LEARNING* E LÓGICA
FUZZY NA BACIA DO RIBEIRÃO INHAÚMA, ICONHA/ES**

PATRÍCIA MORAIS DA MATTA CAMPBELL

Sob a orientação do Professor
Márcio Rocha Francelino

Tese submetida como requisito parcial
para obtenção do grau de **Doutor em
Ciências**, no Curso de Pós-Graduação
em Ciências Ambientais e Florestais,
Área de concentração Silvicultura.

Seropédica, RJ
Dezembro de 2017

Campbell, Patrícia Morais da Matta, 1988-
C187m Mapeamento digital de solos e predição de atributos
utilizando Machine Learning e lógica fuzzy na bacia do
Ribeirão Inhaúma, Iconha/ES / Patrícia Morais da Matta
Campbell. - 2017.
245 f.

Orientador: Márcio Rocha Francelino.
Tese (Doutorado). -- Universidade Federal Rural do Rio
de Janeiro, Programa de Pós-graduação em Ciências
Ambientais e Florestais, 2017.

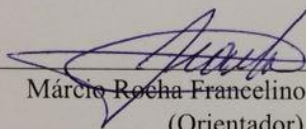
1. Lógica fuzzy-Mapeamento digital de solos. 2.
Aprendizado de máquinas. 3. Técnicas espectrais. 4.
Quimiometria-Fluorescência de raios-x. 5. Dados
aerogeofísicos-Magnetometria-Gamespectrometria. I.
Francelino, Márcio Rocha, 1966-, orient. II
Universidade Federal Rural do Rio de Janeiro.
Programa de Pós-graduação em Ciências Ambientais e
Florestais III. Título.

UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
INSTITUTO DE FLORESTAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS AMBIENTAIS E FLORESTAIS

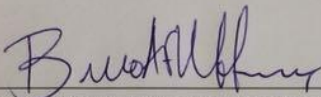
PATRÍCIA MORAIS DA MATTA CAMPBELL

Tese submetida como requisito parcial para obtenção do grau de **Doutor em Ciências**, no Programa de Pós-Graduação em Ciências Ambientais e Florestais, área de Concentração em Silvicultura.

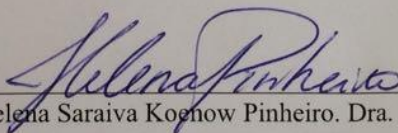
TESE APROVADA EM 07/12/2017



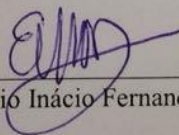
Márcio Rocha Francelino. Dr. UFRRJ
(Orientador)



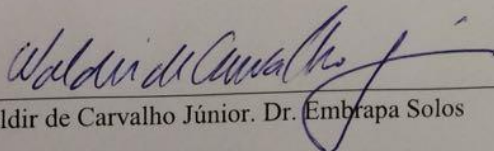
Bruno Araújo Furtado de Mendonça. Dr. UFRRJ



Helena Saraiva Kognow Pinheiro. Dra. UFRRJ



Elpídio Inácio Fernandes Filho. Dr. UFV



Waldir de Carvalho Júnior. Dr. Embrapa Solos

Aos meus pais, Marluce e Neto e ao meu marido Bruno, pelo apoio e amor incondicional em todos os momentos. Sem vocês nenhuma conquista valeria à pena.

“Foi o tempo que dedicastes à tua rosa que a fez tão importante” (Antoine de Saint-Exupéry)

Dedico

AGRADECIMENTOS

A Deus, pela dádiva da vida.

A UFRRJ, por esses anos de aprendizado e desenvolvimento profissional.

Aos Drs. José Barbosa dos Santos e Luzinério Prezotti, que foram responsáveis por meus primeiros passos na pesquisa.

Aos Drs. Claudia Pozzi Jantalia, Bruno José Rodrigues Alves, Robert Michael Boddey, Segundo Urquiaga, Sebastião Manhães Souto, Alexander Rezende e Paulo Francisco Dias†***, pelo convívio de anos me proporcionando grande aprendizado, que me permitiram continuar nessa jornada.

Ao Dr. Victor Hugo Alvarez Venegas, pelo apoio e incentivo de sempre!

Ao meu orientador Dr. Márcio Rocha Francelino, para quem seria impossível expressar todo meu agradecimento. Obrigada pela dedicação, pelos conhecimentos a mim ensinados, pela paciência, amizade, pela solicitude todas as vezes que precisei e por ser esse exemplo de profissional e ser humano.

Aos professores Dr. Elpídio Inácio Fernandes Filho, José Alexandre Melo Demattê e Marcos Gervásio Pereira, por terem contribuído enormemente para o desenvolvimento deste trabalho.

A Clécia, da ESALQ, por todo suporte e empenho.

Aos amigos do INCAPER Alcélcio, Hebert, Alciro, João, Ana, Fábio, Erivelton, Samuel e Renato pelo apoio durante todos esses anos.

Aos amigos da UFV, em especial ao Valdemir, Martin, Pedro, Adriano e Pablo, pela valiosa contribuição nesse trabalho.

Aos meus amigos de sempre Carlos Magno e Felipe, por todo apoio desde o início, vocês foram fundamentais nesse estudo!

Ao Bruno, sem seu incentivo, companheirismo, força e ajuda no desenvolvimento desse trabalho não teria sido possível esse desfecho, muito obrigada por estar sempre ao meu lado, buscando meu crescimento pessoal e profissional.

A outros, que por descuido não tenham sido citados mas que contribuíram com a realização desse sonho!

RESUMO

CAMPBELL, Patrícia Morais da Matta. Mapeamento digital de solos e predição de atributos utilizando Machine Learning e Lógica fuzzy na bacia do Ribeirão Inhaúma, Iconha/ES. 2017. 282p. Tese (Doutorado em Ciências Ambientais e Florestais). Instituto de Florestas, Área de Concentração em Silvicultura, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, 2017.

Obter informações a respeito do solo e seus atributos é essencial para o bom planejamento do uso e ocupação do mesmo. O mapeamento convencional e as análises laboratoriais de rotina são os métodos mais utilizados no Brasil para espacialização do solo e obtenção de informações acerca de seus atributos. No caso do mapeamento de classes de solo esse método é baseado em modelos mentais desenvolvidos pelo pedólogo, o que o torna altamente subjetivo, dependente da experiência do profissional e de difícil repetição, enquanto as análises laboratoriais são dispendiosas e apresentam a possibilidade de geração de impactos ambientais pelo uso de reagentes químicos. Por isso, novas técnicas computacionais tem se apresentado como uma ferramenta para mapeamento de classes e atributos de solos, proporcionando maior rapidez e diminuição do custo em relação aos métodos convencionais. Assim nesse trabalho estudou-se o uso dessas técnicas no Ribeirão Inhaúma, em Iconha/ES. Os capítulos 1 e 2 mostram o uso de diferentes classificadores para mapeamento digital de solos, ao total foram testados 10 modelos, que mostraram-se eficientes, destacando-se a lógica *fuzzy*, *Randon Forest*, *Ranger*, *Extreme Gradient Boosting* e *Weighted Subspace Random Forest*, o capítulo 1 ainda comparou esse modelo ao mapeamento convencional, encontrando 24% de concordância entre os mapas. Os capítulos 3 e 4 testaram o uso de diversos modelos para mapeamento de atributos, destacando-se no capítulo 3 o método *Randon Forest* e no 4 o uso de técnicas espectrais para mapeamento de carbono orgânico, fósforo e argila com boa acurácia. O capítulo 5 analisou o grau de correlação entre informações levantadas pelo sensoriamento remoto através de dados aerogeofísicos e dados obtidos pela análise de fluorescência feita com analisador de raios-X, comparando os valores encontrados com o mapa de uso e ocupação do solo. Foi possível identificar a correlação existente entre o uso e ocupação dos solos com os dados de gamaespectrometria e de temperatura. Em relação aos dados aerogeofísicos de gamaespectrometria foram encontradas altas correlações com altitude e baixa correlação com dados obtidos pela análise de termofluorescência, indicando que dados provenientes da análise de termofluorescência não podem substituir àqueles advindos da gamaespectrometria e vice-versa, para fins de análise de características do terreno. O capítulo 6 propõe um estudo inicial sobre a possibilidade do uso da fluorescência como ferramenta na determinação de óxidos de Fe, Al, Si e Ti, a fim de diminuir o impacto no meio ambiente causado pelos reagentes químicos utilizados no ataque sulfúrico e aumentar a velocidade de análise e preparo da amostra. As análises de fluorescência de raios-X mostraram-se promissoras para determinação dos teores de óxidos de Cambissolos e Latossolos, sendo necessárias análises de uma quantidade maior de amostras utilizando os dois métodos, afim de comprovar a correlação existente entre os teores encontrados.

Palavras-chave: mapeamento digital de solos, técnicas espectrais, fluorescência de raios-X.

ABSTRACT

CAMPBELL, Patrícia Morais da Matta. Mapeamento digital de solos e predição de atributos utilizando Machine Learning e Lógica fuzzy na bacia do Ribeirão Inhaúma, Iconha/ES. 2017. 282p. Tese (Doutorado em Ciências Ambientais e Florestais). Instituto de Florestas, Área de Concentração em Silvicultura, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, 2017.

Obtaining information about the soil and its attributes is essential for proper planning of the use and occupation of it. Conventional mapping and routine laboratory analysis are the most used methods in Brazil for soil spatialization and obtaining information about its attributes. In the case of mapping of soil classes this method is based on mental models developed by the pedologist, which makes it highly subjective, dependent on the professional experience and difficult to replicate, while the laboratory analyzes are expensive and present the possibility of generating impacts through the use of chemical reagents. Therefore, new computational techniques have been presented as a tool for mapping classes and attributes of soils, providing greater speed and decrease of cost in relation to conventional methods. Thus in this work the use of these techniques was studied in Ribeirão Inhaúma, in Iconha / ES. In this study, we used 10 different models for the digital mapping of soils. In this case, 10 models were tested, which were efficient, with emphasis on fuzzy logic, Randon Forest, Ranger, Extreme Gradient Boosting and Weighted Subspace Random Forest , chapter 1 also compared this model to conventional mapping, finding 24% agreement between the maps. Chapters 3 and 4 tested the use of several models for mapping attributes, highlighting in chapter 3 the Randon Forest method and in 4 the use of spectral techniques for mapping organic carbon, phosphorus and clay with good accuracy. Chapter 5 analyzed the degree of correlation between information collected by remote sensing through aerogeophysical data and data obtained by fluorescence analysis using an X-ray analyzer, comparing the values found with the land use and occupation map. It was possible to identify the correlation between the use and occupation of soils with gamma spectrometry and temperature data. High correlation with elevation and low correlation with thermofluorescence data was observed in the aerogeophysical data of gamma spectrometry, indicating that data from the thermofluorescence analysis can not replace those from the gamma spectrometry and vice versa, for the purpose of analyzing characteristics of the land. Chapter 6 proposes an initial study on the possibility of using fluorescence as a tool in the determination of Fe, Al, Si and Ti oxides in order to reduce the impact on the environment caused by the chemical reagents used in the sulfuric attack and to increase the speed analysis and preparation of the sample. X-ray fluorescence analyzes were promising for the determination of the oxides of Cambisols and Oxisols, and a larger number of samples were required using the two methods, in order to prove the correlation between the contents found.

Keywords: digital mapping of soils, spectral techniques, X-ray fluorescence.

LISTA DE TABELAS

Tabela 1- Distribuição dos perfis, amostras extras e pontos de observação por classes de solos na área de estudo.....	20
Tabela 2- Intervalos de Kappa com respectiva interpretação de desempenho	22
Tabela 3- - Estatística descritiva das covariáveis utilizadas para predição das classes de solo	23
Tabela 4- Valores e tipos de curva utilizados para os atributos do terreno para determinação de cada classe de solo	25
Tabela 5- Matrix de confusão para o mapeamento de solos utilizando lógica <i>fuzzy</i>	28
Tabela 6- Matriz de confusão para o mapeamento convencional de solos.....	31
Tabela 7- Áreas definidas para cada classe pelo mapeamento convencional e digital usando lógica <i>fuzzy</i>	31
Tabela 8- Distribuição dos perfis, amostras extras e pontos de observação por classes de solos na área de estudo.....	56
Tabela 9- Descrição das variáveis morfométricas utilizadas no mapeamento digital de solos da bacia do ribeirão Inhaúma, Iconha/ES	58
Tabela 10- Relação dos dados gamaespectrométricos e magnométricos, utilizados como parte das variáveis independentes para predição de classes de solos	60
Tabela 11- Classificadores utilizados para mapeamento de solos na bacia do Ribeirão Inhaúma, Iconha/ES, com seus respectivos métodos e bibliotecas do R	61
Tabela 12- Intervalos de Kappa com respectiva interpretação de desempenho	61
Tabela 13- Desempenho dos classificadores segundo o índice Kappa.....	64
Tabela 14- Matriz de confusão gerada pela classificação segundo o método Adabag.....	66
Tabela 15- Matriz de confusão gerada pela classificação segundo o método MARS.....	66
Tabela 16- Matriz de confusão gerada pela classificação segundo o método CIRF	66
Tabela 17- Matriz de confusão gerada pela classificação segundo o método C5.0	67
Tabela 18- Matriz de confusão gerada pela classificação segundo o método SVM Linear	67
Tabela 19- Matriz de confusão gerada pela classificação segundo o método SVM Polinomial	67
Tabela 20- Matriz de confusão gerada pela classificação segundo o método WSRF	68
Tabela 21- Matriz de confusão gerada pela classificação segundo o método Xgbtree	68
Tabela 22- Matriz de confusão gerada pela classificação segundo o método Random Forest	68
Tabela 23- Matriz de confusão gerada pela classificação segundo o método Ranger.....	69
Tabela 24- Áreas definidas para cada classe pelos classificadores Random forest, Ranger, XGBTREE, WSRF.....	69
Tabela 25- Ordem de importância das variáveis utilizadas pelos classificadores <i>Randon Forest, Extreme Gradient Boosting e Ranger</i> na predição das classes de solo no Ribeirão Inhaúma, Iconha/ES.....	70
Tabela 26- Teste Z para os coeficientes Kappa de 10 métodos de classificação usados no mapeamento digital de solos do Ribeirão Inhaúma, Iconha/ES	71
Tabela 27- Variáveis independentes utilizadas para predição de atributos químicos do solo.	90
Tabela 28- Teste de normalidade para os elementos Al, Ca, Fe, K, Mg, Na, Si, Ti, Cr, Cu, Mn, Ni, P, Pb, V, Zn, Zr e S nas camadas 0-10 e 10-30 cm.....	92
Tabela 29- Análise descritiva dos atributos químicos do solo nas profundidades 0-10 cm e 10-30 cm.....	93

Tabela 30- Valores de R^2 gerados pelos três métodos de predição dos atributos químicos do solo na profundidade 0-10 cm	94
Tabela 31- Valores de RMSE encontrados para Al, Fe, Mn, P, Pb, S, Ti, V e Zn na camada 0-10 cm, nos diferentes método testados	95
Tabela 32- Valores de R^2 gerados pelos três métodos de predição dos atributos químicos do solo na profundidade 10-30 cm	96
Tabela 33- Valores de RMSE encontrados para Al, Ca, Cr, Fe, K,P, Pb, S, Ti, V e Zr na camada 10-30 cm, nos diferentes método testados	96
Tabela 34- Principais variáveis independentes e número total de variáveis usadas pelos modelos de predição de atributos químicos do solo na profundidade 0-10 cm.....	99
Tabela 35- Principais variáveis independentes e número total de variáveis usadas pelos modelos de predição de atributos químicos do solo na profundidade 10-30 cm.....	100
Tabela 36- Valores máximos e mínimos encontrados para Al, Fe, K, Mn, P, Pb, S, Ti, V e Zn na análise de raios-x e pelos modelos de predição, na profundidade 0-10 cm.....	104
Tabela 37- Valores máximos e mínimos encontrados para Al, Ca, Cr, Fe, K, P, Pb, S, Ti, V e Zr na análise de raios-x e pelos modelos de predição na profundidade 10-30 cm.....	105
Tabela 38- Estatística descritiva para os atributos argila, fósforo (P) e carbono orgânico (CO)	136
Tabela 39- Valores do coeficiente de determinação (R^2), para variável argila, encontrados para os diferentes métodos e pré-processamentos no grupo no treinamento, nos espectros Mir, Vis-NIR e Vis-NIR-MIR	138
Tabela 40- Valores de RMSE, para variável argila., encontrados para os diferentes métodos e pré-processamentos no grupo no treinamento (T) e validação (V), nos espectros Mir, Vis-NIR e Vis-NIR-MIR.....	139
Tabela 41- Valores do coeficiente de determinação (R^2), para variável carbono orgânico, encontrados para os diferentes métodos e pré-processamentos no grupo no treinamento, nos espectros Mir, Vis-NIR e Vis-NIR-MIR	144
Tabela 42- Valores de RMSE, para variável carbono orgânico, encontrados para os diferentes métodos e pré-processamentos no grupo no treinamento (T) e validação (V), nos espectros Mir, Vis-NIR e Vis-NIR-MIR	145
Tabela 43- Valores do coeficiente de determinação (R^2), para variável fósforo, encontrados para os diferentes métodos e pré-processamentos no grupo no treinamento, nos espectros Mir, Vis-NIR e Vis-NIR-MIR	152
Tabela 44- Valores de RMSE, para variável fósforo, encontrados para os diferentes métodos e pré-processamentos no grupo no treinamento (T) e validação (V), nos espectros Mir, Vis-NIR e Vis-NIR-MIR.....	153
Tabela 45- Correlação não linear entre altitude e dados geofísicos	180
Tabela 46- Valores de temperatura, altitude e área para os diferentes usos do solo na bacia do Ribeirão Inhaúma, Iconha/ES.....	182
Tabela 47- Teste de normalidade para a profundidade 0-10 cm	182
Tabela 48- Variância total explicada pelos componentes principais.....	183
Tabela 49- Composição dos fatores após a rotação dos eixos dos dados de solo coletados na bacia do Ribeirão Inhaúma, Iconha/ES	184
Tabela 50- Teores médios de nutrientes no solo, na camada 0-10 cm obtidos por termofluorescência.....	185
Tabela 51- Correlação de Pearson entre dados obtidos por termofluorescência de raios-x, dados gamaespectrométricos e temperatura	187
Tabela 52- Correlação não linear- RDC entre dados obtidos por termofluorescência de raios-x, dados gamaespectrométricos e temperatura	188

Tabela 53- Teores obtidos para SiO ₂ , Al ₂ O ₃ , TiO ₂ e Fe ₂ O ₃ através da análise de fluorescência, e índices Ki e Kr no horizonte B de latossolos e cambissolos	202
Tabela 54- Teores obtidos para SiO ₂ , Al ₂ O ₃ , TiO ₂ e Fe ₂ O ₃ através da análise de ataque total, índices Ki e Kr no horizonte B de latossolos e cambissolos	203
Tabela 55- Análise granulométrica do horizonte B dos perfis amostrados no Ribeirão Inhaúma, Iconha/ES.....	203

LISTA DE FIGURAS

Figura 1- Localização geográfica da área de estudo.	2
Figura 2- Localização dos perfis e amostras extras coletadas na Bacia do Ribeirão Inhaúma-Iconha/ES.	3
Figura 3- Temperatura (A) e precipitação (B) médias mensais no ano de 2016 na área de estudo (INMET, 2017).	4
Figura 4- Balanço hídrico climatológico para o período de 1931 a 1990, em Cachoeiro de Itapemirim/ES (INMET, 2017).	5
Figura 5- Vista parcial do relevo na bacia do Ribeirão Inhaúma – Iconha/ES.	5
Figura 6- Área de estudo.	16
Figura 7- Covariáveis do terreno utilizadas para predição das classes de solo na bacia do Ribeirão Inhaúma-Iconha/ES.	17
Figura 8- Mapa de uso e ocupação do solo.	18
Figura 9- Fluxograma para predição de classes de solo utilizando a Lógica Fuzzy.	19
Figura 10- Boxplots referentes aos atributos do terreno em cada classe em que foram utilizados para predição.	24
Figura 11- Mapas de probabilidade de ocorrência de cada classe de solo obtidos a partir da inferência <i>fuzzy</i>	26
Figura 12- Mapa final de classes de solo obtido por inferência <i>fuzzy</i>	27
Figura 13- Mapa final de classes de solo obtido pelo método convencional de mapeamento.	30
Figura 14- Mapa de concordância entre o método digital (Lógica <i>fuzzy</i>) e o mapa convencional na bacia do ribeirão Inhaúma, Iconha/ES.	32
Figura 15- Fluxograma de funcionamento do algoritmo <i>Boosting</i>	44
Figura 16- Esquema de funcionamento do algoritmo <i>bagging</i> . T é a base de treinamento, H são as hipóteses.	46
Figura 17- Esquema de classificação por meio do Support Vector Machines-SVM.	50
Figura 18- SVM com suavização de margens.	51
Figura 19- Afloramento rochoso na bacia do ribeirão Inhaúma.	52
Figura 20- Cambissolo Háplico latossólico com plantio consorciado de café e bananeira. ...	53
Figura 21- Gleissolo Háplico em seu ambiente de ocorrência na bacia do ribeirão Inhaúma.	54
Figura 22- Latossolo Vermelho-Amarelo.	54
Figura 23- Neossolo Litólico.	55
Figura 24- Localização dos perfis e pontos de observação na bacia do Ribeirão Inhaúma, Iconha/ES.	56
Figura 25- Classificação do solo segundo o método <i>Random Forest</i>	72
Figura 26- Classificação do solo segundo o método <i>Ranger</i>	72
Figura 27- Classificação do solo segundo o método WSRF.	73
Figura 28- Classificação do solo segundo o método XGBTREE.	73
Figura 29- Bacia do Ribeirão Inhaúma, localizada em Iconha/ES.	88
Figura 30- Modelo digital de elevação.	88
Figura 31- Teores totais de Alumínio (Al_2O_3), Ferro (Fe_2O_3), Potássio (K_2O) e Titânio (TiO_2) usando os regressores GLMBOOST (Al), RF (Fe), GMB (K) e RIDGE (Ti), em 0-10 cm. <i>Random Forest</i> -RF, <i>Ridge regression</i> -RIDGE, <i>Generalized Boosted Regression Models</i> -GBM e <i>Gradient Boosting with Component-wise Linear Models</i> – GLMBOOST.	101

Figura 32- Teores totais de Manganês (MnO), Fósforo (P ₂ O ₅), Chumbo (PbO) e Vanádio (V ₂ O ₅) usando regressores RF (Mn e V), Cubist (Pb), PLS (P) em 0-10 cm. <i>Random Forest-RF, Partial Least Squares-PLS</i>	101
Figura 33- Teores totais de Zinco (ZnO), Enxofre (SO ₃) usando o regressor RIDGE em 0-10 cm. <i>Ridge regression-RIDGE</i>	102
Figura 34- Teores totais de Alumínio (Al ₂ O ₃), Cálcio (CaO), Ferro(Fe ₂ O ₃) e Potássio (K ₂ O) usando o regressor Random Forest em 10-30 cm.....	102
Figura 35- Teores totais de Titânio (TiO ₂), Cromo (Cr ₂ O ₃), Fósforo (P ₂ O ₅) e Chumbo (PbO) usando os regressores GBM (Ti), FOBA (Cr), RF (P e Pb) em 10-30 cm. <i>Random Forest-RF, Adaptive Forward-Backward Greedy Algorithm-FOBA, Generalized Boosted Regression Models- GBM</i>	103
Figura 36- Teores totais de Vanádio (V ₂ O ₅), Zircônio (ZrO ₂) e Enxofre (SO ₃), usando os regressores GBM (V), RF (Zr) e PLS (S) em 0-30 cm. <i>Random Forest-RF, Partial Least Squares-PLS, Generalized Boosted Regression Models- GBM</i>	103
Figura 37- Comportamento espectral para um solo siltoso e arenoso, ambos secos.....	121
Figura 38- Curvas espectrais dos horizontes A e B de um Latossolo Vermelho distroférico típico (a) e de um Latossolo Amarelo árcico típico (b).....	122
Figura 39- Curva espectral da gibbsita.	123
Figura 40- Espectrorradiômetro LI-COR 1800.	124
Figura 41- Localização dos pontos de amostragem.	128
Figura 42- Espectro MIR sem pré-processamento	130
Figura 43- Espectro VIS-NIR sem pré-processamento	130
Figura 44- Espectro VIS-NIR-MIR sem pré-processamento	131
Figura 45- Pré-processamentos aplicados no espectro MIR 1- <i>Absorbance (ABS)</i> 2- <i>Continuum removal (CR)</i> 3- <i>Multiplicative Scatter Correction (MSC)</i> 4- <i>Savitzky Golay Derivative (SGD)</i> e 5- <i>Satandard Normal Variate (SNV)</i>	132
Figura 46- Pré-processamentos aplicados no espectro VIS-NIR 1- <i>Absorbance (ABS)</i> 2- <i>Continuum removal (CR)</i> 3- <i>Multiplicative Scatter Correction (MSC)</i> 4- <i>Savitzky Golay Derivative (SGD)</i> e 5- <i>Satandard Normal Variate (SNV)</i>	133
Figura 47- Pré-processamentos aplicados no espectro VIS-NIR-MIR 1- <i>Absorbance (ABS)</i> 2- <i>Continuum removal (CR)</i> 3- <i>Multiplicative Scatter Correction (MSC)</i> 4- <i>Savitzky Golay Derivative (SGD)</i> e 5- <i>Satandard Normal Variate (SNV)</i>	134
Figura 48- Histogramas de distribuição das três propriedades do solo: (a) argila total; (b) carbono orgânico; (c) fósforo	135
Figura 49- Valores do coeficiente de determinação (R ²), para variável argila., encontrados para os diferentes métodos e pré-processamentos no grupo na validação.	137
Figura 50- Valores de RPIQ encontrados para os diferentes métodos e pré-processamentos no grupo na validação, para variável argila.	140
Figura 51- Valores de RPD encontrados para os diferentes métodos e pré-processamentos no grupo na validação, para variável argila.	141
Figura 52- Box splot da distribuição dos dados de RPIQ nos diferentes pré-processamentos e faixas do espectro, para variável argila.	142
Figura 53- Box splot da distribuição dos dados de RPD nos diferentes pré-processamentos e faixas do espectro, para variável argila.	142
Figura 54- Variáveis importantes para predição de argila, no espectro Vis-NIR (Visível e infravermelho próximo) usando pré-processamento <i>Multiplicative Scatter Correction (MSC)</i> e modelo <i>Support Vector Machine (SVM)</i>	143
Figura 55- Valores do coeficiente de determinação (R ²), encontrados para variável carbono orgânico, nos diferentes métodos e pré-processamentos no grupo na validação.	144

Figura 56- Valores de RPIQ encontrados para os diferentes métodos e pré-processamentos no grupo na validação, para variável carbono orgânico.	146
Figura 57- Valores de RPD encontrados para os diferentes métodos e pré-processamentos no grupo na validação, para variável carbono orgânico.	147
Figura 58- Box splot da distribuição dos dados de RPIQ nos diferentes modelos e faixas do espectro, para variável carbono orgânico.	149
Figura 59- Box splot da distribuição dos dados de RPD nos diferentes modelos e faixas do espectro, para variável carbono orgânico.	149
Figura 60- Variáveis importantes para predição de carbono orgânico, no espectro MIR (Infravermelho médio) usando pré-processamento <i>Continuum removal</i> (CR) e modelo <i>Partial Least Squares Regression</i> (PLSR).....	150
Figura 61- Valores do coeficiente de determinação (R^2), encontrados para variável fósforo, nos diferentes métodos e pré-processamentos no grupo na validação.	151
Figura 62- Valores de RPIQ encontrados para os diferentes métodos e pré-processamentos no grupo na validação, para variável fósforo.	154
Figura 63- Valores de RPD encontrados para os diferentes métodos e pré-processamentos no grupo na validação, para variável fósforo.	154
Figura 64- Box splot da distribuição dos dados de RPIQ nos diferentes modelos e faixas do espectro, para variável fósforo.	156
Figura 65- Box splot da distribuição dos dados de RPD nos diferentes modelos e faixas do espectro, para variável fósforo.	156
Figura 66- Variáveis importantes para predição de fósforo, no espectro Vis-NIR (Visível e Infravermelho próximo) usando pré-processamento <i>Satandard Normal Variate</i> (SNV) e modelo <i>Partial Least Squares Regression</i> (PLSR).	157
Figura 67- Principais classes de uso e ocupação do solo definidas na Bacia do Ribeirão Inhaúma em Iconha/ES.....	175
Figura 68- Localização dos pontos de amostragem na bacia do Ribeirão Inhaúma, Iconha/ES.	176
Figura 69- Mapa de uso e ocupação do solo.	178
Figura 70- Mapa de Potássio (%), Tório (ppm) e Urânio (ppm).....	179
Figura 71- Modelo digital de elevação.....	180
Figura 72- Mapa de temperatura da superfície para a bacia do Ribeirão Inhaúma, Iconha/ES.	181
Figura 73- Gráficos de dispersão para correlações maiores que 0,8.	189
Figura 74- Malha de amostragem para análise de Fe, Al, Si e Ti.	200
Figura 75- Correlação entre teores obtidos de Alumínio, Titânio, Ferro e Silício, por ataque total e fluorescência de raios-x.	204

SUMÁRIO

1	INTRODUÇÃO GERAL	1
2	ÁREA DE ESTUDO	2
2.1	Caracterização Do Meio Físico	3
2.2.1	Clima e balanço hídrico	3
2.2.2	Vegetação	5
2.2.3	Relevo	5
2.2.4	Hidrografia	6
2.2.5	Geologia e Geomorfologia	6
3	REFERÊNCIAS BIBLIOGRÁFICAS	6
	CAPÍTULO I	9
	O USO DA LÓGICA FUZZY PARA O MAPEAMENTO DIGITAL DE SOLOS NO ESPÍRITO SANTO	9
	RESUMO	10
	ABSTRACT	10
1	INTRODUÇÃO	11
2	REVISÃO BIBLIOGRÁFICA	12
2.1	Mapeamento Digital De Solos	12
2.2	Lógica Fuzzy	13
2.3	Uso da Lógica <i>fuzzy</i> na predição de classes de solo	14
3	MATERIAL E MÉTODOS	16
4	RESULTADOS E DISCUSSÃO	23
5	CONCLUSÕES	32
6	REFERÊNCIAS BIBLIOGRÁFICAS	33
	CAPÍTULO II	40
	USO DE <i>MACHINE LEARNING</i> PARA PREDIÇÃO DE CLASSES DE SOLOS APLICADOS A UMA BACIA HIDROGRÁFICA	40
	RESUMO	41
	ABSTRACT	41
1	INTRODUÇÃO	42
2	REVISÃO BIBLIOGRÁFICA	43
2.1	Classificadores Utilizados Para Predição De Classes De Solo	43
2.1.1	<i>Bagged AdaBoost</i>	43
2.1.2	C5.0	44

2.1.3	<i>Conditional Inference Random Forest</i>	45
2.1.4	<i>Extreme Gradient Boosting</i>	46
2.1.5	<i>Multivariate Adaptive Regression Splines</i>	48
2.1.6	<i>Random Forest</i>	48
2.1.7	<i>Ranger</i>	49
2.1.8	<i>Support Vector Machine com Kernel Linear e Polynomial</i>	50
2.1.9	<i>Weighted Subspace Random Forest</i>	51
2.2	Principais Classes De Solo Encontradas Na Área De Estudo	52
2.2.1	Afloramentos rochosos com e sem vegetação	52
2.2.2	Cambissolo Háplico	52
2.2.3	Gleissolos Háplico	53
2.2.4	Latossolos	54
2.2.5	Neossolo Litólico	55
3	MATERIAL E MÉTODOS	55
4	RESULTADOS E DISCUSSÃO	63
5	CONCLUSÕES	74
6	REFERÊNCIAS BIBLIOGRÁFICAS	74
	CAPÍTULO III	81
	MAPEAMENTO DIGITAL DE ATRIBUTOS DO SOLO UTILIZANDO APRENDIZADO DE MÁQUINAS	81
	RESUMO	82
	ABSTRACT	82
1	INTRODUÇÃO	83
2	REVISÃO BIBLIOGRÁFICA	84
2.1	Classificadores Utilizados Para Predição De Atributos	84
2.1.1	<i>Random Forest</i>	84
2.1.2	<i>Ridge Regression</i>	84
2.1.3	<i>Cubist</i>	85
2.1.4	<i>Partial Least Squares- PLS</i>	85
2.1.5	Regressão por componentes principais- PCR	86
2.1.6	<i>Adaptive Forward-Backward Greedy Algorithm - FOBA</i>	86
2.1.7	<i>Generalized Boosted Regression Models- GBM</i>	86
2.1.8	<i>Gradient Boosting with Component-wise Linear Models - GLMBOOST</i>	87
3	MATERIAL E MÉTODOS	87
4	RESULTADOS E DISCUSSÃO	91
5	CONCLUSÕES	105
6	REFERÊNCIAS BIBLIOGRÁFICAS	105

CAPÍTULO IV	116
PREDIÇÃO DE ATRIBUTOS DO SOLO UTILIZANDO DADOS ESPECTRAIS	116
RESUMO	117
ABSTRACT	117
1 INTRODUÇÃO	118
2 REVISÃO BIBLIOGRÁFICA	120
2.1 Comportamento Espectral Do Solo	120
2.2 Espectroscopia De Reflectância	122
2.3 Infravermelho Próximo e Médio (NIR E MIR)	124
3 MATERIAL E MÉTODOS	127
4 RESULTADOS E DISCUSSÃO	130
5 CONCLUSÕES	158
6 REFERÊNCIAS BIBLIOGRÁFICAS	158
CAPÍTULO V	168
LEVANTAMENTO DE ATRIBUTOS DO SOLO UTILIZANDO ESPECTROMETRIA DE RAIOS-X : COMPARAÇÃO COM DADOS AEROGEOFÍSICOS E MAPA DE USO E OCUPAÇÃO DO SOLO	168
RESUMO	169
ABSTRACT	169
1 INTRODUÇÃO	170
2 REVISÃO BIBLIOGRÁFICA	172
2.1 Espectrometria Por Fluorescência De Raios-X	172
2.2 Estudo Da Correlação Entre Dados Aerogeofísicos E Dados Obtidos Por Termofluorescência	173
3 MATERIAL E MÉTODOS	174
4 RESULTADOS E DISCUSSÃO	177
5 CONCLUSÕES	190
6 REFERÊNCIAS BIBLIOGRÁFICAS	190
CAPÍTULO VI	195
CORRELAÇÃO DE TEORES DE SILÍCIO, ALUMÍNIO, TITÂNIO E FERRO OBTIDOS POR ATAQUE TOTAL E FLUORESCÊNCIA DE RAIOS-X	195
RESUMO	196
ABSTRACT	196
1 INTRODUÇÃO	197
2 REVISÃO BIBLIOGRÁFICA	198
2.1 Ataque Sulfúrico	198
2.2 Índices Ki e Kr	198
2.3 Fluorescência De Raios-X	199

3	MATERIAL E MÉTODOS	200
4	RESULTADOS E DISCUSSÃO	201
5	CONCLUSÕES	204
6	REFERÊNCIAS BIBLIOGRÁFICAS	205
	CONCLUSÕES GERAIS	207
	ANEXOS	208
	APÊNDICES	213

1 INTRODUÇÃO GERAL

O mapa de solos é uma ferramenta essencial para otimização do uso e ocupação dos solos. O processo de produção convencional dos mapas de solos é demorado e apresenta um alto custo, não permitindo uma rápida atualização de forma acurada (ZHU et al., 2010), trata-se ainda de uma atividade com elevado grau de subjetividade, o que impede a avaliação do resultado de forma sistemática.

O método convencional, estabelecido por Dokuchaev nos primórdios da Ciência do Solo, descreve que o solo é o resultado da interação entre cinco fatores: clima (*Climate* - Cl), organismos (*Organisms* - O), relevo (*Relief* - R), material de origem (*Parent material* - P) e tempo (*Time* - T) (MENDONÇA-SANTOS E SANTOS, 2003). Os quatro primeiros fatores interagindo no tempo, criam uma série de processos específicos que levam à diferenciação em horizontes e, conseqüentemente, à formação do solo.

No método Clorpt, a taxonomia de solos, é o modelo usado pelo pedólogo, para criar e distinguir as classes de solos e estabelecer seus limites. É um método intuitivo, determinístico, onde considera-se a existência de alta correlação entre o tipo de solo e o ambiente onde ele ocorre (NEUMANN, 2012).

Dessa forma, a interpretação será mais adequada quanto melhores forem as informações disponíveis (SILVA, 2012). Assim, o levantamento pedológico convencional é baseado no modelo discreto e descreve características dos solos de uma determinada área, classificados de acordo com um sistema taxonômico vigente, estabelece limites entre classes definidas no mapa e permite fazer inferências sobre o comportamento dos solos quanto ao uso e ao manejo (MENDONÇA-SANTOS E SANTOS, 2003).

Um dos levantamentos de solos realizados segundo este método foram os de reconhecimento executados pela Divisão de Pedologia do Ministério da Agricultura, que corresponde hoje ao Centro Nacional de Pesquisas de Solos/Embrapa Solos (CNPS) (CHAGAS, 2006).

Embora bastante difundidos, existem alguns gargalos, como o fato dos conceitos empregados serem gerais e descritivos e não haver explicação sobre o fato de se realizar o mapeamento do solo, algo com tamanha variabilidade, realizando-se pouca amostragem (HUDSON, 1990). Muitos autores têm criticado o método convencional, devido aos seus aspectos subjetivos (McBRATNEY et al., 1981; BURROUGH, 1989). O método convencional não considera a dependência espacial entre as unidades de mapeamento, a qual pode ser grande, especialmente em levantamentos detalhados e/ou em áreas onde os limites entre os solos não são óbvios (SILVA, 2012).

Atualmente, cerca de 35% do território brasileiro é coberto por mapas de solos, obtidos pelo método convencional, em escalas variando entre 1:100 000 a 1:600 000 (SANTOS, 2007), escalas essas que não permitem um planejamento mais pontual, visto que não é possível um nível de detalhamento adequado para tal.

Dessa forma, o mapeamento digital de solo (MDS) tem sido impulsionado devido a grande demanda por informações sobre a distribuição espacial de propriedades e classes de solos (CATEN, 2011). Segundo Sanchez et al. (2009), essa predição é realizada por meio de equações matemáticas, onde é possível obter classes e propriedades de solos e mapeá-las de forma contínua e espacial, o que possibilita a organização de extensos conjuntos de dados para análise e interpretações em qualquer época. Outra vantagem desses métodos, consiste na possibilidade de se realizar a avaliação da acurácia dos mapas digitais de solos concomitantemente à sua geração (SILVA, 2012).

O uso de técnicas computacionais para manipulação, análise e distribuição de dados espaciais tem contribuído para maior eficiência e diminuição do custo de execução dos levantamentos de solos no mundo inteiro (PINHEIRO, 2012). No Brasil, o Mapeamento Digital de Solos (MDS), está deixando de ser uma possibilidade para pesquisa e caminhando para sua consolidação como técnica de auxílio ao mapeamento de solo (CATEN et al, 2011).

O objetivo geral deste trabalho é mapear digitalmente os solos em bacia hidrográfica localizada no estado do Espírito Santo utilizando a lógica fuzzy e outros métodos e validar o mapa através da validação de campo, índice Kappa, matriz de confusão e exatidão global.

Os objetivos específicos são: Mapear os solos da bacia hidrográfica do Ribeirão Inhaúma no município de Iconha/ES utilizando a Lógica *Fuzzy* comparando-a com o mapa obtido de forma convencional; Avaliar a acurácia do modelo de lógica *fuzzy* e de outros métodos na predição de classes do solo em uma bacia hidrográfica no estado do Espírito Santo; Aferir a confiabilidade do mapa digital de classes de solos por meio de validação de campo; Determinar a acurácia do mapa digital de classes de solos por meio do índice Kappa, matriz de confusão e exatidão global; propor o uso de técnicas digitais para predição de classes e atributos de solos no Espírito Santo; Mapear digitalmente os atributos do solo na área utilizando diferentes técnicas.

2 ÁREA DE ESTUDO

O presente trabalho foi realizado no estado do Espírito Santo, na bacia hidrográfica do Ribeirão Inhaúma, coordenada $21^{\circ}10'58.82''S$ e $41^{\circ}00'08.87''O$, localizada quase que totalmente no município de Iconha, no sul do estado do Espírito Santo (Figura 1), com área total de 2403,9 ha.

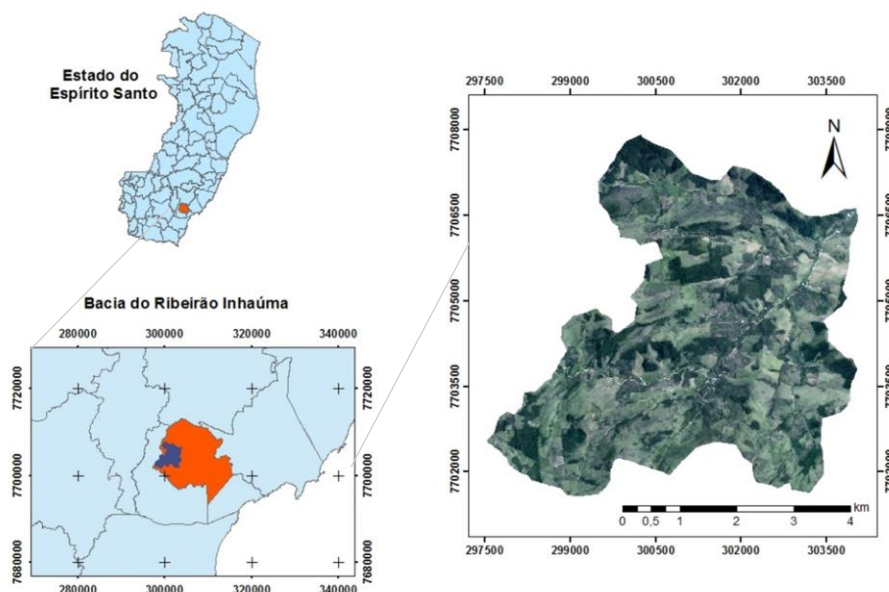


Figura 1- Localização geográfica da área de estudo.

A comunidade Inhaúma fica na porção Oeste do município, confrontando com Rio Novo do Sul (Oeste) e com as comunidades de Alto Rio Novo (Sul), Nova Esperança (Leste) e São José (Norte). A nascente do Ribeirão Inhaúma fica a 770 m de altitude sob o bioma da Mata Atlântica. O uso do solo deste na região é prioritariamente de agricultura, pastagem e mata. Segundo o Instituto Capixaba de Pesquisa e Extensão Rural (INCAPER, 2016) nas áreas plantadas predominam as culturas da banana e café.

A agricultura é predominantemente familiar, caracterizada por pequenas propriedades, não estando presente na região assentamentos rurais (INCAPER, 2011). Segundo o INCRA (2011), a estratificação fundiária é dada por 897 minifúndios, 332 pequenas propriedades, 23 médias propriedades e nenhuma grande propriedade, totalizando 1 252.

De acordo com o mapa de solos, elaborado pela Embrapa, 2001, na região estudada encontram-se Latossolo Vermelho-Amarelo, Cambissolo Háplico e afloramentos de rocha.

As imagens ortorretificadas da área, utilizadas neste estudo, foram cedidas pelo INCAPER, por meio do Sistema Integrado de Bases Geoespaciais do Estado do Espírito Santo- GEOBASES, na escala 1: 8 000.

Foram descritos na área 12 perfis e coletadas duas amostras extras. A localização desses encontra-se na Figura 2. A descrição encontra-se nos apêndices.

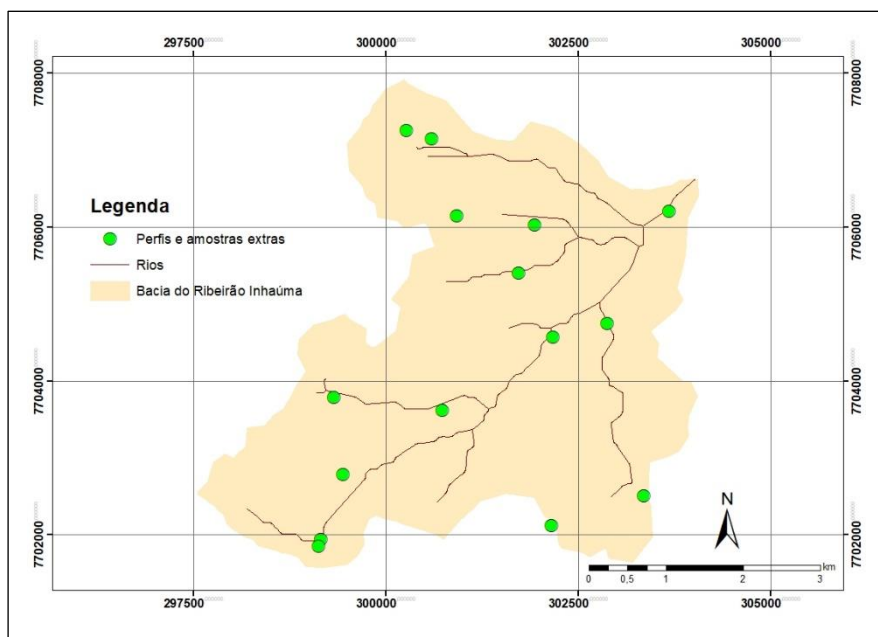


Figura 2- Localização dos perfis e amostras extras coletadas na Bacia do Ribeirão Inhaúma- Iconha/ES.

2.1 Caracterização Do Meio Físico

2.2.1 Clima e balanço hídrico

De acordo com a classificação de Köppen, o clima da área de estudo é Aw, o qual é caracterizado por apresentar clima quente e úmido, temperatura média entre 19 e 28 °C,

pluviosidade inferior a 2 000 mm/ano e duas estações bem definidas, sendo o verão chuvoso e o inverno seco.

De acordo com a Figura 3 é possível perceber que no ano de 2016 a temperatura oscilou entre 17 e 31°C aproximadamente. E os meses mais chuvosos ocorreram em março e dezembro (INMET, 2017).

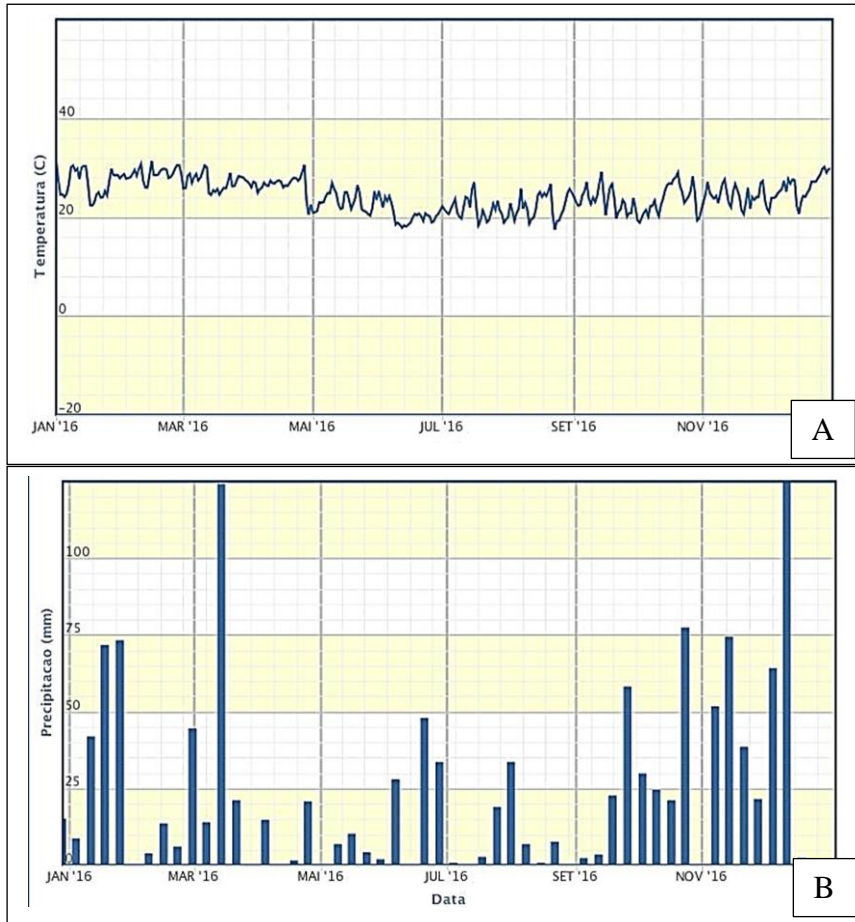


Figura 3- Temperatura (A) e precipitação (B) médias mensais no ano de 2016 na área de estudo (INMET, 2017).

O balanço hídrico para as séries históricas de 1931 a 1990, são apresentados na Figura 4. Os dados são da estação de Cachoeiro de Itapemirim (INMET, 2017) e representam a área de estudo.

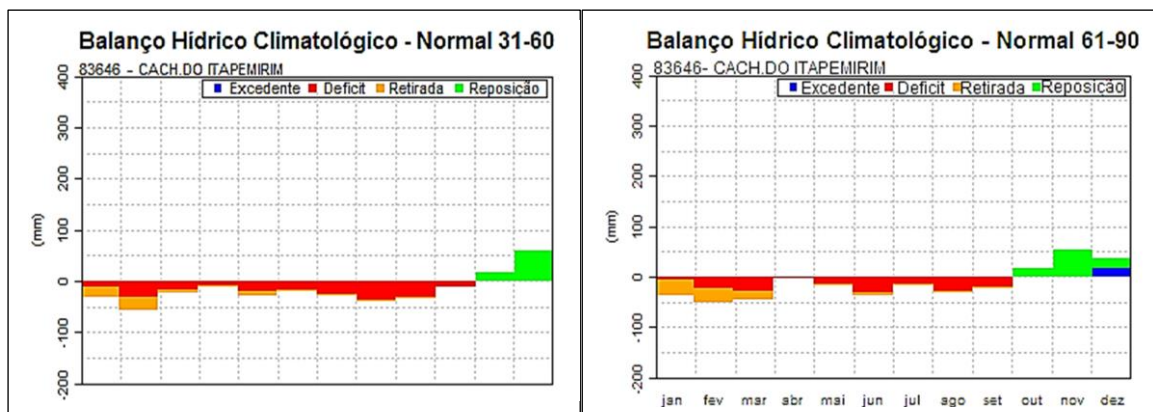


Figura 4- Balanço hídrico climatológico para o período de 1931 a 1990, em Cachoeiro de Itapemirim/ES (INMET, 2017).

2.2.2 Vegetação

A área de estudo encontra-se no bioma mata atlântica (BRASIL, 2008). A cobertura Florestal da área é a floresta Ombrófila Densa (Floresta Tropical Pluvial) próximo a transição para áreas de floresta Estacional Semidecidual (Floresta Tropical Subcaducifólia) (IBGE, 2017). Essa área de transição poderia explicar o motivo da existência de um balanço hídrico com tantos meses no negativo, mesmo a bacia estando em área classificada como Ombrófila Densa.

É possível perceber, através de visitas a campo, que a vegetação local é representada, principalmente, por áreas de floresta e agricultura, destacando-se nesta café e banana, além de pastagens.

2.2.3 Relevo

O relevo é uma característica relevante na área, visto que se apresenta bastante montanhoso, com áreas escarpadas. A altitude varia de uma cota de 249 m a 1 033m. A Figura 5 mostra uma visão do relevo de parte da bacia.



Figura 5- Vista parcial do relevo na bacia do Ribeirão Inhaúma – Iconha/ES.

2.2.4 Hidrografia

A região de Iconha está inserida na octobacia do Rio Benevente. A área de estudo é banhada pelo Ribeirão Inhaúma, que é um afluente do Rio Iconha, e este desagua diretamente no mar (IJSN, 2017).

2.2.5 Geologia e Geomorfologia

O estado do Espírito Santo é dividido em três grandes unidades geológicas: Unidades do pré-cambriano, unidades terciárias e unidades quaternárias. A área de estudo está localizada na unidade pré-cambriana, onde se localizam rochas muito antigas, sendo sua geomorfologia a região serrana, onde o embasamento cristalino compreende os remanescentes de cadeias dobradas e faixa de dobramentos remobilizados (BRASIL, 2005).

De acordo com o Instituto Jones dos Santos Neves- IJSN (2012), a morfoestrutura do local de estudo são faixas de dobramentos remobilizados, que são caracterizadas pelas evidências de movimentos crustais, com marcas de falhas, deslocamentos de blocos e falhamentos transversos, o que impõe claro controle estrutural sobre a morfologia que se tem atualmente.

Ainda de acordo com IJSN (2012), a região é conhecida por planaltos da Mantiqueira setentrional, que tem apecto montanhoso fortemente dissecado, com altitudes variadas, onde os níveis altimétricos são relacionados com as fases de dissecação comandadas pelos rios. A unidade em que se encontra a área de estudo são os maciços do Caparaó I, cujas altitudes médias são em torno de 600 m, com algumas grandes elevações superiores a 2 000 m de altitude.

Segundo a CPRM (2015), existe no município o Maciço Iconha, que é representado por dois corpos, estando um a norte e outro ao sul, o município de Iconha está situado ao sul. Esse maciço é composto por rochas dioríticas a granodioríticas rodeadas por rochas graníticas (FÉBOLI, 1993).

Em campo foi possível observar a predominância de metatonalitos.

3 REFERÊNCIAS BIBLIOGRÁFICAS

BRASIL. Decreto nº 6.660, de 21 de novembro de 2008. **Regulamenta dispositivos da Lei nº 11.428, de 22 de dezembro de 2006, que dispõe sobre a utilização e proteção da vegetação nativa do Bioma Mata Atlântica.** Disponível em < https://www.planalto.gov.br/ccivil_03/_ato2007-2010/2008/decreto/d6660.htm> Acesso em: Mar, 2017.

BRASIL. Ministério do Meio Ambiente. **Panorama sobre desertificação no estado do Espírito Santo.** 2005. Disponível em: < http://www.mma.gov.br/estruturas/sedr_desertif/_arquivos/panorama_espiritosanto.pdf > Acesso em: Mar., 2017.

BURROUGH, P. A. Fuzzy mathematical methods for soil survey and land evaluation. Journal of Soil Science, v. 40, n.3,p. 477-482, 1989.

CATEN, A. T . Mapeamento digital de solos: metodologias para atender a demanda por informação espacial em solos. 106 p . 2011. (Tese de Doutorado). Universidade Federal de Santa Maria.

CATEN, A.T.; DALMOLIN, R.S.D; PEDRON, F.A.; SANTOS, M.L.M. Extrapolação das relações solo-paisagem a partir de uma área de referência. Ciência Rural, v.41, n.5,p. 812-816, 2011.

CHAGAS,C. DA S. Mapeamento digital de solos por correlação ambiental e redes neurais em uma bacia hidrográfica de domínio de mar de morros. 223 p. 2006 (Tese de Doutorado).Universidade Federal de Viçosa.

CPRM- Companhia de Pesquisa de Recursos Minerais/Serviço Geológico do Brasil. **Geologia e Recursos Minerais do Estado do Espírito Santo: texto explicativo do mapa geológico e de recursos minerais.** Belo Horizonte : CPRM, 2015.

FÉBOLI, W. L. **Piúma, folha SE.24-Y-C-VI, estado do Espírito Santo:** texto explicativo. Brasília: DNPM; CPRM. 140 p., 1993.

HUDSON, B.D. Concepts of soil mapping and interpretation. Soil Survey Horizons. v. 31,n.3 p.63-72, 1990.

IBGE - EMBRAPA - Mapa de Solos do Brasil. Rio de Janeiro: IBGE, 2001 - Escala 1:5.000.000.

IBGE. Instituto Brasileiro de Informações Geográficas. Disponível em <www.cidades.ibge.gov.br>. Acesso em: Mar., 2017.

IJSN. Instituto Jones dos Santos Neves. Disponível em <www.ijsn.es.gov.br>. Acesso em: Mar., 2017.

IJSN. Instituto Jones dos Santos Neves. Mapeamento Geomorfológico do estado do Espírito Santo. Nota técnica 28. Governo do Espírito Santo: Vitória, ES. 19 p. 2012.

INCAPER. Instituto Capixaba de pesquisa, assistência técnica e extensão rural. Programa de assistência técnica e extensão rural 2011-2013. Iconha. 29 p. 2011.

INCRA. Instituto nacional de colonização e reforma agrária. 2011. Disponível em <www.incra.gov.br>. Acesso em: Mai, 2017.

INMET. Instituto Nacional de Meteorologia. Disponível em <<http://www.inmet.gov.br>>. Acesso em: Mar., 2017.

INSTITUTO CAPIXABA DE PESQUISA E EXTENSÃO RURAL- INCAPER. Município de Iconha. Disponível em <<http://incaper.es.gov.br/media/incaper/proater/municipios/Caparao/Iconha.pdf>>. Acesso em Nov, 2016.

McBRATNEY, A. B.; WEBSTER, R. Spatial dependence and classification of soil along a transect in northeast Scotland. *Geoderma*, v. 26,n.1, p. 63-82, 1981.

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

NEUMANN, M. R. B.. Mapeamento digital de solos, no Distrito Federal. 123 p. 2012. (Tese de Doutorado). Universidade de Brasília. Brasília.

PINHEIRO, H.S.K.; CHAGAS, C.S.; CARVALHO JÚNIOR, W. ANJOS, L.H.C. Modelos de elevação para obtenção de atributos topográficos utilizados em mapeamento digital de solos. *Pesquisa Agropecuária Brasileira*. v.47, n.9, p.1384-1394, 2012.

SANCHEZ, P.A.; AHAMED, S.; CARRÉ, F.; HARTEMINK, A.E.;HEMPEL, J.; HUISING, J.; LAGACHERIE, P.; MCBRATNEY,A.B.; MCKENZIE, N.J.; MENDONCA-SANTOS, M. de L.;MINASNY, B.; MONTANARELLA, L.; OKOTH, P.; PALM, C.A.;SACHS, J.D.; SHEPHERD, K.D.; VÅGEN, T. -G.; VANLAUWE, B.; WALSH, M.G.; WINOWIECKI, L.A.; ZHANG, G. -L. Digital soil map of the world. *Science*, v.325,n 5941, p.680-681, 2009.

SANTOS, H. G. dos. Importância e evolução dos levantamentos de solo no Brasil. *Boletim Informativo da Sociedade Brasileira de Ciência do Solo*. Viçosa, MG, v. 32, n. 1 - Jan/Abr, 2007.

SILVA, Cristiano Cassiano. Mapeamento digital de classes de solo: aplicação de metodologia na folha Botucatu (SF-22-Z-B-IV-3) e validação de campo. 117 p. 2012. (Dissertação de Mestrado). Instituto Agrônomo de Campinas.

ZHU, Q.; LIN, H. S. Comparing ordinary kriging and regression kriging for soil properties in contrasting landscapes. *Pedosphere*, v. 20, n. 5, p. 594-606, 2010.

CAPÍTULO I

O USO DA LÓGICA FUZZY PARA O MAPEAMENTO DIGITAL DE SOLOS NO ESPÍRITO SANTO

RESUMO

Técnicas de mapeamento digital de solos têm sido empregadas em várias regiões de modo eficiente e apresenta-se como uma técnica rápida, que permite um mapeamento menos oneroso e por vezes mais hábil em relação ao convencional. Este trabalho objetivou mapear digitalmente os solos do Riberião Inhaúma em Iconha/ES, a partir de covariáveis ambientais obtidas por meio do MDE, utilizando-se a lógica *fuzzy*, comparando a eficiência do mapeamento digital e convencional na área, a partir de matrizes de confusão. O mapeamento foi realizado utilizando-se o software ArcGis e SAGA, onde a partir do Modelo digital de Elevação, foram gerados mapas de atributos do terreno. Desses mapas foram selecionados quatro (Modelo digital de elevação-MDE; *Channel Network - Chan_net*; *Topographic Wetness Index-TWI*; *Multiresolution index of valley bottom flatness-MRVBF*), que foram utilizados na extensão *ArcSIE* para modelagem utilizando a lógica *fuzzy*. O mapa convencional foi elaborado com base no MDE e em perfis coletados na área. A acurácia do mapa de solo produzido foi medida pelo índice Kappa e pela exatidão global determinadas à partir da matriz de confusão, para comparação dos dois métodos foi utilizado o teste Z. O mapa digital apresentou valores de Kappa de 0,75, sendo classificado como muito bom (Landis e Kochi, 1977), com acurácia global de 84%; Para o mapeamento convencional o índice Kappa encontrado foi de 0,35, e a acurácia global de 58%. O teste Z foi igual a 6,53 demonstrando a diferença estatística entre os dois métodos, onde o mapeamento digital mostrou-se superior. O mapeamento digital da região estudada foi eficiente utilizando-se lógica *fuzzy*; Há divergências entre o mapeamento digital e convencional, sendo necessário o uso de critérios diferenciados para a aplicação dos mesmos na área de estudo.

Palavras-chave: mapeamento digital, lógica difusa, *fuzzy logic*

ABSTRACT

Soil digital mapping techniques have been used in several regions in an efficient way and it is presented as a fast technique, which allows a mapping less costly and sometimes more skilful than the traditional one. This work aimed to digitally map the soils of the Riberião Inhaúma in Iconha / ES, from environmental covariates obtained through the MDE, using the fuzzy logic, comparing the efficiency of the digital and traditional mapping in the area, from matrices of confusion. The mapping was performed using the software ArcGis and SAGA, where from the Elevation digital model, terrain attribute maps were generated. These models were selected using four models (MDE elevation model, Channel Network - Chan_net, Topographic Wetness Index-TWI), which were used in the ArcSIE extension for modeling using fuzzy logic. The traditional map was elaborated based on the MDE and on profiles collected in the area. The accuracy of the produced soil map was measured by the Kappa index and the global accuracy determined from the confusion matrix. For the comparison of the two methods the Z test was used. The digital map presented Kappa values of 0.75, being classified as very good (Landis and Kochi, 1977), with an overall accuracy of 84%; For the traditional mapping, the Kappa index found was 0.35, and the overall accuracy was 58%. The Z test was equal to 6.53, showing the statistical difference between the two methods, where the digital

mapping showed to be superior. The digital mapping of the studied region was efficient using fuzzy logic; There are divergences between the digital and traditional mapping, and it is necessary to use differentiated criteria for the application of the same in the study area.

Key words: digital mapping, fuzzy logic, fuzzy logic

1 INTRODUÇÃO

O mapeamento convencional é o método mais utilizado no Brasil para a espacialização de classes de solo. São baseados em modelos mentais desenvolvidos pelo pedólogo que buscam compreender as relações entre solos e a paisagem. São altamente subjetivos, dependente da experiência do profissional e de difícil repetição. Entretanto, nos últimos anos vem evoluindo novas técnicas de mapeamento que se utilizam de modelos matemáticos, sustentados por recursos computacionais robustos e que utilizam covariáveis ambientais oriundas principalmente de modelos digitais de elevação e produtos de sensoriamento remoto.

Esses métodos possuem muitas interfaces, se distanciando na etapa de delineamento das unidades de mapeamento e na possibilidade de espacialização não somente de classe, mas também de atributos dos solos. Outra vantagem do método digital é a possibilidade de repetição do método, a determinação estatística do erro e a possibilidade de atualização fácil do mapa quando novos dados são introduzidos.

Para McBratney (2007), o mapeamento digital pode ser então definido como a criação e manipulação de sistemas de informação espaciais de solos, onde se usa modelos numéricos/matemáticos, associados a observações, conhecimento dos solos e variáveis ambientais correlacionadas, para inferência das variações espaciais e temporais dos tipos de solos e suas propriedades.

Dessa forma, diferentes técnicas de mapeamento digital de solos (MDS) vem sendo propostas, desde o uso de algoritmos classificadores, como aquelas baseadas em processos de interpolação, como a geoestatística (McBRATNEY et al., 2003). Dentre outras técnicas estão a lógica *fuzzy*, alvo de estudo deste trabalho, que foi introduzida por Zadeh (1965) e permite trabalhar com classes indefinidas e com limites indeterminados (NEUMANN, 2012).

Rodríguez et al. (2016) apontam o crescente número de aplicações bem-sucedidas da lógica *fuzzy* e da teoria dos conjuntos difusos para lidar com a incerteza, a imprecisão e a subjetividade inerentes às avaliações da qualidade ambiental.

A lógica *fuzzy* tem por objetivo modelar, de modo aproximado, o raciocínio humano, visando manipular informações em um ambiente de incerteza e imprecisão, fornecendo uma resposta aproximada para uma questão baseada em um conhecimento inexato, incompleto ou não totalmente confiável (NEUMANN, 2012; BÖNISCH et al., 2004).

Devido ao fato do relevo ser um dos fatores de formação do solo mais importantes (JENNY, 1941) e os modelos digitais de elevação (MDE) serem encontrados facilmente, sendo amplamente disponíveis em diferentes resoluções (BEHRENS et al., 2010), os atributos do terreno são os preditores mais utilizados no mapeamento digital do solo (McBRATNEY et al., 2003).

Assim, este trabalho teve como objetivo realizar mapeamento digital de solos na bacia hidrográfica do Riberião Inhaúma, Iconha/ES, a partir de covariáveis ambientais e utilizando-se a lógica *fuzzy*, comparando a eficiência do mapeamento digital e convencional na área, a partir de matrizes de confusão.

2 REVISÃO BIBLIOGRÁFICA

2.1 Mapeamento Digital De Solos

Conforme Lagacherie, 2008 mapeamento digital de solos é definido como a criação e manipulação de sistemas espaciais de informação de solos, através do uso de modelos numéricos para a inferência das variações espaciais e temporais dos tipos de solos e de suas propriedades, a partir de observações e conhecimento dos solos e de variáveis ambientais correlacionadas.

Um importante propósito do mapeamento digital de solos é auxiliar o mapeamento convencional, reduzindo os recursos humanos, materiais para trabalhos de campo e, conseqüentemente, gastos e tempo necessários para a elaboração dos mapas de solos (ARRUDA et al., 2013). Sarmiento (2010) relata que ao analisar a rotina dos trabalhos com mapeamentos digitais é possível constatar que a grande maioria segue algumas etapas que, podem ser consideradas equivalentes àquelas que são usadas nos levantamentos convencionais. O mapeamento digital, assim como o levantamento convencional, faz uso de informações coletadas ou disponíveis em pontos de observação de solos.

Essas etapas foram descritas por MacMillan (2008), que mostra que independentemente da abordagem, dos objetivos e dos métodos que são empregados, o processo de mapeamento digital de solos envolve seis passos. O primeiro consiste em definir o objetivo do mapeamento: prever classes de solos ou propriedades individuais de solos. O segundo é identificar e obter dados de entrada relevantes para se alcançar o objetivo proposto, incluindo pontos com informação de referência sobre classes ou propriedades de solos e variáveis que podem ser usadas como preditoras, tais como mapas existentes. O terceiro passo envolve o desenvolvimento de equações para prever propriedades dos solos ou de regras de classificação. O quarto passo consiste em aplicar as equações ou as regras de classificação previamente desenvolvidas e o quinto passo corresponde à avaliação do sucesso ou acurácia da predição das propriedades ou classes comparando-se o resultado com dados de referência coletados em pontos de observação. O último passo, é a geração dos mapas finais.

Atualmente, o Brasil tem 80% de seu território coberto pelo levantamento de solos produzido pelo Projeto RADAMBRASIL (1986) em escala 1:250.000 e publicado em 1:1.000.000 (FIGUEIREDO, 2008). Embora estejam disponíveis, esses dados têm escala muito pequena, não adequada para o gerenciamento de propriedades rurais ou regiões de bacias hidrográficas, e mapas de solos mais detalhados, em maiores escalas, são escassos. (COELHO e GIASSON, 2010).

O modelo para o mapeamento digital de solos é baseado no modelo *Clorpt* do método convencional e foi proposto por McBRATNEY et al (2003), eles propuseram a inclusão de dois fatores dos quais o solo pode ser inferido, assim o primeiro novo fator é o espacial (o solo pode ser predito a partir de informações da sua vizinhança) e o segundo

é o solo (o solo pode ser predito a partir de seus próprios atributos, ou os atributos podem ser preditos a partir das classes ou outras propriedades).

Dados derivados de modelos digitais de elevação (MDE), vêm sendo usados para compreensão das relações espaciais e temporais entre classes de solos e diferentes variáveis ambientais (SANCHEZ et al., 2009). Vários estudos mostram que o uso de atributos de terreno derivados do MDE é eficiente na predição de propriedades e classes do solo. (CARVALHO JÚNIOR et al., 2011).

2.2 Lógica Fuzzy

A base de conjuntos *fuzzy* foi introduzida por Zadeh (1965), como uma teoria matemática aplicada a conceitos difusos. A partir daí, a pesquisa e a aplicação dessa teoria em sistemas decorrente dos termos linguísticos convencionalmente vistas como não-científicas são admissíveis na abordagem *fuzzy*, aproximando muitas vezes a representação do sistema à própria realidade (ROVEDA et al., 2011).

A pertinência de um elemento num conjunto fuzzy não é uma questão de afirmação ou negação, mas uma questão de grau. Este grau representa um nível de compatibilidade do elemento sobre o conjunto, na qual o valor 0 (zero) significa não participação, o valor 1 (um) significa participação plena e demais valores entre 0 e 1 significam participação parcial. Os graus de pertinência são obtidos através das funções que são representadas por formas geométricas diversas. As principais formas utilizadas são: triangular, trapezoidal e gaussiana. (BARIN et al., 2010).

Segundo Silva (2013) o raciocínio *fuzzy* é formado por três etapas: a *fuzzificação*, a *inferência* e a *defuzzificação*. A *Fuzzificação* transforma as variáveis do problema em valores *fuzzy*, ou seja, valores numéricos em termos da linguagem natural. A *Inferência* é a etapa mais importante do raciocínio *fuzzy*, pois é mediante ela que é feita a tomada da decisão. Também é efetuado o raciocínio permitindo conclusões a partir de fatos conhecidos e das variáveis linguísticas de entrada e saída. E a *defuzzificação* transforma resultado *fuzzy* em resultado escalar, ou seja, o transforma em dado quantitativo (consiste em retornar valor numérico na faixa estipulada pela lógica *fuzzy*).

A lógica *fuzzy*, possibilita gerar e disponibilizar informação aos tomadores de decisões sem, contudo, determinar uma solução ótima ou a melhor de todas, trabalhando com a informação dos diversos atores ou intervenientes e permitindo verificações passo a passo do processo (MEDEIROS, 2007).

Assim, a modelagem *fuzzy* tem sido extensamente utilizada, por basear-se na caracterização de classes que não possuem, ou não podem definir, limites rígidos entre si (SILVA et al., 2009) sendo indicada para lidar com ambiguidades, abstrações e ambivalências em modelos matemáticos complexos que representam limites difusos comuns em processos naturais (BONISCH et al. 2004).

Segundo Moraes (2014), as características da lógica fuzzy que merecem destaque são: modelagem de problemas complexos, modelagem cognitiva, complexidade reduzida, modelagem de sistemas envolvendo múltiplos especialistas e manipulação de incertezas. Já as desvantagens conforme citado por Andrade et al. (2014) são a imprecisão e a dificuldade em prever qual será a interação entre as diferentes condicionantes do problema.

2.3 Uso da Lógica *fuzzy* na predição de classes de solo

O mapeamento digital de solos começou na década de 1970 e acelerou-se significativamente nos anos 1980, com o advento de novas tecnologias de coleta, processamento e análise de informações espaciais, como os sistemas de informações geográficas (SIG), os sistemas de navegação por satélite e o sensoriamento remoto. A partir do início da década de 1990, avanços sensíveis na capacidade de processamento dos computadores e em softwares de estatística e modelagem deram grande impulso aos estudos de mapeamento digital de solos (SARMENTO, 2010).

O levantamento de solos pode fornecer subsídios para o planejamento de diversas áreas de atividade agrícola ou não (COSTA et al., 2009), apresentando o mapa de solos como seu produto final. (SILVA et al., 2013).

Uma abordagem que tem sido utilizada para mapeamento de solos e predição de seus atributos e que tem apresentado bom desempenho é a lógica *fuzzy* (LIAO, 2010). Esta técnica tem a vantagem de utilizar uma amostragem de solos pequena incorporando a relação solo-paisagem e o conhecimento de especialistas na modelagem ao invés de usar simplesmente técnicas estatísticas. Este método baseia-se na premissa que o conhecimento do especialista em solos e o entendimento das relações solo-paisagem atuam como um modelo mental que pode predizer classes e propriedades do solo (ASHTEKAR e OWENS, 2013). Segundo Stolle et al. (2009) os valores de pertinência *fuzzy* dados de forma relativa e os operadores *fuzzy* permitem uma grande flexibilidade na modelagem.

A modelagem por lógica *fuzzy* permite simular, através do uso de possibilidades, as incertezas e transições tão comuns aos sistemas pedológicos, que é feita correlacionando-se o grau de certeza à experiência e ao conhecimento acumulados; assim, a atribuição dos pesos (que variam de 0 a 1) às possibilidades de ocorrência das classes de solo em determinada área, é totalmente controlada pelo julgamento subjetivo de especialista experiente e que detenha um conhecimento profundo sobre a área de estudo (CARVALHO et al., 2009).

Outros autores corroboram com Liao (2010), como Bonisch et al. (2004), que citam que a modelagem *fuzzy* tem sido amplamente utilizada para classificação de solos, pois é indicada para lidar com ambiguidades, abstrações e ambivalências de modelos matemáticos complexos que representam limites difusos comuns em processos naturais. E, de acordo com Sarmiento (2010), a incerteza, imprecisão e ambiguidade são inevitáveis e inerentes em sistemas naturais como o solo. Segundo Alvez Sobrinho et al. (2011) frente às dificuldades para precisar, por exemplo, o potencial de erosão dos solos através de modelos empíricos, a lógica *fuzzy* vem sendo utilizada no aperfeiçoamento desses modelos.

Segundo McBratney et al. (2003), a teoria de conjuntos difusos oferece uma rica base matemática para entender problemas de decisão e para construir regras de decisão na avaliação e combinação de critérios para o mapeamento digital de solos. Assim a lógica *fuzzy* permite o estabelecimento de relações matemáticas entre covariáveis ambientais e classes de solos (COELHO e GIASSON, 2010).

Zhu et al. (2009), mostra que os atributos do solo podem ser preditos com base na utilização da lógica *fuzzy* (ZHU et al., 1997), a qual requer *softwares* para que seja realizada as inferências, além da integração dos conhecimentos a respeito dos solos (SHI et al., 2009). ZHU et al., (2010), demonstraram a eficiência do uso desta técnica na predição de atributos dos solos.

Softwares que permitem a integração do conhecimento do pedólogo e esse tipo de lógica apresentam um grande potencial quanto a criação de mapas mais realísticos. E ainda essa técnica necessita de uma menor quantidade de pontos amostrados no campo (SHI et al., 2009).

Há basicamente duas abordagens de conjuntos difusos em mapeamento digital de solos (SARMENTO, 2010). A primeira procura particionar as observações feitas sobre um espaço multivariado em grupos naturais relativamente estáveis, ou seja, em grupos formados por similaridade de características. A segunda consiste em utilizar conjuntos difusos para classificação de solos, onde uma função de pertinência é definida a priori, com base no conhecimento especialista ou em definições convencionalmente impostas, para então alocar os dados conforme o grau com que eles se ajustam aos requerimentos das classes (QI e ZHU, 2003).

De acordo com Menezes et al. (2013), técnicas de mapeamento digital de solos baseada em sistemas sob uso de lógica *fuzzy*, permitem a produção mais rápida de mapas de solos, além de ajustar a distribuição espacial contínua das propriedades do solo em categorias discretas, de acordo com a complexidade inerente da variabilidade dos mesmos, aumentando a acurácia de informações espaciais. Portanto, sob essas condições, as relações solo-paisagem podem auxiliar na espacialização de informações para a confecção de mapas de solos (SILVA, 2013).

Existem vários trabalhos testando a lógica *fuzzy* em solos. Silva et al., 2009, fizeram uma ponderação por meio da modelagem *fuzzy* para estudar, com base em atributos químicos do solo, a propagação de incertezas da acidez de um Latossolo Vermelho-Amarelo húmico cultivado com café arábica. Benini (2012) utilizando a Teoria *Fuzzy*, teve por proposta avaliar a densidade do solo via sistema de inferência *fuzzy* sendo os dados de entrada: resistência à penetração no solo, teor de água, teor de argila, e a partir destes, obtiveram a saída densidade do solo para prever a compactação do solo.

Carvalho et al. (2009) apresentaram uma proposta metodológica para delineamento de unidades de mapeamento de solos na região de Mucugê (BA), utilizando dados ambientais, integrados via SIG sob inferência *fuzzy*, visando a confecção de mapa digital de solos, e verificaram ser possível fazer predição de unidades de solos a partir de dados auxiliares pré-existente e do conhecimentos de especialistas de solos, demonstrando ser uma técnica potencialmente capaz de operacionalizar os trabalhos em levantamentos de solo (NEUMANN, 2012). Roveda (2011) desenvolveu um estudo sobre a permeabilidade de solos de uma região impactada da Baixada Santista utilizando sistemas de inferência *fuzzy*.

Ippoliti R, et al. (2005) propuseram um procedimento simples para a extração de informação morfológica relevante para fins de mapeamento pedológico a partir dos dados digitais de elevação, baseado nas ferramentas de um Sistema de Informação Geográfica (SIG). Eles objetivaram a análise quantitativa da superfície de uma microbacia, visando à identificação e classificação das geoformas da paisagem e a sua posterior associação aos diferentes tipos de solos.

Zhu et al. (2008) também usaram um procedimento de classificação por lógica de conjuntos difusos para identificar combinações únicas de fatores ambientais associadas a classes típicas de solos. O objetivo foi avaliar formas de racionalizar a coleta de dados de campo para desenvolver modelos de solo-paisagem para mapeamento digital, de forma a conduzir futuras amostragens em locais de classes mais puras. Barreto-Neto e Souza Filho (2008) apresentaram um modelo baseado na lógica *fuzzy* para estimar o escoamento superficial em uma bacia hidrográfica tropical.

Lobão et. al. (2006) avaliaram a vulnerabilidade natural à erosão do solo no município de Morro do Chapéu-BA, a partir do modelamento das variáveis, geologia,

geomorfologia, solo, clima, vegetação e uso do solo. As variáveis foram integradas por operadores de lógica fuzzy. Os autores concluíram que modelamentos baseados em lógica *fuzzy* possibilitam a codificação do conhecimento, numa forma mais coerente ao modo como os especialistas pensam, aproximando-se assim, do modo cognitivo utilizado por eles na análise de problemas, tornando a apreensão da realidade muito mais confiável. Arruda (2013) desenvolveu um trabalho que teve por objetivos gerar um mapa digital de solos a partir de covariáveis ambientais que apresentem relação significativa com a distribuição dos solos.

Silva (2013), utilizou os princípios de uma modelagem *fuzzy* para elaboração de mapas temáticos da vulnerabilidade costeira baseada em variáveis disponíveis para estudos abrangendo um trecho da praia do Pina e as praias da Boa Viagem e Piedade/Recife-PE. Alvez Sobrinho, et al. (2011), desenvolveram uma rede neural artificial capaz de estimar, com precisão satisfatória, a erosividade da chuva em qualquer localidade do Estado de Mato Grosso do Sul, demonstrando assim a viabilidade do uso da lógica *fuzzy* em sistemas naturais.

3 MATERIAL E MÉTODOS

O trabalho foi conduzido na bacia hidrográfica do Ribeirão Inhaúma (Figura 6), com área de 2403,9 ha, coordenadas $21^{\circ}10'58.82''S$ e $41^{\circ}00'08.87''O$, localizada quase que totalmente no município de Iconha, no sul do estado do Espírito Santo.

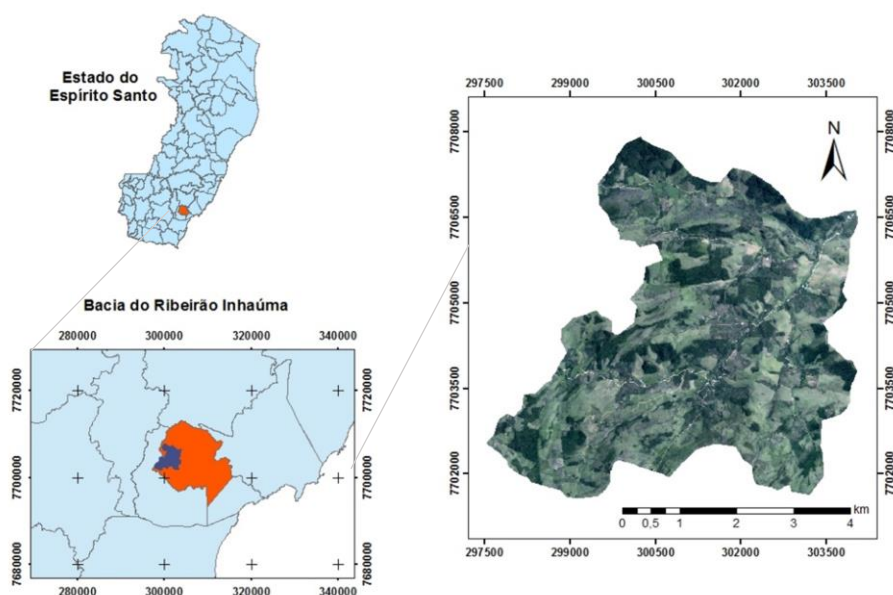


Figura 6- Área de estudo.

Primeiramente o modelo digital de elevação (MDE) foi obtido a partir de dados do satélite ALOS-Palsar com resolução de 12,5 m. A partir deste modelo foram gerados

os mapas de atributos do terreno, utilizando-se o software SagaGis. Os mapas de atributos gerados foram: *Aspect*, *Catchment área*, *Channel Network*, *Convergence Index*; *Convexity*; *Entropy*; *Flow accumulation*; *Flow Direction*; *General Curvature*; *Gradient*; *Hill Index*; *Landform*; *Longitudinal Curvature*; *LS fator*; *Mass Balance Index*; *Maximal Curvature*; *Maximum Curvature*; *Maximum Height*; *Maximum Membership*; *Melton Ruggedness Number*; *Mid-Slope Positon*; *Minimal Curvature*; *Multiresolution index of the ridge top flatness MRRTF*; *Multiresolution index of valley bottom flatness-MRVBF*; *Normalized Height*; *Plan Curvature*; *Profile Curvature*; *Protection Index*; *Slope Height*; *Slope Length*; *Slope*; *Standardized Height*; *Surface Area*; *Tangential Curvature*; *Terrain Ruggedness Index*; *Texture*; *Topographic Position Index*; *Topographic Wetness Index*; *Total Curvature*; *Valley Depth*; *Vector Terrain Ruggedness*; *Vertical Distance*.

Todas essas covariáveis foram testadas para se obter àquelas que tivessem maior capacidade de prever as classes, reduzindo assim o espaço de análise a quatro variáveis: *Channel Network*, *MRVBF*, *Topographic Wetness Index* e o próprio MDE (Figura 7). Essas foram definidas a partir de análise visual, onde observou-se a relação da variável com o local de ocorrência de uma ou mais classes de solo. Essas propriedades foram também utilizadas em outros trabalhos como Silva et al, 2016 e Vaysse e Lagacherie, 2015 e podem ser chamadas de atributos do terreno. Foi ainda utilizado como variável preditora o mapa de uso e ocupação do solo (Figura 8). Todos os mapas foram padronizados com o mesmo tamanho de células e número de linhas e colunas.

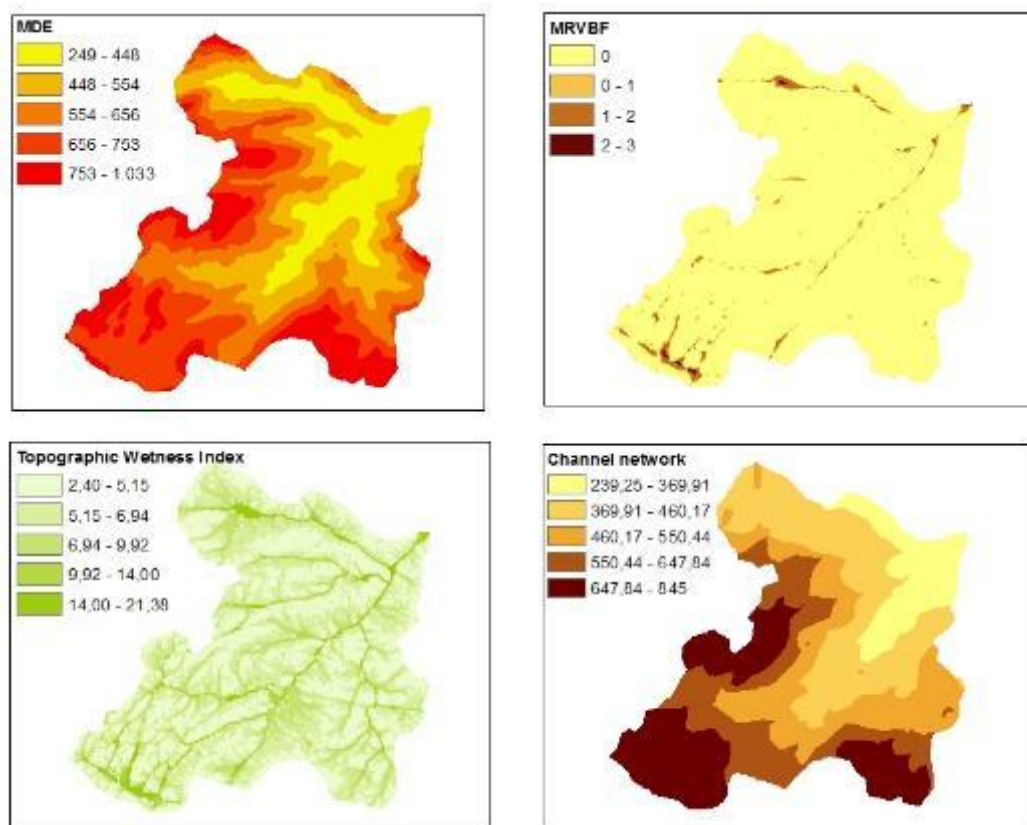


Figura 7- Covariáveis do terreno utilizadas para predição das classes de solo na bacia do Ribeirão Inhaúma-Iconha/ES.

Modelo digital de elevação-MDE; *Multiresolution index of valley bottom flatness-MRVBF*

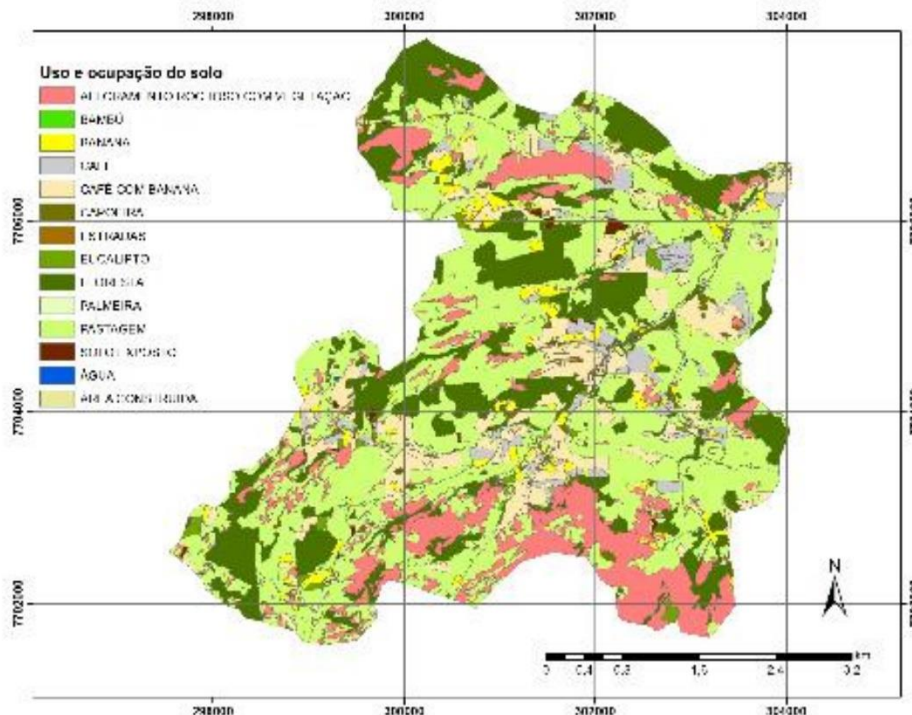


Figura 8- Mapa de uso e ocupação do solo.

Para espacialização foi utilizado a lógica *fuzzy*. Formalmente, um conjunto *fuzzy* A pode ser expresso por: (FUJIMOTO, 2005):

$$A = \{x, \mu_A(x) / x \in X\} \quad (1)$$

Onde X é o universo de discurso ao qual o conjunto pertence, x um componente do conjunto *fuzzy*, e μ_A a função de pertinência. O universo de discurso indica o domínio em que o modelamento do sistema *fuzzy* é válido.

A função de pertinência μ_A indica o grau de pertinência (ou compatibilidade) entre x e o conceito expresso por A (SANDRI E CORREA, 1999):

- $\mu_A(x) = 1$ indica que x é completamente compatível com A;
- $\mu_A(x) = 0$ indica que x é completamente incompatível com A; e
- $0 < \mu_A(x) < 1$ indica que (x) é parcialmente compatível com A, com grau $\mu_A(x)$.

A determinação das classes de solos utilizando a técnica da *lógica fuzzy*, foi feita de acordo com a metodologia proposta por Tanscheit (2014) (Figura 9).

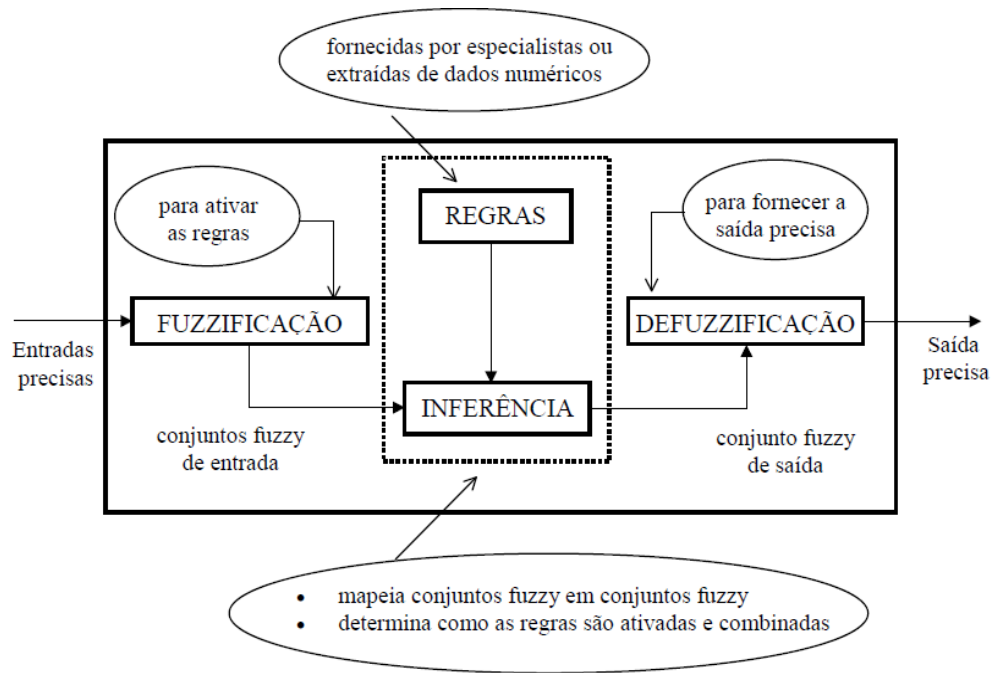


Figura 9- Fluxograma para predição de classes de solo utilizando a Lógica Fuzzy.

Fonte: TANSCHKEIT, 2014.

Para proceder à modelagem sob inferência *fuzzy*, são atribuídos pesos variando de 0 a 1, para cada variável preditora, estes valores constituem o peso que a variável tem para cada classe de solo e são determinados de forma subjetiva, baseados no conhecimento técnico do especialista que irá atribuir os pesos. Ou seja, os pesos caracterizam o grau de importância de cada variável no controle da definição do tipo de solo.

Posteriormente, são gerados os mapas de classes de solo a partir de diferentes proposições de operadores *fuzzy*. Foi utilizado o operador lógico intersecção (*Fuzzy AND*), onde a intersecção significa uma sequência de “E” e é obtida através do operador MIN.

$$\mu = \text{MIN} (\mu_a, \mu_b, \mu_c, \dots) \quad (2)$$

em que, μ_a , μ_b , e μ_c são valores de pertinência dos mapas A, B e C, para uma dada localização.

Este operador produz estimativas conservadoras implicando, pedologicamente, na modelagem de associações de fatores diagnósticos, ou seja, na seleção mais rigorosa de áreas mais propícias à ocorrência de determinada classe de solo (CARVALHO, 2009). Por exemplo, se a possibilidade de ocorrer Latossolo em um local tiver os valores $\mu_a = 0,50$, $\mu_b = 0,90$ e $\mu_c = 0,30$ (respectivamente para os fatores a, b e c), o mapa resultante assumirá o valor $\mu_{\text{comb}} = 0,30$ neste mesmo local, deduzindo-se que somente onde os três mapas combinados possuam simultaneamente valores elevados se consideraria locais mais prováveis para a ocorrência de Latossolos (CARVALHO, 2009).

Nesse trabalho, para o mapeamento utilizando a lógica *fuzzy*, foi utilizado o ArcSIE (SHI et al., 2009), que consiste em uma extensão ArcGIS que utiliza lógica fuzzy para identificar os locais dentro da área de interesse que corresponde a condição ambiental de cada classe de solo (SILVA et al. 2016).

Inicialmente se utilizou a extensão *case-based reasoning* (CBR), cuja a inferência das classes de solo é feita com base em pontos, linhas e polígonos. Nesse caso, utilizou-se 200 pontos, onde as classes de solo eram conhecidas. Estes pontos foram determinados em campo, com coleta das coordenadas geográficas por meio do uso de receptores de sinais GNSS (GS08 Plus). Foram separados 150 pontos para o treinamento e 50 pontos para a validação do método. Estes mapas serviram de base para a construção das curvas de pertinência *fuzzy* pelo método *rule-based reasoning* (RBR), onde utilizou-se os 200 pontos para validação.

Dentro da extensão RBR, as inferências são baseadas em regras, ou seja, consiste numa função de pertinência *fuzzy* que define a relação entre os valores de uma covariável ambiental (*environment feature*) e valores ótimos (*optimality values*) para um dado tipo de solo. Após se estabelecer as regras para cada tipo de solo (Neossolo, Cambissolo, Latossolo, Gleissoloe afloramento rochoso), foi gerado um mapa para cada um deles; posteriormente foi utilizada a ferramenta *Harden map* onde foi gerado o mapa de polígonos contendo todas as classes de solo da área. Esse mapa é obtido a partir dos chamados vetores de similaridade onde cada pixel é classificado com o valor mais elevado de pertinência *fuzzy* no vetor de similaridade.

Para que um mapa digital de solos produzido tenha confiabilidade, sua validação em campo é necessária, identificando-se e coletando-se o solo de forma a representar as unidades de mapeamento delimitadas. Dessa forma, a validação do mapa digital de solos foi feita com base em pontos de observação determinados em toda a área, totalizando 200 pontos. Foram ainda utilizados cortes de estradas para coleta de amostras de solo e descrição do perfil do solo. Ao total coletou-se 12 perfis e 3 amostras extras utilizadas para confecção do mapa convencional de solos.

Tanto nos pontos de amostragem quanto nos pontos de observação os solos foram caracterizados morfologicamente, de acordo com Santos et al. (2015). Para os pontos de observação, foi feita uma amostragem aleatória na área, de modo a se obter a maior representatividade da mesma (Tabela 1).

Tabela 1- Distribuição dos perfis, amostras extras e pontos de observação por classes de solos na área de estudo

Classe de solo	Número de amostras	Representatividade (%)
RL	28	14,00
CX	100	50,00
GX	24	12,00
AR	33	16,50
LVA	15	7,50
Total	200	100

RL- Neossolo Litólico, CX- Cambissolo Háplico, GX-GleissoloHáplico, AR- Afloramento rochoso, LVA-Latossolo Vermelho-Amarelo

A localização dos perfis no campo foi feita pelo método de amostragem do Hipercubo Latino Condicionado, sendo considerado as possibilidades de acesso devido às dificuldades de deslocamento na área em função do relevo acentuado, nesse método a seleção dos valores da amostra é altamente controlada e através disto são necessárias menos amostras para obter-se uma distribuição mais representativa (SALIBY, 1997), além de ser uma técnica que proporciona uma redução de variância.

A acurácia do mapa de solo produzido foi medida pelo índice Kappa e pela exatidão global, determinadas a partir da matriz de confusão. A exatidão global é definida como a estimativa da porcentagem de área mapeada que foi corretamente classificada, quando comparada com os dados de referência ou verdade de campo. Foi obtida pela divisão da soma da diagonal principal (número de classificações corretas ou concordância real) pelo número total de amostras tomadas (SUAREZ e CANDEIAS, 2012).

Na matriz de confusão os dados de referência são representados pelas colunas da matriz e foram comparados com os dados classificados, representados pelas linhas da matriz. Os elementos da diagonal principal indicam o nível de concordância entre os dois conjuntos de dados.

Para a determinação de categorias individuais de exatidão o número total de pixels correto de uma categoria é dividido pelo número total de pixels desta categoria, proveniente dos dados de referência (ou seja, do total da coluna), esta medida de acurácia indica a probabilidade de um pixel de referência ser corretamente classificado e é realmente uma medida de erro de omissão (SUAREZ e CANDEIAS, 2012), que representa a proporção de uma unidade de mapeamento ter sido classificada corretamente.

Também foi determinada a exatidão do usuário que indica a probabilidade de um ponto no mapa representar a verdade de campo, é obtido com a divisão do número de registros classificados corretamente de uma determinada categoria pelo número total de registro nos dados de classificação (total da linha para a determinada categoria).

O coeficiente Kappa é uma medida da concordância real (indicada pelos elementos da diagonal da matriz de confusão) menos a possibilidade de concordância (indicada pelo produto das linhas e colunas totais, que não incluem entradas não reconhecidas) (CONGALTON e GREEN, 1999). O índice Kappa geralmente varia de 0 a 1, sendo que 0 indica a ausência de concordância ou seja, a estimativa e a referência são estatisticamente independentes (ELNAGGAR, 2007), e 1 indica total concordância. Valores de Kappa de 0,75 ou maiores mostram um desempenho do classificador de muito bom a excelente, enquanto valores entre 0,4 e 0,75 representam desempenho moderado a bom, e valores menores que do que 0,4, uma baixa concordância (baixo desempenho) (LANDIS e KOCH (1977); MONSERUD e LEEMANS (1992)).

O Coeficiente Kappa foi obtido segundo a equação 3:

$$K = \frac{\sum_{i=1}^c \chi_{ii} - \sum_{i=1}^c (\chi_{i+} * \chi_{+i} + 1)}{n^2 - \sum_{i=1}^c (\chi_{i+} * \chi_{+i} + 1)} \quad (3)$$

em que, K é uma estimativa do coeficiente Kappa; χ_{ii} é o valor na linha i e coluna i; linhas totais χ_{i+} é a soma de linhas i e colunas totais; χ_{+i} é a soma das colunas i da matriz de confusão; n é o número total de amostras e c o número total de classes.

Nesse trabalho foi utilizado para interpretação do índice Kappa a tabela proposta por Landis e Kochi (1977) (Tabela 2).

Tabela 2- Intervalos de Kappa com respectiva interpretação de desempenho

Kappa	Desempenho
< 0,00	Péssimo
0,00-0,20	Ruim
0,21-0,40	Regular
0,41-0,60	Bom
0,61-0,80	Muito bom
0,81-1,00	Excelente

Fonte: Adaptado de Landis & Kochi (1977).

Foi gerado mapa na escala 1:50 000, cuja área mínima mapeável (AMM) foi igual a 10 ha, calculada de acordo com a equação:

$$AMM = \frac{E^2 \times 0,4}{10^8} \quad (4)$$

Onde E é a escala, 0,4 é a área mínima mapeável especificada no Brasil em cm² (IBGE, 2007) e AMM é a área mínima mapeável em ha.

Os mapas digitais de solo foram ainda comparados ao mapa convencional, sendo este elaborado com base nos perfis amostrados e no modelo digital de elevação da área. Para comparação foi utilizado o teste Z.

O teste Z compara coeficientes Kappa de dois classificadores e é dado pela equação (CONGALTON e MEAD ,1983):

$$Z = \frac{\hat{k}_1 - \hat{k}_2}{\sqrt{\hat{\sigma}_1 + \hat{\sigma}_2}} \quad (5)$$

Onde: Z = valor Z calculado;

\hat{k}_1 = coeficiente Kappa do classificador 1;

\hat{k}_2 = coeficiente Kappa do classificador 2;

$\hat{\sigma}_1$ = variância de Kappa do classificador 1;

$\hat{\sigma}_2$ = variância de Kappa do classificador 2

A variância de Kappa pode ser estimada por:

$$\sigma(\hat{k}) = \frac{1}{n} \left[\frac{\theta_1 (1 - \theta_1)}{(1 - \theta_2)^2} + \frac{2(1 - \theta_1)(2\theta_1\theta_2 - \theta_3)}{(1 - \theta_2)^3} + \frac{(1 - \theta_1)^2(\theta_4 - 4\theta_2^2)}{(1 - \theta_2)^4} \right] \quad (6)$$

Em que:

$$\theta_1 = \frac{1}{n} \sum_{i=1}^c x_{ii}$$

$$\theta_2 = \frac{1}{n^2} \sum_{i=1}^c x_{1+} x_{+1}$$

$$\theta_3 = \frac{1}{n^2} \sum_{i=1}^c x_{ii}(x_{i+} + x_{+i})$$

$$\theta_4 = \frac{1}{n^3} \sum_{i=1}^c \sum_{j=1}^c x_{ij}(x_{j+} + x_{+i})^2$$

4 RESULTADOS E DISCUSSÃO

A estatística descritiva dos atributos utilizados é mostrada na Tabela 3, enquanto os gráficos são mostrados na Figura 10. Os gráficos mostram boa distribuição dos dados, onde estes não apresentam nenhum valor de *outlier*, e possuem valor de mediana no centro, determinando uma distribuição simétrica. A covariável MRVBF apresentou a maior amplitude interquartilica, e menores valores máximos e mínimos.

Tabela 3- - Estatística descritiva das covariáveis utilizadas para predição das classes de solo

	MDE	TWI	Channel_net	MRVBF
Min	249	2,40	239	0
Max	1033	21,38	845	3
Mean	599,51	6,05	523,27	0,06
Desvio padrão	138,27	2,44	128,95	0,31

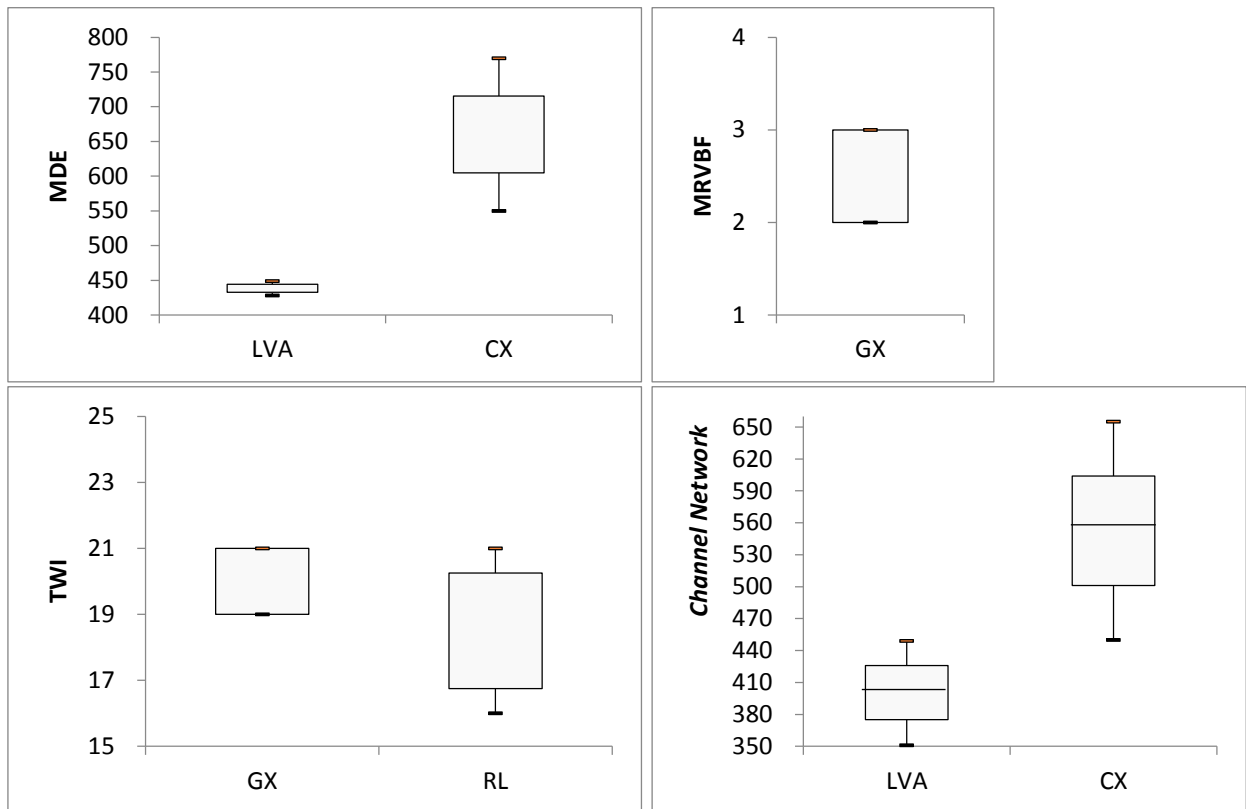


Figura 10- Boxplots referentes aos atributos do terreno em cada classe em que foram utilizados para predição.

CX- Cambissolo Háplico; GX- Gleissolo Háplico; LVA- Latossolo Vermelho-Amarelo; RL- Neossolo Litólico; Topographic Wetness Index-TWI; Modelo digital de elevação-MDE

Para mapeamento das áreas de Gleissolofoi utilizado a covariável *Topographic Wetness Index- TWI* e *Multiresolution index of valley bottom flatness-MRVBF*; para o Neossolo foi utilizado o TWI e mapa de uso e ocupação do solo; para o Latossolo e Cambissolos foram utilizados o *Channel Network (Channel_Net)* e o MDE; e para Afloramento rochoso o mapa de uso e ocupação do solo (Tabela 4). Essas variáveis mostraram-se importantes pois a partir de diferentes posições do relevo e segmentos das vertentes formam-se distintos tipos de solos (MOORE et al., 1993). Assim, a função de associação *fuzzy* descreve como a semelhança entre um solo local e o caso típico desse tipo de solo mudará à medida que as condições ambientais mudam (ZHU et al., 2010).

Previamente a escolha dessas covariáveis, foram testadas todas as demais (42), chegando-se a conclusão que as mais satisfatórias para utilização seriam: *Vertical distance*; MDE; MRVBF; TWI; Channel_Net e *Slope*. Então, estes passaram por outra seleção, chegando-se àquelas variáveis efetivamente usadas no mapeamento final, por promoverem melhor classificação dos solos na região. Silva et al., 2016 usaram para mapear Latossolo vermelho e Cambissolo as covariáveis: *Vertical distance e TWI*; e para Gleissolo: *Vertical distance, slope e TWI*. Xing-Zhu et al. (2010) utilização as covariáveis *slope gradient, contour curvature, profile curvature e topographic wetness index* para mapear solos em sua área de estudo.

O mapa de uso e ocupação foi imprescindível para classificação dos afloramentos rochosos. Quanto ao uso deste para classe dos Neossolos Litólicos, percebe-se que essa

foi uma peculiaridade encontrada na área, que facilitou o mapeamento. Ao fazer o reconhecimento de campo, observou-se que áreas ocupadas com fragmentos florestais, estavam sobre Neossolos Litólicos. Dessa forma, o mapa de uso se tornou importante para a determinação também dessa classe. Para que este critério possa ser usado em outras regiões é necessário a análise prévia do local, para definir se há casos particulares, como esse.

A Tabela 4 apresenta os valores e os tipos de curva utilizado na predição dos mapas utilizando-se as covariáveis escolhidas. Utilizando esse método no ArcSIE foi necessário inserir o valor típico de cada atributo do terreno para cada classe de solo; valor esse conhecido como V1 e V2, e os desvios deste valor (W1 e W2).

W1 e W2 são calculados multiplicando o desvio padrão dos valores de cada atributo do terreno por 0,2 e representa o valor a ser subtraído (W1) ou acrescentado (W2) ao valor típico (V1 e V2 respectivamente), e representam 50% do grau de semelhança em relação às condições típicas daquela classe de solo ocorrer, enquanto que V1 representa a 100 % de grau de semelhança, ou seja, a condição típica para que classe de solo ocorra (SHI, 2013) .

Dessa forma, quanto maior a semelhança , maior a chance de determinado local conter àquela classe de solo. Os shapes da curva podem ser bell, S ou Z, sendo que a curva bell é usada quando àquela classe ocorre numa faixa, S quando ocorre em um valor superior ao pré-estabelecido e Z a um valor menor ao determinado.

Xing-Zhu et al. (2010) citam um exemplo de escolha da curva em seu trabalho de mapeamento de solos na China, em seu estudo o valor típico do gradiente de inclinação para os Cambissolos está localizado na extremidade mais à esquerda do eixo do gradiente da inclinação, então a função de associação difusa em relação a essa variável ambiental (gradiente de inclinação) é considerada como em forma de Z.

Tabela 4-Valores e tipos de curva utilizados para os atributos do terreno para determinação de cada classe de solo

Classe de solo	Atributo do terreno	Valores do gráfico				Shape da curva
		V1	W1	V2	W2	
Cambissolo	MDE	550	28	770	28	Bell
	Channel_Net	450	28	655	28	Bell
Gleissolo	TWI	19	2	-	-	S
	MRVBF	2,39	0,09	3,98	0,09	Bell
Latossolo	Channel_Net	351	28	449	28	Bell
	MDE	428	28	449	28	Bell
Neossolo	TWI	16	0.2	-	-	S
	Uso e ocupação*	1	0.1	1	0.1	Bell
Afloramento rochoso	Uso e ocupação*	4	0.1	4	0.1	Bell

*Mapa de uso e ocupação do solo, classe 1 corresponde a florestas e 4 a afloramentos rochosos. Modelo digital de elevação-MDE; *Channel Network* - Chan_net; *Topographic Wetness Index*-TWI; *Multiresolution index of valley bottom flatness*-MRVBF.

Após realizar-se as inferências para as classes de solo, gerou-se um mapa de pertinência para cada uma delas (Figura 11). A partir desses é possível perceber que a

classe de maior probabilidade de ocorrência na área é a dos Cambissolos, seguida pelos Neossolo, afloramento rochoso, Latossolos e, em menores áreas, os Gleissolo.

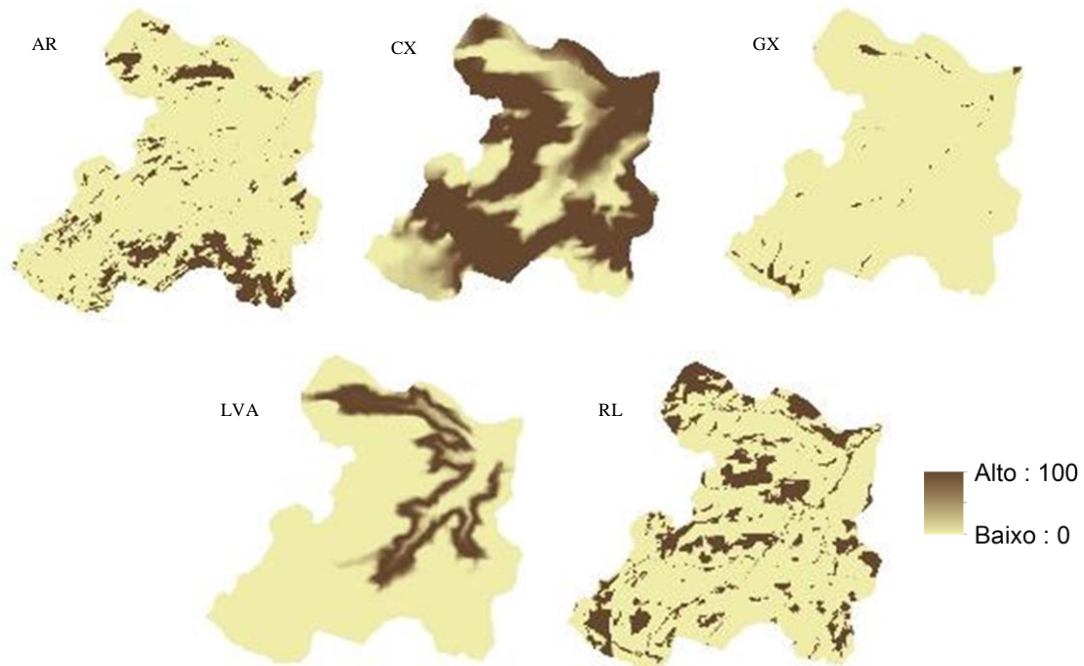


Figura 11- Mapas de probabilidade de ocorrência de cada classe de solo obtidos a partir da inferência *fuzzy*.

AR- Afloramento rochoso; CX- Cambissolo Háplico; GX- Gleissolo Háplico; LVA- Latossolo Vermelho-Amarelo; RL- Neossolo Litólico

O mapa final é apresentado na Figura 12. A maior parte da área foi ocupada por Cambissolo, enquanto a menor por Gleissolo, que ficou concentrado em áreas de várzea, nas partes mais baixas da paisagem, por exemplo na parte sudoeste da bacia, que apresenta relevo em torno bastante acidentado, permitindo assim acúmulo de água nessa área mais baixa.

O Neossolo e os Afloramentos rochosos foram encontrados nas partes mais altas da bacia, próximos aos domínios dos Cambissolos. Já os Latossolos foram mais comuns nos terços médios e inferiores das encostas.

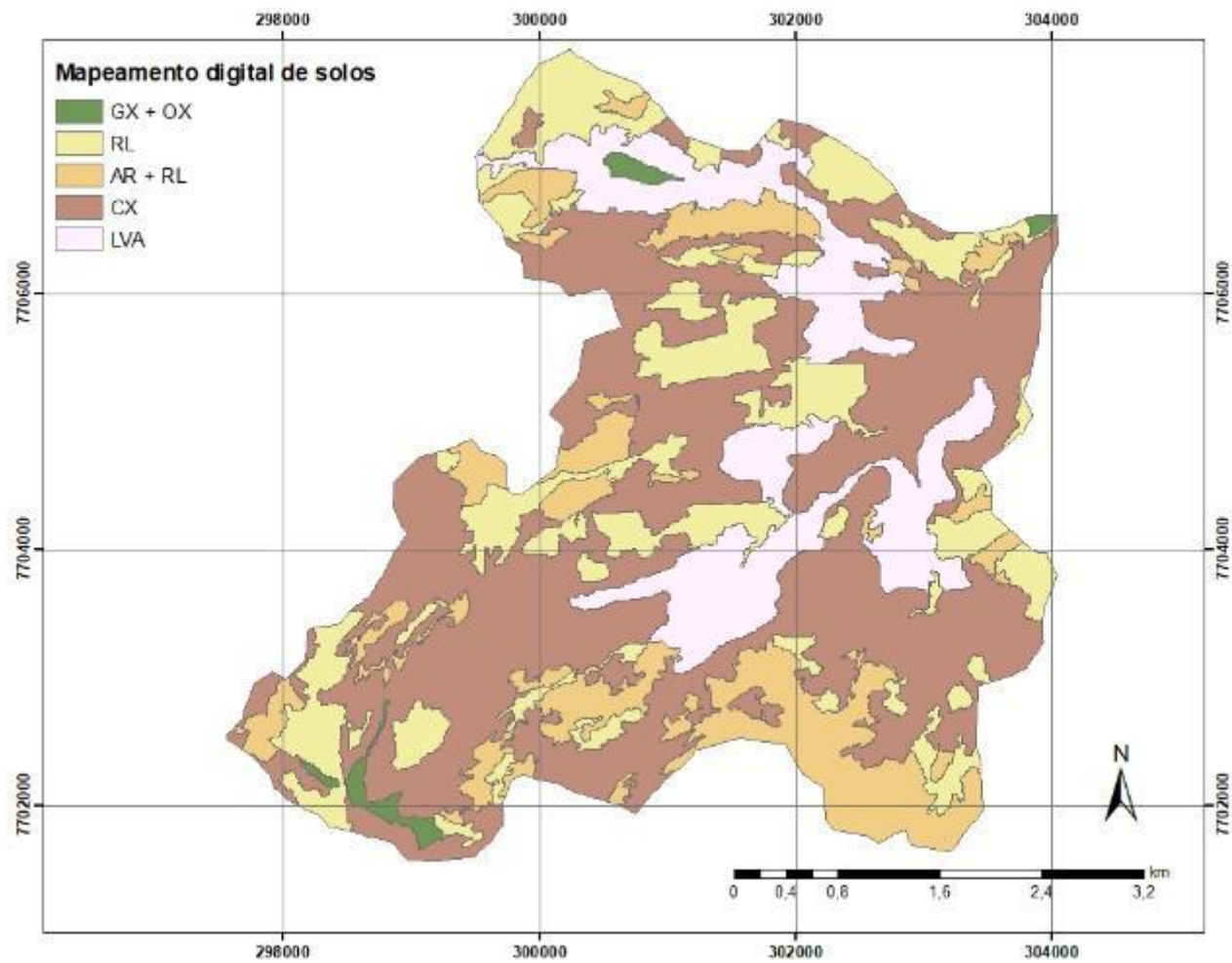


Figura 12- Mapa final de classes de solo obtido por inferência *fuzzy*.

AR- Afloramento rochoso; CX- Cambissolo Háptico; GX- Gleissolo Háptico; LVA- Latossolo Vermelho-Amarelo; RL- Neossolo Litólico

A Tabela 5 mostra a matriz de confusão obtida para o mapeamento digital, utilizando lógica *fuzzy* para inferência das classes de solo. O coeficiente Kappa obtido foi de 0,75, sendo classificado como muito bom (LANDIS E KOCHI, 1977), com acurácia global de 84%. Akumu et al. (2016), mapeando classes de profundidade do solo no Canadá alcançaram acurácia global de 94%. Bui et al. (2017) mapeando classes de solo no Vietnã, utilizando a covariável relevo para tal, conseguiram acurácia global de 72%. Enquanto Akumu et al. (2015) mapeando classes de textura do solo encontraram acurácia global de 79%. Silva et al. (2016) mapeando classes de solo obtiveram 80% de acurácia global e um índice Kappa de 0,65.

Esses valores de precisão mostram-se aceitáveis segundo o encontrado em outros trabalhos como de Zhu et al., 2010 e Schuler et al., 2010.

A maior acurácia do usuário (AU) foi alcançada pelo Afloramento rochoso com 100%, seguido por Neossolo Litólico (96%) e Gleissolo (91%). Os menores valores foram obtidos para Latossolo Vermelho-Amarelo (61%), enquanto Cambissolos obtiveram acurácia de 81%. Segundo Congalton e Rekas (1985) a AU mede a probabilidade de um ponto no mapa representar a verdade de campo.

Quanto a acurácia do produtor (AP) os melhores valores foram encontrados para Cambissolos com 95%, seguido por Neossolo Litólico (89%), Afloramento rochoso (82%), Latossolo Vermelho-Amarelo (73%) e com menor percentual Gleissolo com 42%. Gleissolos foram confundidos principalmente com Cambissolos, e isso se deve, provavelmente, a maior área ocupada por esse. A confusão das demais classes também ocorreu principalmente com CX, pelo mesmo motivo, já que esses estão distribuídos majoritariamente em toda bacia. Dias et al. (2015) encontrou maiores percentagens de acurácias do produtor para a classe de Neossolos litólicos, chegando a 100%.

Segundo Bui et al. (2017), a lógica *fuzzy* tem sido amplamente estudada e demonstrado eficiência para o mapeamento de áreas com condições de relevo suaves. Porém estes estudaram uma região do Vietnã, com condições de relevo acentuado e obtiveram boas precisões, assim como neste trabalho.

Tabela 5- Matrix de confusão para o mapeamento de solos utilizando lógica *fuzzy*

Classe de solo	RL	AR	CX	GX	LVA	Σ Linhas	AU (%)
RL	25	1	0	0	0	26	96
AR	0	27	0	0	0	27	100
CX	3	5	95	11	4	118	81
GX	0	0	1	10	0	11	91
LVA	0	0	4	3	11	18	61
Σ Colunas	28	33	100	24	15	200	
AP (%)	89	82	95	42	73		

AR- Afloramento rochoso; CX- Cambissolo Háplico; GX- Gleissolo Háplico; LVA- Latossolo vermelho-amarelo; RL- Neossolo Litólico; AU- Acurácia do usuário; AP- Acurácia do produtor

Considerando que no Brasil, a maior parte da informação sobre o solo encontra-se em mapas convencionais, realizou a comparação entre os mapas obtidos pelo mapeamento digital e convencional.

Nota-se que o mapa confeccionado aplicando o método convencional (Figura 13), utilizou-se intensamente do recurso de associações, não sendo possível a identificação da ocorrência e localização de uma única classe de solo dentro da unidade de mapeamento, deixando assim implícita a heterogeneidade desse tipo de técnica, conforme observado por Dias (2015). Ao contrário, o mapa digital apresenta as unidades simples, com classes de solo individuais, o que permite a obtenção de um mapa mais detalhado (TESKE, 2014).

Segundo Seyfried (1998), a variação espacial do solo retratada em mapas de solos convencionais é muitas vezes altamente generalizada e muitas vezes discreta. Outra questão relevante é apontada por Hudson (1992), que revela que o levantamento convencional de solos, é exemplo de uma ciência baseada nas relações solo-relevo, que por sua vez são baseadas no conhecimento tácito, obtido através da experiência do pedólogo, o que leva este conhecimento a ser de difícil compreensão, uma vez que este não é explícito, dificultando seu aprendizado e sua disseminação. A literatura mostra diversos estudos desde a década de 70, que apontam o caráter subjetivo desse método (Beckett & Webster, 1971; McBratney et al., 1981; Burrough 1989), levando este a ser uma técnica que apresenta grande incerteza (Mendonça-Santos et al., 2003).

Segundo Hodza (2010) para uma determinada área, pedólogos inevitavelmente produzirão diferentes mapas de solos, independentemente de usar ferramentas e técnicas similares ou diferentes, sendo a extensão e a gravidade das diferenças entre estes mapas de solo da área amplamente desconhecidas. Hodza (2010) usou uma abordagem de lógica difusa para quantificar a magnitude e a seriedade das diferenças entre dois mapas de solos interpretativos, sendo que um mapa foi criado por um cientista individual do solo usando métodos convencionais de mapeamento do solo e o outro foi produzido por quatro cientistas usando Sistema de Informação Geográfica, chegando ao resultado de dois mapas substancialmente diferentes.

De acordo com Heuvelink e Webster (2001), independentemente de como um ou outro prevê, um cientista do solo assume algum tipo de modelo de como o solo se comporta no espaço e no tempo, podendo haver grandes diferenças entre esses modelos, mas o que todos eles têm em comum é que envolvem algum grau de incerteza, sendo por isso, os modelos estatísticos melhores opções de mapeamento, já que os mesmos reconhecem a incerteza associada.

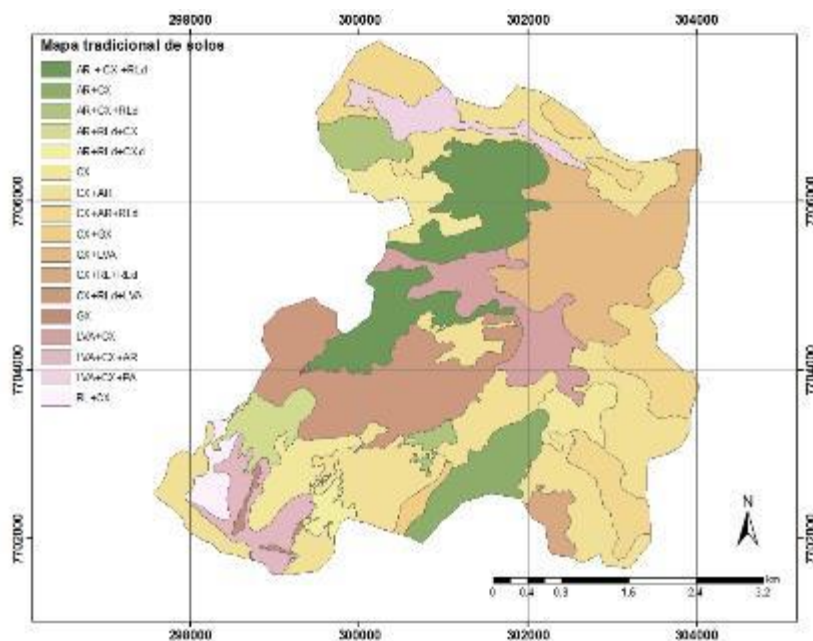


Figura 13- Mapa final de classes de solo obtido pelo método convencional de mapeamento.

AR- Afloramento rochoso; CX- Cambissolo Háptico; GX- Gleissolo Háptico; LVA- Latossolo Vermelho-Amarelo; RLd- Neossolo Litólico distrófico; PA- Argissolo Amarelo.

Poderando-se que o mapa convencional foi produzido a partir dos dados dos perfis de solo e observações no campo, obteve-se a matriz de confusão a partir dos 200 pontos de observação utilizados na validação do mapa digital de forma a garantir a não tendenciosidade, uma vez que se espera ao avaliar um mapa, independentemente da forma pelo qual foi produzido, que se encontre informações condizentes com a realidade. Para elaboração da mesma, foi considerado concordância entre inferência e realidade, quando a classe predominante foi classificada.

Sendo assim, a matriz de confusão é apresentada na Tabela 6. O índice Kappa encontrado foi de 0,35, e a acurácia global de 58%. A acurácia do usuário (AU) mostra valores elevados para RL (100%) e GX (90%), enquanto a menor acurácia se deu para CX com 62%. O que indica que devido a grande distribuição dessa classe na área, as demais classes confundiram-se, tendo todas apresentado amostras erroneamente classificadas como Cambissolos.

Quanto a acurácia do produtor (AP), assim como o encontrado no mapeamento digital, Cambissolos foram a classe de maior acurácia (77%). Porém divergentemente, Neossolo Litólico foi a classe de menor precisão, com apenas 11%, isso provavelmente se deve a não ter sido possível inferir no mapeamento convencional a presença de neossolos em áreas de mata, já que não existe uma relação direta. Porém, em uma área prioritariamente agricultável, acredita-se que quando se deu o desbravamento da região com provável desmatamento de florestas nativas, foram preservadas essas áreas hoje encontradas, devido ao fato de provavelmente essas terem sido consideradas impróprias para agropecuária.

Tabela 6- Matriz de confusão para o mapeamento convencional de solos

Classe de solo						Σ	AU
	RL	AR	CX	GX	LVA	Linhas	(%)
RL	3	0	0	0	0	3	100
AR	6	18	3	0	0	27	67
CX	19	14	77	8	6	124	62
GX	0	0	1	9	0	10	90
LVA	0	1	19	7	9	36	25
Σ Colunas	28	33	100	24	15	200	
AP (%)	11	55	77	38	60		

AR- Afloramento rochoso; CX- Cambissolo Háplico; GX- GleissoloHáplico; LVA- Latossolo Vermelho-Amarelo; RL- Neossolo Litólico.

O teste Z foi igual a 6,53 demonstrando a diferença estatística entre os dois métodos, onde o mapeamento digital mostrou-se superior. Valores acima de 1,96 mostram significância do teste a 5 %.

A Tabela 7 mostra a comparação das áreas delimitadas para cada classe no mapeamento digital e convencional. As áreas do mapeamento convencional foram determinadas considerando-se àquela de principal ocorrência na área mapeada em cada unidade.

As maiores concordâncias foram encontradas para as áreas de afloramento rochoso, devido a relativa facilidade de se mapear essa classe, e para Latossolo vermelho-amarelo. Quanto as divergências, elas foram mais significativas para as classes de Gleissoloe Neossolo Litólico.

A precisão da área de Gleissolos, torna-se mais difícil pelo fato dessa classe estar presente na menor parte da área de estudo. Quanto aos RL acredita-se haver uma subestimação no mapeamento convencional, considerando o fato da área apresentar relevo extremamente acidentado, com grandes áreas de afloramento rochoso, o que sugere a presença dessa classe em maiores áreas do que aquelas mapeadas pelo método, mesmo essa tendo tido outras classes a ela associada na classificação.

Tabela 7- Áreas definidas para cada classe pelo mapeamento convencional e digital usando lógica *fuzzy*

Método de classificação	Classe de solo				
	RL	CX	GX	AR	LVA
	(ha)				
Lógica fuzzy	486,6	1174,2	26,1	367,2	349,5
Convencional	32,4	1577,8	8,9	441,1	329,6

AR- Afloramento rochoso; CX- Cambissolo Háplico; GX- GleissoloHáplico; LVA- Latossolo Vermelho-Amarelo; RL- Neossolo Litólico.

A Figura 14 mostra o mapa de concordância entre os mapas convencional e digital. A concordância foi de 26%, sendo encontrada principalmente em áreas de Afloramento Rochoso, Cambissolo e em menor proporção para Latossolos, corroborando assim com os dados apresentados na Tabela 7. O mapa evidencia as áreas em que deverão ser concentrados os esforços para melhoria da acurácia dos mapas, mostrando que as técnicas devem ser complementares.

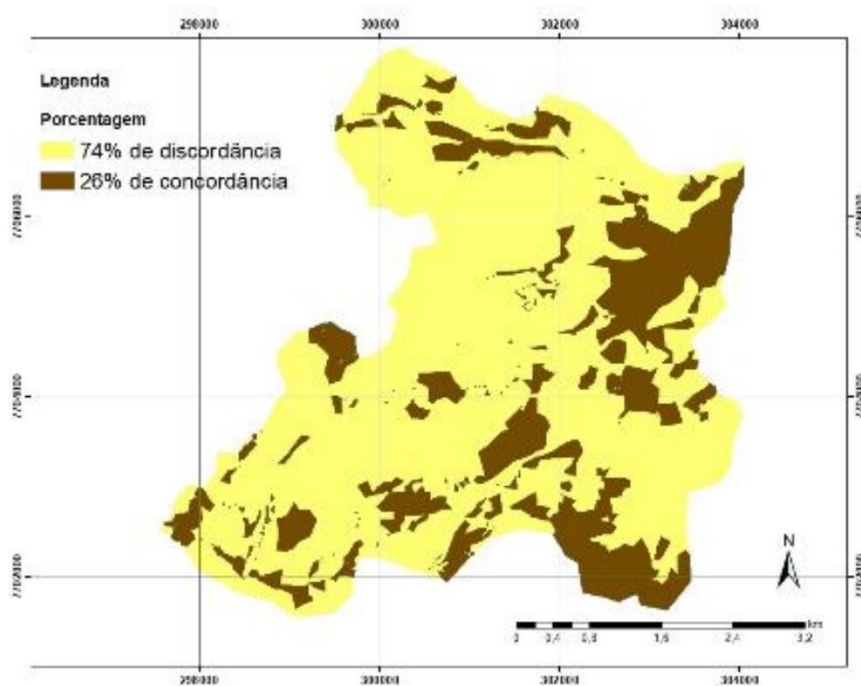


Figura 14- Mapa de concordância entre o método digital (Lógica *fuzzy*) e o mapa convencional na bacia do ribeirão Inhaúma, Iconha/ES.

5 CONCLUSÕES

1. É possível o mapeamento digital da região estudada utilizando-se lógica *fuzzy* com boa precisão;
2. Há divergências entre o mapeamento digital e convencional, sendo o índice de concordância de 26% entre si para a área, e por isso é necessário o uso de critérios diferenciados para a aplicação dos mesmos na área de estudo;
3. É necessário realizar a validação criteriosa nas áreas de maior discordância nas classificações para melhorar a acurácia dos mapas;
4. Lógica *fuzzy* é um método eficiente de mapeamento digital de solos em regiões com condições similares e mesmo padrão ambiental da área de estudo;

5. A integração entre o conhecimento tácito dos pedólogos sobre as relações solo-paisagem e a otimização dos procedimentos via mapeamento digital de classes de solos, poderiam levar aos melhores resultados para mapeamento de uma área.

6 REFERÊNCIAS BIBLIOGRÁFICAS

AKUMU, C. E. ; WOODS M.; JOHNSON J.A. ; PITT D.G.; UHLIG P. ; McMURRAY S. GIS-fuzzy logic technique in modeling soil depth classes: Using parts of the Clay Belt and Hornepayne region in Ontario, Canada as a case study. *Geoderma*, [s.l.], v. 283, p.78-87, dez. 2016. Elsevier BV. <http://dx.doi.org/10.1016/j.geoderma.2016.07.028>.

AKUMU, C; JOHNSON J.A.; ETHERIDGE D; UHLIG P. ; WOODS M; PITT D.G.; MCMURRAY S. GIS-fuzzy logic based approach in modeling soil texture: Using parts of the Clay Belt and Hornepayne region in Ontario Canada as a case study. *Geoderma*, [s.l.], v. 239-240, p.13-24, fev. 2015. Elsevier BV. <http://dx.doi.org/10.1016/j.geoderma.2014.09.021>.

ALVES SOBRINHO, T.; PERTUSSATTI, C. A.; REBUCCI, L.C. S.; OLIVEIRA, P. T. S. Estimativa da erosividade local das chuvas, utilizando redes neurais artificiais. *Revista Ambiente e Água, Taubaté*, v. 6, n. 2, p.246-254, 2011.

ANDRADE, F.; SANTOS, J.; CARNEIRO,L.; PAULO, R. Notas de aula. Disponível em:<http://paginas.fe.up.pt/~eol/IA/IA0708/APONTAMENTOS/IAlogica_conjuntos_difusos.pdf> . Acesso em : Ago, 2014.

ARRUDA, G. O; DEMATTÊ, J. A. M. CHAGAS, C. S. Mapeamento digital de solos por redes neurais artificiais com base na relação solo-paisagem. *Revista Brasileira Ciência do Solo*, v. 37, n.2, p.327-338, 2013.

ASHTEKAR, J.M.; OWENS, P.R. Remembering Knowledge: Na expert knowledge based approach to digital soil mapping. *Soil Horizons*. v. 54, n 5, p. 1-6, 2013.

BARIN, A.; CANHA, L.N. MAGNAGO, K.F.; ABAIDE, A.R. Seleção de fontes alternativas de geração distribuída utilizando uma análise multicriterial baseada no método ahp e na lógica fuzzy. *Revista Controle & Automação*. v.21 n.5, p.477-486, 2010.

BARRETO-NETO, A. A.; SOUZA FILHO, C. R. de. Application Of Fuzzy Logic To The Evaluation Of Runoff In A Tropical Watershed. *Environmental Modeling & Software*, v23, n1, p.244-253. 2008.

BECKETT, P. H. T.; WEBSTER, R. Soil variability: a review. *Soils and Fertilizers*, Wallingford, v. 34, p. 1-15, 1971.

BEHRENS T.; Zhu A-X; Schmidt K; Scholten T. Multi-scale digital terrain analysis and feature selection for digital soil mapping. *Geoderma*, [s.l.], v. 155, n. 3-4, p.175-185, mar. 2010. Elsevier BV. <http://dx.doi.org/10.1016/j.geoderma.2009.07.010>.

BENINI, L.C. Modelagem neuro-fuzzy para precisão da densidade do solo. *Anais do XXXIV Congresso Nacional de Matemática aplicada e computacional*. p.856-857. 17 a 21 de Set, 2012.

BONISCH, S.; LOPES ASSAD, M.L.; CÂMARA, G.; MONTEIRO, A.M.V. Representação e propagação de incertezas em dados solos. *Revista Brasileira de Ciência do Solo*. v.28, n.1, p.21-32, 2004.

BUI, Le Vinh; STAHR, Karl; CLEMENS, Gerhard. A fuzzy logic slope-form system for predictive soil mapping of a landscape-scale area with strong relief conditions. *Catena*, [s.l.], v. 155, p.135-146, ago. 2017. Elsevier BV. <http://dx.doi.org/10.1016/j.catena.2017.03.001>.

BURROUGH, P. A. Fuzzy mathematical methods for soil survey and land evaluation. *Journal of Soil Science, Oxford*, v. 40, p. 477-482, 1989.

CARVALHO JÚNIOR, W. de; CHAGAS, C. da 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*, v.68, n.6, p.691-696, 2011.

CARVALHO, A. de P.; MORAES NETO, J. M. de; LIMA, V. L. A. de; & SILVA, D. G.K. C. Estudo da degradação ambiental do açude Bodocongó. *Engenharia Ambiental - Espírito Santo do Pinhal*, v.6,n.2, p. 293-305, 2009.

COELHO, FF; GIASSON, E. Comparação de métodos para mapeamento digital de solos com utilização de sistema de informação geográfica. *Ciência Rural*, [s.l.], v. 40, n. 10, p.2099-2106, out. 2010. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0103-84782010005000156>.

CONGALTON, R.G.; GREEN, K. Assessing the accuracy of remotely sensed data:Principles and practices. *International Journal of Applied Earth Observation and Geoinformation*, v.11, n.2,p.448-449,1999.

CONGALTON, R.; A. REKAS. COMPAR: A computerized technique for the indepth comparison of remotely sensed data. *Proceedings of the Fifty First Annual Meeting of the American Society of Photogrammetry*, Washington, DC. 1985. p. 98-106.

COSTA, A. M.; CURI, N.; MENEZES, M. D.; ARAÚJO, E. F.; MARQUES, J. J. Levantamento detalhado de solos da microbacia hidrográfica do Horto Florestal Terra Dura (RS) e considerações sobre escala de mapeamento. *Ciência e Agrotecnologia, Lavras*, v. 33, n. 5, p. 1272-1279, set./out., 2009.

DIAS, LM da S. Predição de classes de solo por atributos do meio físico e de sensoriamento remoto em área da bacia sedimentar do São Francisco. [Dissertação]. Campinas, São Paulo. 127 p. 2015.

ELNAGGAR, A. A. Development of Predictive Mapping Techniques for Soil Survey and Salinity Mapping. 148 p. 2007. Thesis (PhD) – Oregon State University.

EMBRAPA. EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA -Centro Nacional de Pesquisa de Solos. Sistema brasileiro de classificação de solos. 2.ed. Rio de Janeiro, Embrapa Solos, 2006.316p.

FIGUEIREDO, S.R.; GIASSON, E. TORNQUIST, C.G.; NASCIMENTO, P.C. Uso de regressões logísticas múltiplas para mapeamento digital de solos no planalto médio do RS. Revista Brasileira de Ciência do Solo. v. 32, n. Spe, p.2779-2785, 2008.

FUJIMOTO, R.Y. Diagnóstico automático de defeitos em rolamentos baseado em Lógica Fuzzy. 158 p. 2005. (Dissertação de Mestrado). Escola Politécnica, Universidade de São Paulo.

HEUVELINK, G.B.M; WEBSTER, R. Modelling soil variation: past, present, and future. Geoderma, [s.l.], v. 100, n. 3-4, p.269-301, maio 2001. Elsevier BV. [http://dx.doi.org/10.1016/s0016-7061\(01\)00025-8](http://dx.doi.org/10.1016/s0016-7061(01)00025-8).

HODZA, Paddington. Fuzzy logic and differences between interpretive soil maps. Geoderma, [s.l.], v. 156, n. 3-4, p.189-199, maio 2010. Elsevier BV. <http://dx.doi.org/10.1016/j.geoderma.2010.02.016>.

HUDSON, B.D. The soil survey as paradigm-based science. Soil Sci. Soc. Am. J., v. 56, p.836–841. 1992.

IPPOLITI, G. A. R.; COSTA, L. M. da; SCHAEFER, C. E. G. R., FERNANDES FILHO, E. I. ; GAGGERO, M. R. Análise digital do terreno: ferramenta na identificação de pedoformas em microbacia na região de "mar de morros" (MG). Revista Brasileira de Ciência do Solo, vol. 29, núm. 2, pp. 269-276, 2005.

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

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

LAGACHERIE, P. 2008. Digital Soil Mapping: A State of the Art. In: HARTEMINK, A. E.; McBRATNEY, A.B., MENDONÇA-SANTOS M. L. (Eds) Digital Soil Mapping with Limited Data. Springer, p. 3-14.

LANDIS, J. R. & KOCH, G. G. The measurement of observer agreement for categorical data. *Biometrics*, v.33, n.1, p.159-174, 1977.

LIAO, W.H. Region Description Using Extended local Ternary Patterns. *International Conference on Pattern Recognition*. p. 1003-1006, 2010.

LOBÃO, J. S. B.; ROCHA, W. de J. S. da F; SILVA, A. de B. Utilização De Lógica Fuzzy Na Modelagem De Vulnerabilidade À Erosão No Município de Morro Chapéu – BA. In: III SIMPÓSIO REGIONAL DE GEOPROCESSAMENTO E SENSORIAMENTO REMOTO, Anais, Aracaju –SE. 2006.

MacMILLAN, R. A. Experiences with Applied DSM: Protocol, Availability, Quality and Capacity Building. In: HARTEMINK, A.E.; McBRATNEY, A.; MENDONÇA-SANTOS, M. DE L. (Ed.), *Digital Soil Mapping with Limited Data*. New York: Springer. Ed 1, cap. 10. p.113-135, 2008.

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

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

McBRATNEY, A. B.; WEBSTER, R. Spatial dependence and classification of soil along a transect in northeast Scotland. *Geoderma*, Amsterdam, v. 26, p. 63-82, 1981.

McBRATNEY, A.; MENDONÇA-SANTOS, M.L. (Ed.). *Digital Soil Mapping With limited Data*. Springer, p. 233-245, 2008.

MEDEIROS, S.; MELLO, R.; CAMPOS FILHO, P. Análise de projetos para unidades de conservação usando lógica fuzzy. *Produção*. v.17, n.2, p.317-329, 2007.

MENDONÇA-SANTOS, M de L. *Mapeamento digital de classes e atributos de solos: métodos, paradigmas e novas técnicas*. Rio de Janeiro: Embrapa Solos, (Embrapa Solos. Documentos; n. 55), 2003. 19p.

MENEZES, M. D.; SILVA, S. H. G.; OWENS, P. R.; CURI, N. Digital soil mapping approach based on fuzzy logic and field expert knowledge. *Ciência e Agrotecnologia*. 2013, vol.37, n.4, pp. 287-298.

MONSERUD, R. A. & LEEMANS, R. Comparing global vegetation maps with the Kappa statistic. *Ecological Modelling*, v.62,n.4, p.275-293, 1992.

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

MORAES, C.M.M. Fuzzycom- Componente de lógica fuzzy. Puc-Departamento de Engenharia Elétrica, p.1-15. Disponível em: <http://www.pucrio.br/pibic/relatorio_resumo2007/relatorios/ele/ele_claudio_magno_martins_moraes.pdf>. Acesso em: Ago, 2014.

NEUMANN, Marina Rolim Bilich. Mapeamento digital de solos, no Distrito Federal. 2012. 123 f. Tese (Doutorado em Geociências aplicadas). Instituto de Geociências. Universidade de Brasília. Brasília, DF. 2012.

PELEGRINO, MHP; SILVA, SHG; MENEZES MD de; SILVA E da; OWENS PR; CURIN. Mapping soils in two watersheds using legacy data and extrapolation for similar surrounding areas. *Ciência e Agrotecnologia*, [s.l.], v. 40, n. 5, p.534-546, out. 2016. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/1413-70542016405011416>.

PROJETO RADAMBRASIL. Folha SH 22, SH 21 e SI 22. Rio de Janeiro, IBGE. Levantamento dos Recursos Naturais,Série III, nº33. p.303-404, 1986.

QI, F.; ZHU, A.X. Knowledge Discovery from soil maps using inductive learning. *International Journal of Geographical Information Science*. v. 17, n 8, p. 771-795, 2003.

ROVEDA, R.G.A.; COSTA, L.M.; SHAEFER,C.E.G.R.; FERNANDES FILHO, E.J.; GAGGERO, M.R. Análise digital do terreno ferramenta na identificação de pedofomas em microbacia na região de mar de morros (MG). *Revista Brasileira de Ciência do Solo*. v.29,n.1.p.269-276, 2005.

SALIBY, E. Descriptive Sampling: An Improvement Over Latin Hypercube Sampling. In: *Winter Simulation Conference*, v. 1, n 1, p. 1-6, 1997.

SANCHEZ, P.A.; AHAMED, S.; CARRÉ, F.; HARTEMINK, A.E.;HEMPEL, J.; HUISING, J.; LAGACHERIE, P.; MCBRATNEY,A.B.; MCKENZIE, N.J.; MENDONC-SANTOS, M. de L.;MINASNY, B.; MONTANARELLA, L.; OKOTH, P.; PALM, C.A.;SACHS, J.D.; SHEPHERD, K.D.; VÅGEN, T.-G.; VANLAUWE, B.; WALSH, M.G.; WINOWIECKI, L.A.; ZHANG, G.-L. Digital soil map of the world. *Science*, v.325, n 5941, p.680-681, 2009.

SANDRI, S.; CORREA, C. *Lógica Nebulosa*. São José dos Campos: ITA, 1999. Disponível em: <<http://www.ele.ita.br/cnrm/minicursos-5ern/log-neb.pdf>>. Acesso em: 22 jul.2014

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. Viçosa, MG, Sociedade Brasileira de Ciência do Solo/Rio de Janeiro, Embrapa, 2005. 100p.

SARMENTO, E. C. Comparação entre quatro algoritmos de aprendizagem de máquina no mapeamento digital de solos no Vale dos Vinhedos, RS, Brasil. 109 p. 2010. (Tese de Doutorado). Universidade Federal do Rio Grande do Sul.

SCHULER, U., HERRMANN, L., INGWERSEN, J., ERBE, P., STAHR, K. Comparing mapping approaches at subcatchment scale in northern Thailand with emphasis on the Maximum Likelihood approach. *Catena* 81, 137–171.2010.

SEYFRIED, M. Spatial variability constraints to modeling soil water at different scales. *Geoderma*, [s.l.], v. 85, n. 2-3, p.231-254, ago. 1998. Elsevier BV. [http://dx.doi.org/10.1016/s0016-7061\(98\)00022-6](http://dx.doi.org/10.1016/s0016-7061(98)00022-6).

SHI, A.X; LONG, R.; DEKETT, R.; PHILIPPE, J. Integrating different types of knowledge for digital soil mapping. *Soil Science Society of America Journal*. v. 73, n 5, p. 1682-1692, 2009.

SHI, X. ArcSIE User's Guide. Spatial Inference Enterprises-SIE LLC. 29 Heidi Ln Bow, NH 03304, USA. 120 p. 2013. Disponível em :<http://www.arcsie.com/Download/ArcSIE_UsersGuide_130319.pdf> , Acesso em Out 2017.

SHI, X; Long R., Dekett R.; Philippe J. Integrating Different Types of Knowledge for Digital Soil Mapping. *Soil Science Society Of America Journal*, [s.l.], v. 73, n. 5, p.1682-1692, 2009. Soil Science Society of America. <http://dx.doi.org/10.2136/sssaj2007.0158>.

SILVA, Cristiano Cassiano. Mapeamento digital de classes de solo: aplicação de metodologia na folha Botucatu (SF-22-Z-B-IV-3) e validação de campo. 117 p. 2012. (Dissertação de Mestrado). Instituto Agrônômico de Campinas.

SILVA, S. H. G. Cambisol (Inceptisol) Solum thickness digital mapping based on expert knowledge with limited data from a watershed in Minas Gerais, Brazil. 41 p. Monografia – Universidade Federal de Lavras, Lavras, 2013.

SILVA, SHG, MENEZES, MD de; OWENS, PR; CURI, N. Retrieving pedologist's mental model from existing soil map and comparing data mining tools for refining a larger area map under similar environmental conditions in Southeastern Brazil. *Geoderma*, [s.l.], v. 267, p.65-77, abr. 2016. Elsevier BV. <http://dx.doi.org/10.1016/j.geoderma.2015.12.025>.

SILVA, SHG; MENEZES MD de; OWENS PR; CURI N. Retrieving pedologist's mental model from existing soil map and comparing data mining tools for refining a larger area map under similar environmental conditions in Southeastern Brazil. *Geoderma*, [s.l.], v. 267, p.65-77, abr. 2016. Elsevier BV. <http://dx.doi.org/10.1016/j.geoderma.2015.12.025>.

SILVA,S.A.; LIMA,J.S. de; SOUZA, G.S.; XAVIER, A.C. Distribuição e incerteza da acidez de um Latossolo vermelho-amarelo húmico sob o cultivo do café. *Revista Brasileira de Ciência do Solo*. v. 33; n.1; p.1053-1060, 2009.

STOLLE,L; LINGNAU,C.; ARCE, J.E.; BOGNOLA, I.A. Aplicação de um sistema baseado em conhecimento e lógica fuzzy para o mapeamento da fragilidade ambiental em

relação às atividades florestais. Anais do XIV Simpósio Brasileiro de Sensoriamento Remoto.p. 3103-3110. 25 a 30 de Abr, 2009.

SUAREZ,A.F. E CANDEIAS, A.L.B. Avaliação da acurácia da classificação de dados de sensoriamento remoto para o município de Maragogipe. IV Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação. Recife/PE.p 1-10. 2012.

TANSCHKEIT, R. Sistemas fuzzy. PUC-Rio, Rio de Janeiro, 2014. Disponível em < <http://www2.ica.ele.puc-rio.br/Downloads/41/LN-Sistemas%20Fuzzy.pdf>> Acesso em 21 jul, 2014.

TESKE, R. Teste metodológicos para o mapeamento digital de classes de solo utilizando árvores de decisão.[Tese]. Universidade Federal do Rio Grande do Sul, Porto Alegre, RS. 2014.

ZADEH, L.a.. Fuzzy sets. Information And Control, [s.l.], v. 8, n. 3, p.338-353, jun. 1965. Elsevier BV. [http://dx.doi.org/10.1016/s0019-9958\(65\)90241-x](http://dx.doi.org/10.1016/s0019-9958(65)90241-x).

ZHU, Q.; LIN, H. S. Comparing ordinary kriging and regression kriging for soil properties in contrasting landscapes. Pedosphere, v. 20, n. 5, p. 594-606, 2010.

ZHU, A.X., YANG, L., LI, B., QIN, C.Z., PEI, T., LIU, B. Construction of membership functions for predictive soil mapping under fuzzy logic. Geoderma 155, 164–174.2010.

ZHU, A.X.; YANG, L.; LI,B.; QIN, C. ENGLISH, E.; BURT, J.; ZHOU, C. Purposive Sampling for digital soil mapping for areas with limited data. In: HARTEMINK, A.E.; McBRATNEY, A.B. & MENDONÇA SANTOS, M.L., eds. Digital soil mapping with limited data. New York, Springer-Verlag, 2008. p.233-245.

ZHU, A.X; YANG L; LI B; QIN C; PEI T; LIU B. Construction of membership functions for predictive soil mapping under fuzzy logic. Geoderma, [s.l.], v. 155, n. 3-4, p.164-174, mar. 2010. Elsevier BV. <http://dx.doi.org/10.1016/j.geoderma.2009.05.024>.

ZHU, A-X A similarity model for representing soil spatial information. Geoderma, [s.l.], v. 77, n. 2-4, p.217-242, jun. 1997. Elsevier BV. [http://dx.doi.org/10.1016/s00167061\(97\)00023-2](http://dx.doi.org/10.1016/s00167061(97)00023-2).

ZHU, A-X; FENG Q; MOORE A; BURT J E. Prediction of soil properties using fuzzy membership values. Geoderma, [s.l.], v. 158, n. 3-4, p.199-206, set. 2010. Elsevier BV. <http://dx.doi.org/10.1016/j.geoderma.2010.05.001>.

ZHU,A.X.; BAND, L.; VERTESSY, R.; DUTTON, B. Derivation of soil properties using a soil land inference model (SoLIM). Soil Science Society of American Journal. v. 61, n 2, p. 523-533, 1997.

CAPÍTULO II

USO DE *MACHINE LEARNING* PARA PREDIÇÃO DE CLASSES DE SOLOS APLICADOS A UMA BACIA HIDROGRÁFICA

RESUMO

Diversas técnicas de mapeamento digital vêm sendo desenvolvidas para melhorar a eficiência deste em várias áreas de estudo. Diante disso, tem-se um gama de métodos a serem escolhidos, cuja o desempenho deve ser testado. Para um mapeamento satisfatório, é necessário a boa seleção de variáveis preditoras, dessa forma, este trabalho objetivou testar diferentes modelos para mapeamentos de classes de solo na bacia do Ribeirão Inhaúma, em Iconha/ES, utilizando como variáveis preditoras: covariáveis do terreno, bandas do Landsat 8, Índice de vegetação da diferença normalizada- NDVI e dados de magnetometria e gamaespectrometria. Inicialmente, foi realizado o levantamento de solos que consistiu em 285 pontos de observação, 12 perfis e três amostras extras, totalizando 300 pontos . Em cada ponto de observação foi determinado a classe de solo, enquanto nos perfis e amostras extras foi feita a descrição morfológica, para predição das classes de solos foram usadas como variáveis independentes: 38 variáveis morfométricas derivadas do Modelo Digital de Elevação-MDE, NDVI e dados gamaespectrométricos e magnométricos obtidos através da Companhia de Pesquisa de Recursos Minerais -CPRM. O desempenho do modelo foi avaliado pelo método de amostragem *bootstrap* 632, através dos valores de Kappa, além das acurácias do produtor e do usuário e teste Z. Os maiores valores de Kappa foram encontrados para RF (0,65), XGBTREE (0,65), RANGER (0,64) e WSRF (0,63). As classes que obtiveram maior acurácia tanto do produtor quanto do usuário foram Cambissolo e afloramento rochoso. As variáveis mais importantes na predição das classes para os métodos RF, XGBTREE e RANGER foram o uso do solo, bandas do Landsat 8, seguidas por covariáveis do terreno e dados de magnetometria e gamaespectrometria. O teste Z demonstrou que RF e XGBTREE foram superiores em relação a MARS, porém mostraram-se igualmente eficientes aos demais.

ABSTRACT

Several techniques of digital mapping have been developed to improve the efficiency of this in several areas of study. Given this, there is a range of methods to be chosen, whose performance must be tested. For a satisfactory mapping, a good selection of predictor variables is necessary, so this work aimed to test different models for soil class mapping in the Ribeirao Inhaúma basin, in Iconha / ES, using as predictor variables: soil covariates, bands of Landsat 8, vegetation index of normalized difference - NDVI and data of magnetometry and gamma spectrometry. Initially, the soil survey consisted of 285 observation points, 12 profiles and three extra samples, totaling 300 points. At each observation point, the soil class was determined, while in the profiles and extra samples the morphological description was made, to predict the soil classes were used as independent variables: 38 morphometric variables derived from the Digital Elevation Model-MDE, NDVI and gamma-spectrometric and magnometric data obtained through the Company of Research of Mineral Resources -CPRM. The performance of the model was evaluated by the bootstrap 632 sampling method, using Kappa values, as well as producer and user acuity and Z test. The highest Kappa values were found for RF (0.65), XGBTREE (0, 65), RANGER (0.64) and WSRF (0.63). The classes that obtained greater

accuracy of both the producer and the user were Cambissolo and rock outcrop. The most important variables in the prediction of the classes for the RF, XGBTREE and RANGER methods were soil use, Landsat 8 bands, followed by terrain covariates and data from magnetometry and gamma spectrometry. The Z test showed that RF and XGBTREE were superior in relation to MARS, but they were equally efficient in the others.

1 INTRODUÇÃO

A técnica convencional de mapear solos ainda persiste como a forma mais aplicada para mapeamento deste compartimento da paisagem (SCULL et al., 2003). Porém, segundo Dias (2015), trata-se de método que demanda tempo, tem custo elevado e é subjetivo, dependendo quase que exclusivamente do conhecimento e julgamento do pedólogo, além de estar mais suscetível a erros durante o esboço das classes (ZHU et al., 2001), resultando em uma dependência excessiva de conhecimento tácito, fazendo com que muitas vezes se crie informações incompletas relativa à derivação do produto final da pesquisa de solo (HUDSON, 1992).

Dessa forma, têm sido cada vez mais adotadas novas técnicas para se obter informações sobre o solo de modo a diminuir os gastos do mapeamento convencional (CHAGAS et al., 2011). Em função da elevada complexidade em determinar as relações das classes de solos com as respectivas paisagens em que se encontram, torna-se necessário a adoção de técnicas mais objetivas e eficientes para o delineamento das unidades de mapeamento (LIMA et al., 2013). Nesse contexto a Pedometria apresenta-se como uma alternativa viável principalmente devido aos avanços tecnológicos nas últimas décadas, o que criou um grande potencial para a melhoria da maneira que os mapas de solo são produzidos (McKENSIE et al., 2000).

McBratney em dois momentos (1996 e 2007) define a Pedometria como sendo uma ciência fundamentada nas relações matemáticas, modelos numéricos ou estatísticos, através dos quais é possível estabelecer relações entre classes de solos e as covariáveis usadas para predição, delimitando assim a distribuição das mesmas no ambiente.

O mapeamento digital de solos, entra como uma das interfaces possíveis dentro da pedometria e é definido como sistemas de informações espaciais de solos, que são criados e manipulados, pelo uso de modelos numéricos de modo a inferir as variações espaciais e temporais de classes e propriedades do solo, utilizando observações, conhecimento dos solos e variáveis ambientais correlacionadas (LAGACHERIE & McBRATNEY 2007).

Alguns trabalhos demonstram o uso de diferentes métodos para o mapeamento digital de solos (TEN CATEN et al., 2011; CHAGAS et al., 2011; CHAGAS et al., 2013; TEN CATEN et al., 2009; SIRTOLI, 2008; BEHRENS e SCHOLTEN, 2007; Giasson et al., 2006;), e a escolha de qual método ou modelos aplicados à espacialização das classes de solos dependerá parcialmente de como o solo é percebido no espaço geográfico (SCULL et al., 2003).

Este trabalho tem por objetivo testar diferentes modelos para mapeamentos digital de classes de solo na bacia do Ribeirão Inhaúma, em Iconha/ES, utilizando como variáveis preditoras covariáveis ambientais disponíveis gratuitamente.

2 REVISÃO BIBLIOGRÁFICA

Primeiramente serão sucintamente apresentados os Classificadores que foram testados neste trabalho.

2.1 Classificadores Utilizados Para Predição De Classes De Solo

2.1.1 *Bagged AdaBoost*

No software R o pacote Adabag implementa o algoritmo Adaboost.M1 de Freund e Schapire (1996) e o algoritmo Bagging, de Breiman (1996) usando árvores de classificação como classificadores individuais. Uma vez que esses classificadores tenham sido treinados, eles podem ser usados para prever novos dados onde antes não existiam. O Adaboost.M1 é considerado o algoritmo multiclasse do classificador AdaBoost, sendo que o número de classes podem assumir 3 ou mais categorias, e é considerado um algoritmo facilmente implementado (FREUND e SCHAPIRE, 1997).

De acordo com Oliveira (2016), o algoritmo Boosting a cada iteração vai atribuindo pesos maiores às observações classificadas incorretamente na iteração anterior. Sendo assim ele aplica de forma sequencial uma regra de classificação final a amostra de treinamento, que são iterativamente reponderadas. Dessa forma a regra de classificação final é obtida pela combinação das regras dos classificadores parciais que foram construídos.

Segundo Fernandes et al.(2014), o AdaBoost, é o mais representativo algoritmo de Boosting, este cria classificadores em série, e a cada nova iteração é dado mais ênfase a aquelas instâncias classificadas erroneamente na iteração feita anteriormente. Assim, o peso das amostras classificadas incorretamente aumenta enquanto daquelas classificadas corretamente diminui.

A Figura 15, mostra de forma simplificada como essa classificação é feita.

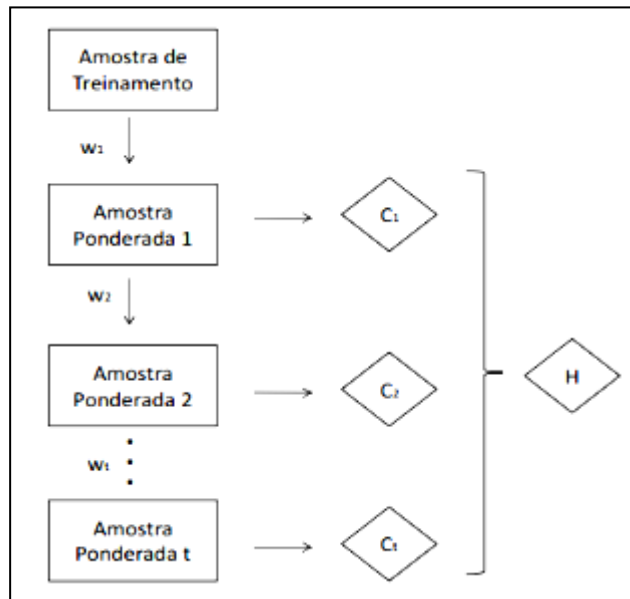


Figura 15- Fluxograma de funcionamento do algoritmo *Boosting*.

(w_i são os pesos atribuídos às observações em cada iteração t ; os C_i são os classificadores parciais encontrados em cada amostra ponderada e H é a hipótese final de classificação).

Fonte: Oliveira (2016).

Quando feita a primeira iteração, o classificador base tem como objetivo realizar a correção dos pesos atribuídos aos objetos classificados incorretamente com fins de se obter o erro mínimo (OLIVEIRA, 2016). De acordo com Schapire e Freund (2012), ao melhorar ligeiramente o erro de treino de um classificador fraco, o erro de treinamento será decrescido exponencialmente.

2.1.2 C5.0

Método criado por Max Kuhn, Steve Weston, Nathan Coulter, Mark Culp (2015). As árvores de decisão C5.0 são modelos baseados em regras para reconhecimento de padrões.

Este modelo amplia as aplicações dos algoritmos de classificação C4.5, descritos em Quinlan (1992). O algoritmo C4.5 constrói árvores de decisão a partir de um conjunto de dados de treinamento, em cada nó da árvore, o algoritmo escolhe o atributo dos dados que mais efetivamente divide seu conjunto de amostras em subconjuntos enriquecidos em uma classe ou outra. O critério de divisão é o ganho de informação normalizado. O atributo com o maior ganho de informação normalizado é escolhido para tomar a decisão.

Os detalhes das extensões do C4.5 para o C5.0 são em grande parte não documentados. O modelo C5.0 pode assumir a forma de uma árvore de decisão completa ou de uma coleção de regras.

Este modelo particular manipula dados não-numéricos de alguns tipos (como caracteres, fator e dados ordenados). A matriz de custo deve ser $C \times C$, onde C é o número de classes. Os elementos diagonais são ignorados. As colunas devem corresponder às classes verdadeiras e as linhas são as classes previstas. Por exemplo, se $C = 3$ com classes

Vermelho, Azul e Verde (RGB nessa ordem), um valor de 5 no elemento (2,3) da matriz indicaria que o custo de prever uma amostra Verde como Azul é cinco vezes o valor usual (de um) (PACKAGE C5.0).

Algumas das melhorias implantadas de C5.0 em relação a C4.5 são a velocidade, onde o C5.0 é significativamente mais rápido do que C4.5 (PATIL et al., 2012); uso de memória mais eficiente; Árvores de decisão menores, C5.0 consegue obter resultados semelhantes ao C4.5 com árvores de decisão consideravelmente menores. De acordo com Ruiz (2015), foi incorporado ao C5.0 um método reforço (*boosting*) na determinação do modelo final, e uma abordagem para seleção de atributos mais informativos na construção das árvores, denominado *winnow*.

Segundo Kung e Jonhson (2013) a seleção de atributos é feita através de uma divisão aleatória dos dados de treinamento em duas partes iguais, posteriormente é gerado o modelo de árvore, chamada árvore teste com uma dessas partes, essa tem o objetivo de avaliar a utilidade de cada atributo utilizado como preditor. De forma, que atributos preditivos que não estão presentes em nenhuma divisão da árvore são considerados sem importância para o modelo e são removidos do conjunto de dados. A outra metade não utilizada para gerar a árvore teste, é usada para se estimar a taxa de erro da árvore, essa taxa é calculada com e sem cada atributo preditivo, se a taxa de erro aumenta na ausência do atributo, esse permanece no conjunto de dados.

De acordo com IBM (2017):

Um modelo C5.0 funciona dividindo a amostra com base no campo que fornece o máximo de **ganho de informações**. Cada subamostra definida pela primeira divisão é, então, dividida novamente, geralmente com base em um campo diferente, e o processo é repetido até que as subamostras não possam ser divididas ainda mais. Por último, as divisões de nível inferior são reexaminadas e aquelas que não contribuírem significativamente com o valor do modelo são removidas ou **podadas**.

2.1.3 *Conditional Inference Random Forest*

O algoritmo Cforest (*Conditional Inference Random Forest*) tem como algoritmo base o conditional inference *trees* (ctrees), cujo meta-algoritmo é o *bagging* (SANTOS, 2013). O algoritmo ctree emprega a inferência condicional como processo de partição em subconjuntos de forma binária e recursiva (HOTHORN et al, 2006). O pacote R é o *party*, esse pode ser considerado uma caixa de ferramentas computacional para particionamento recursivo.

Corroborando com Santos (2013) e Horthorn et al. (2006), de acordo com o PACKAGE PARTY (2017), o núcleo do pacote *ctree*, é uma implementação de árvores de inferência condicional que incorporam modelos de regressão estruturados em árvore, em uma teoria bem definida de procedimentos de inferência condicional. Esta classe não-paramétrica de árvores de regressão é aplicável a todos os tipos de problemas de regressão, incluindo variáveis nominais, ordinais, numéricas.

O procedimento *Bagging* (BREIMAN, 1996) versa em usar amostras *bootstrap* do conjunto de treinamento, e assim dá origem a diversos classificadores, que irão ser usados para formação do classificador final (OLIVEIRA,2016). Amostragem *bootstrap* é uma técnica na qual é feita uma amostragem com reposição, assim, partindo de um primeiro conjunto de treinamento, são selecionados de forma aleatória preditores para um novo conjunto de treinamento (OSHIRO, 2013).

Ao criar preditores em amostras *bootstrap* e posteriormente agregá-los, espera-se formar um preditor que seja superior, sendo que no caso de classificação as combinações

são feitas através de votos, sendo que a classe mais votada é escolhida, enquanto no caso da regressão é utilizada a média dos preditores para tomada de decisão (RUBESAM, 2004).

A Figura 16, mostra o esquema de funcionamento do *bagging*.

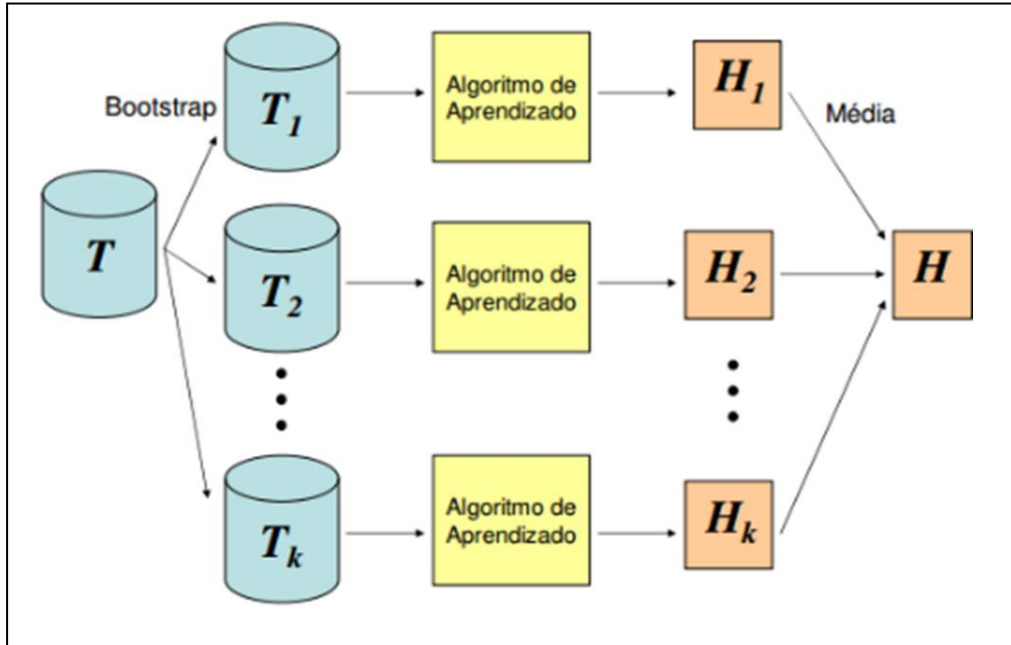


Figura 16- Esquema de funcionamento do algoritmo *bagging*. T é a base de treinamento, H são as hipóteses.

Fonte: Zadrozny, 2010

Em síntese, o principal foco de *Bagging* é dar origem a vários classificadores em um dado conjunto de treinamento, que posteriormente serão utilizados para originar o classificador final (OLIVEIRA, 2016). Sendo *cforest*, segundo o PACKAGE PARTY (2017) uma implementação dos algoritmos de conjuntos aleatórios de floresta utilizando inferência condicional como base de aprendizado.

2.1.4 *Extreme Gradient Boosting*

Xgboost é a abreviação do pacote *eXtreme Gradient Boosting*. Trata-se de uma implementação eficiente e escalonável da estrutura de aumento de gradiente proposta por (FRIEDMAN, 2001; FRIEDMAN et al., 2000). O pacote inclui algoritmo de resolução de modelos linear e algoritmo de aprendizagem por árvore. Suporta várias funções objetivas, incluindo regressão e classificação (CHEN e HE, 2017).

Os modelos de árvore foram introduzidos por Breiman et al. (1984) como Árvore de Classificação e Regressão (CART). O objetivo dos métodos de árvore são criar partições recursivas para o espaço preditor x , a fim de criar grupos homogêneos para o resultado y (ESCABIAS, 2017). Estas partições recursivas ocorrem até que haja apenas uma observação deixada ou até que um critério de parada seja alcançado. As árvores têm partes diferentes; Os nós finais em um algoritmo de árvore são as folhas e o número de níveis de partição que uma árvore consegue é a profundidade de uma árvore.

Escabias (2017), apresenta algumas das vantagens apresentadas pelo algoritmo, sendo elas:

- **Computação rápida:** A capacidade de fazer computação paralela em uma única máquina, bem como a implementação de um algoritmo de dispersão, o torna cerca de 10 vezes mais rápido que outras implementações (CHEN E GUESTRIN, 2016);
- **Flexível:** pode trazer funções objetivas e facilmente lidar com valores perdidos;
- **XGBoost** permite executar a validação cruzada em cada iteração do algoritmo.
- **Open source:** Existe uma comunidade que melhora o algoritmo na *Distributed (Deep) Machine Learning Community* (2017).

O resumo dos parâmetros mais importantes para os aprendizes de árvores que devem ser ajustados são apresentados abaixo (ESCABIAS, 2017):

- **Eta (η):** taxa de aprendizagem do algoritmo. Intervalo: [0, 1]. Valor padrão: 0.3.
- **Gamma (γ):** redução mínima de perda dividida.
É a redução de perda mínima para fazer outra partição em um nó folha. Para valores maiores o algoritmo será mais conservador. Intervalo: [0, inf]. Valor padrão: 0.
- **max depth:** profundidade máxima de uma árvore.
Um valor baixo evita o *overfitting*. Maior valor poderia aprender relacionamentos específicos para os dados de treinamento. Intervalo: [0, inf]. Valor padrão: 6
- **min child weight (hi):** proxy de observação mínima para uma criança.
É a soma mínima exigida do Hessian. Se a soma for menor que o parâmetro, não haverá mais particionamento. Intervalo: [0, inf]. Valor padrão: 1.
- **subsample:** proporção de amostras aleatórias a serem usadas para treinamento.
Equivalente da porcentagem de observações aleatórias coletadas para o crescimento de cada árvore. Evita o *overfitting*. Intervalo: [0, 1]. Valor padrão: 1.
- **colsample bytree:** relação de colunas usadas para treinar cada árvore
Equivalente da porcentagem de colunas aleatórias coletadas para o cultivo de árvores. Intervalo: (0, 1]. Valor padrão: 1.
- **colsample bylevel:** Relação de colunas utilizadas para treinar cada divisão de uma árvore.
O mesmo que antes, mas a randomização das colunas são para cada divisão. Intervalo: (0, 1) Valor predefinido: 1.
- **lambda (λ):** Regularização nos pesos.
L2 termo de regularização para os pesos. Aumentar esse valor aumentará o parâmetro de regularização. Valor padrão: 1.

- **alpha (α):** Regularização nos pesos.
L1 termo de regularização para os pesos. Aumentar esse valor aumentará o parâmetro de regularização. Valor padrão: 0
- **tree method:** algoritmo de construção da árvore
- **sketch eps:** Número de candidatos para o algoritmo
Este parâmetro seleciona o número de candidatos por recurso a ser avaliado em cada divisão. Valores mais baixos vão propor mais candidatos e ser mais precisos. Intervalo: (0, 1). Valor padrão: 0.03.
- **max delta step and scale pos weight:** Útil para classes extremamente desequilibradas.

2.1.5 *Multivariate Adaptive Regression Splines*

MARS é um método desenvolvido por Friedman (1991), sendo considerado um modelo de regressão flexível (ANDRIOLO, 2006), onde é possível a análise de dados de grande dimensão. É um teste não paramétrico, onde não se supõe anteriormente a forma como as variáveis dependentes estão relacionadas com os preditores. Uma das vantagens do método, apontado por Briand et al. (2000) é que divide o espaço amostral em sub-regiões, em que cada uma contém um conjunto de variáveis independentes, onde foram identificadas interações e ajustada uma regressão linear, limitando assim a complexidade do modelo e evitando o excesso de aproximação dos dados.

Segundo Friedman (1991) MARS é usado para modelagem por regressão de dados dimensionais elevados. O modelo assume a forma de uma expansão nas funções de base *spline* do produto, onde o número de funções de base, bem como os parâmetros associados a cada uma (posições do grau e do nó do produto) são automaticamente determinados pelos dados. Este procedimento é motivado pela abordagem de divisão recursiva para a regressão.

Ao contrário da partição recursiva, no entanto, este método produz modelos contínuos com derivadas contínuas. Tem mais poder e flexibilidade para modelar relações que são quase aditivas ou envolvem interações de no máximo algumas variáveis. Além disso, o modelo pode ser representado de uma forma que identifique separadamente as contribuições aditivas e as associadas, às diferentes interações multivariáveis.

2.1.6 *Random Forest*

Random Forest (RF) ou Floresta Aleatória é uma classe de métodos de conjuntos desenvolvidos para classificadores de árvore de decisão (LOPES, 2013). Ele combina previsões feitas por múltiplas árvores de decisão, onde cada árvore é gerada a partir dos valores de um vetor aleatório. Estes vetores são gerados a partir de uma distribuição de probabilidade fixa (TAN et al., 2009).

Random forest, foi definido de acordo com Breiman, 2001, o desenvolvedor do método, como um classificador composto por uma coleção de árvores, ou seja, várias árvores de decisão. Cada uma dessas árvores de decisão dá um voto que indica sua decisão

sobre a classe à qual pertencerá um determinado objeto. Determinado objeto pertencerá a classe que obtiver o maior número de votos (RODRIGUES E AMARAL, 2012).

O modelo RF combina um conjunto de árvores de regressão, a fim de melhorar a precisão da predição, cada árvore é construída a partir da seleção aleatória de covariáveis preditoras e pontos de treinamento (TAQUES, 2014). O algoritmo de RF é bastante robusto, insensível a ruídos nos valores das variáveis e ao *overfitting* (super ajuste); Outra característica importante é a capacidade de fornecer medidas de importância das covariáveis na predição (HEUNG et al., 2014).

Sá Lucas (2011), cita que esta técnica consiste essencialmente em três partes:

- *Bagging*: gerar, através de reamostragem, um conjunto de sub-amostras provenientes da amostra original, selecionadas de maneira aleatória simples com reposição;
- *Boosting*: construir, para cada subamostra, um modelo de árvore de decisão, aumentando a ponderação das observações incorretamente classificadas com base nos modelos criados para as outras subamostras; e
- *Randomizing*: desenvolver os modelos de árvore de decisão utilizando, em cada nó, um conjunto de variáveis selecionadas aleatoriamente, diferente para cada nó.

2.1.7 Ranger

O nome do pacote ranger vem da abreviação de *RANdom forest GEnerator* (ranger) (Wright e Ziegler, 2015).

Trata-se de uma implementação rápida do algoritmo Random Forest (BREIMAN, 2001), particularmente adequado para dados dimensionais elevados. É possível trabalhar com técnicas de classificação e regressão, as quais são implementadas como no *Random Forest* original (BREIMAN, 2001).

De acordo com o Wright (2017), o algoritmo consiste de 3 etapas:

1. Criar uma floresta aleatória nos dados de treinamento;
2. Para cada observação de interesse (dados de teste), os pesos de todas as observações de treinamento são computados contando o número de árvores nas quais ambas as observações estão no mesmo nó terminal;
3. Para cada observação de teste, crescer uma floresta aleatória ponderada nos dados de treinamento, usando os pesos obtidos na etapa 2. Preveja o resultado da observação do teste.

Wright e Ziegler (2015) realizando testes de simulação demonstraram a eficiência computacional e de memória do *ranger*. O tempo de execução foi melhor dimensionado para o número de recursos, amostras, árvores e o valor de *mtry* (número de atributos utilizados para construir cada árvore), e nenhuma implementação de RF se mostrou mais rápida para dados dimensionais elevados com muitos recursos.

O número de árvores necessárias para a análise depende de vários fatores, como o tamanho do conjunto de dados e o objetivo do estudo. No entanto, o tempo de execução e uso de memória também pode afetar a escolha do número de árvores e outros parâmetros (WRIGHT e ZIEGLER, 2015). Com implementações mais rápidas disponíveis, como o *ranger*, esses desafios computacionais podem ser enfrentados.

2.1.8 Suport Vector Machine com Kernel Linear e Polynomial

De acordo com Santos (2002), o *Support Vector Machines* (SVM) é uma técnica de aprendizagem de máquina proveniente de duas fundamentações: Teoria da Aprendizagem Estatística (VAPNIK, 1999) e Otimização Matemática (MATTERA et al., 1999). O SVM pode ser considerado uma técnica computacional de aprendizado cujo enfoque é o reconhecimento de padrão, assim esse algoritmo tem como objetivo determinar limites de decisão onde haja separação ótima entre as classes com minimização dos erros (NASCIMENTO et al., 2009), permitindo dessa forma a maximização dos limites do hiperplano de separação ótima (*optimal hyperplane*) (FARIA e FERNANDES FILHO, 2013).

Segundo Faria e Fernandes Filho (2013), os pontos próximos ao limite de separação são definidos vetores de suporte (*support vectors*), como mostrado na Figura 17.

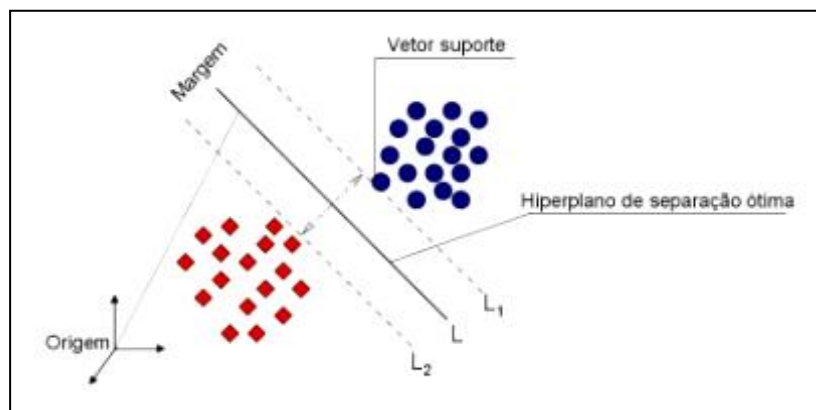


Figura 17- Esquema de classificação por meio do Support Vector Machines-SVM.

Fonte: Nascimento et. al, 2009 modificado de Huang et al. (2002) e Melgani e Bruzzone (2004).

É permitido ao usuário controlar os limites de aceitação e rejeição dos erros de treinamento do conjunto de dados (TSO e MATHER, 2009).

O SVM usam as funções kernel para separação dos dados, ondem Santos (2002), relata que :

As funções Kernel fazem o mapeamento dos dados do espaço de entrada para um espaço de características com uma dimensão muito superior, possibilitando que os dados não separáveis linearmente, tornem-se separáveis no espaço de características .

Quando se tem dados que possuam uma distribuição linear ou aproximadamente linear, as SVMs serão eficazes na classificação, já que estes dados serão linearmente separáveis (LORENA e CARVALHO, 2007).

O Kernel linear é a função kernel mais simples possível. Lida com problemas linearmente separáveis. Ele é dado pelo produto interno $x; x'$ somado a uma constante c (SOUZA, 2016):

$$k(x, \tilde{x}) = (\alpha x' \tilde{x} + c) \quad (1)$$

O grande entrave do uso de Kernel linear, vem da dificuldade de se encontrarem em casos reais, dados que sejam linearmente separáveis, isso devido a fatores como a presença de *outliers*, ruídos ou a própria natureza dos dados, que em muitos casos não se distribui de forma linear (LORENA e CARVALHO, 2007). Por isso usa-se o procedimento para suavização das margens do classificador, permitindo que alguns dados fiquem entre um hiperplano e outro (Figura 18), ao contrário do que acontece quando utiliza-se margens rígidas.

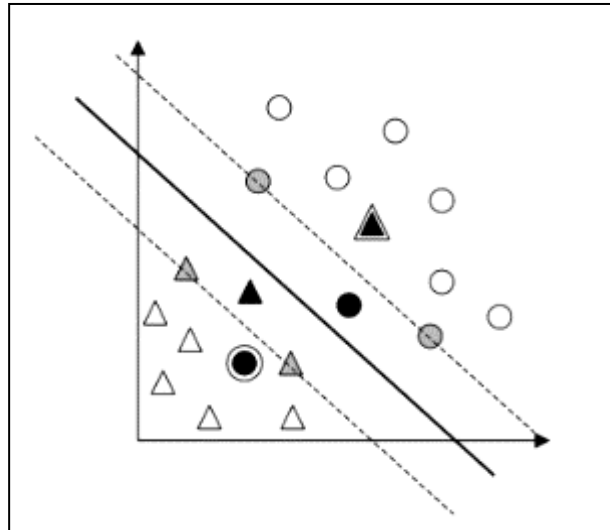


Figura 18- SVM com suavização de margens.

Fonte: Lorena e Carvalho (2007)

O Kernel polinomial é bastante utilizado em dados cuja regra de decisão possui um padrão polinomial. É descrito por (SOUZA , 2016):

$$k(x, \tilde{x}) = (\alpha x' \tilde{x} + c)^d \quad (2)$$

no qual α e c são parâmetros de ajuste e d representa o grau do polinômio.

A classificação feita pelos SVM, são baseadas na separação ótima entre classes, ou seja, considerando-se que as classes são passíveis de separação, haverá a delimitação do máximo de classes possível (NASCIMENTO et. al, 2009).

Estudos apontam que o classificador SVM tem se mostrado eficiente para classificação utilizando imagens de sensoriamento remoto, mostrando-se até mesmo superior a outros métodos já consolidados como Redes Neurais e Árvores de decisão (FOODY e MATHUR, 2004; PAL e MATHER, 2005).

2.1.9 *Weighted Subspace Random Forest*

O algoritmo *Weighted Subspace Random Forest* foi proposto no International Journal of Data Warehousing and Mining por Xu et al. (2012). O algoritmo pode classificar dados dimensionais muito altos usando florestas construídas com pequenos

subespaços. Esse novo método de ponderação é usado para a seleção de subespaços em lugar da amostragem tradicional de variável aleatória (MENG et al., 2017).

Esta nova abordagem é particularmente útil na construção de modelos a partir de dados de alta dimensão, muitas vezes constituídos por milhares de variáveis (MENG et al., 2017). A computação paralela é usada para tirar proveito de máquinas *multi-core* e *clusters* de máquinas para construir modelos de floresta aleatória com tempos reduzidos (PACKAGE WSRF, 2017).

A importância de cada variável é medida da seguinte forma: para cada árvore, é registrado o erro de predição em amostras fora do banco de dados usado no treinamento. Em seguida, o mesmo é feito após a permuta de cada variável de previsão. A partir da diferença entre os dois é feita a média de todas as árvores, e normalizada pelo desvio padrão das diferenças. A segunda medida é a diminuição total das impurezas do nó decorrentes da divisão da variável, sobre todas as árvores. A impureza do nó é medida pelo Índice de Ganho de Informação.

2.2 Principais Classes De Solo Encontradas Na Área De Estudo

Após trabalhos de campo, foram determinadas as principais classes de solos observadas na área de estudo, além do tipo de terreno predominante. Os itens a seguir é uma descrição sucinta do que foi encontrado.

2.2.1 Afloramentos rochosos com e sem vegetação

Os afloramentos de rocha são muito recorrentes na área, e por isso, apesar de não se apresentarem como uma classe de solo, tornam-se muito relevantes na bacia estudada. No local de estudo, apresentam-se, principalmente em forma de *inselbergs* em grande parte, com vegetação rasteira em sua superfície (Figura 19). Também são encontrados na forma de blocos rochosos em superfície.

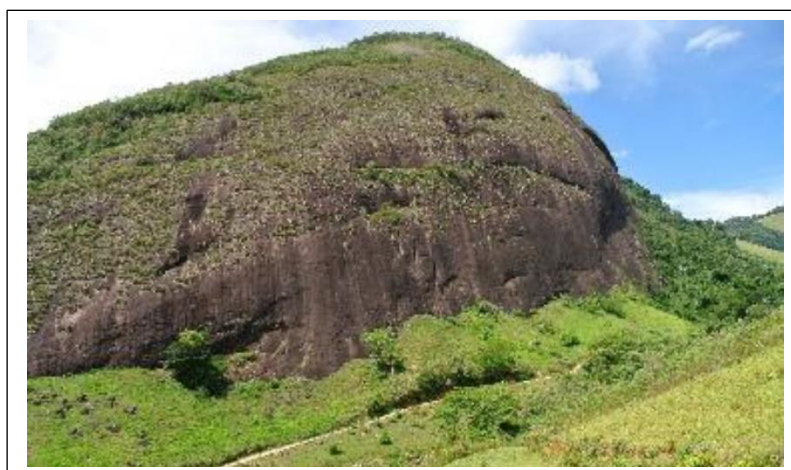


Figura 19- Afloramento rochoso na bacia do ribeirão Inhaúma.

Segundo o sistema brasileiro de classificação de solos (EMBRAPA, 2006), os Cambissolos possuem pedogênese pouco avançada, são constituídos por material

mineral, têm o horizonte B incipiente (Bi) com textura franco-arenosa ou mais argilosa. São solos que podem se formar diferentes locais, incluindo portanto, diferentes materiais de origem, formas de relevo e condições climáticas, e por isso apresentam características bastante variáveis de um local para o outro.

É a classe de solo mais comum na área de estudo e podem ocorrer associados a Neossolos Litólicos e afloramentos rochosos. Apresenta-se, em muitos casos, como latossólico (Figura 20) e estão geralmente utilizados com plantações de cafés.



Figura 20- Cambissolo Háplico latossólico com plantio consorciado de café e bananeira.

2.2.3 Gleissolos Háplico

São solos hidromórficos, ou seja, esses solos encontram-se de forma permanente ou periódica, saturado por água. Salvo se forem artificialmente drenados. O processo de gleização faz com que esses apresentem colorações acinzentadas, azuladas ou esverdeadas, devido ao ambiente redutor, que causa a redução e solubilização do ferro. Comumente desenvolvem-se a partir de sedimentos depositadas próximos aos cursos d'água, em áreas de relevo plano de terraços fluviais, marinhos ou lacustres e em áreas de depressão (EMBRAPA, 2006).

Na bacia são encontrados principalmente nos fundos dos vales, nos terraços fluviais que sofrem alagamento sazonais. Foram também observados em terraços suspensos, na parte alta da bacia (Figura 21).



Figura 21- Gleissolo Háplico em seu ambiente de ocorrência na bacia do ribeirão Inhaúma.

2.2.4 Latossolos

Os Latossolos, de maneira geral, são solos profundos, muito evoluídos e que já sofreram fortemente o processo de intemperização. Por isso, em sua constituição há poucos ou quase nenhum mineral primário ou secundário menos resistente ao intemperismo, sua capacidade de troca catiônica também é baixa, podendo comportar solos caulíníticos a oxídicos, além disso são geralmente, ácidos, com baixa saturação por bases, distróficos ou alumínicos. São recorrentes em regiões equatoriais e tropicais, podendo ser originados a partir de diferentes rochas e sedimentos, em diversos climas e tipos de vegetação (EMBRAPA, 2006).

Os Latossolos encontram-se entre as subordens Vermelho-Amarelo e Amarelo localizados em rampas suaves e topo da paisagem (Figura 22). São ocupados por pastagens, cafés e bananais.



Figura 22- Latossolo Vermelho-Amarelo.

2.2.5 Neossolo Litólico

Os Neossolos apresentam pouco desenvolvimento pedogenético, sendo considerados solos jovens, pouco evoluídos. Este é um solo em formação, devido a ação reduzida dos processos pedogenéticos, em virtude das características intrínsecas ao material de origem, como grande resistência ao intemperismo ou composição químico-mineralógica, ou até mesmo condições de formação (relevo, clima, tempo) que limita a evolução dos mesmos. Pode ser formado por material mineral ou orgânico pouco espesso, e não apresenta grandes alterações em relação ao material originário (EMBRAPA, 2006).

No local de estudo, foi observado a predominância desse tipo de solos em locais abrangidos por áreas de mata em função das restrições de uso. Geralmente apresentam afloramento rochoso associado (Figura 23).



Figura 23- Neossolo Litólico.

3 MATERIAL E MÉTODOS

O estudo foi realizado na Bacia do Ribeirão Inhaúma, na cidade de Iconha/ES. A mesma tem área de 2.403,9 ha, com coordenada central de 21°10'58.82"S e 41°00'08.87"O, localizada quase que totalmente no município de Iconha, no sul do estado do Espírito Santo.

Inicialmente, foi realizado o levantamento de solos que consistiu em 285 pontos de observação, 12 perfis completos e três amostras extras, totalizando 300 pontos (Figura

24). Em cada ponto de observação foi determinado a classe de solo, enquanto nos perfis e amostras extras foi feita a descrição morfológica (SANTOS et al., 2015) e classificação conforme Embrapa (2013), com coleta dos horizontes e posterior análise química e física.

Os perfis seguiram uma malha de amostragem determinada pelo método do hipercubo latino condicionado considerando à dificuldade de acesso, pelo relevo extremamente montanhoso da área. As coordenadas de cada perfil e amostras extras foram registradas com auxílio de receptor GNSS modelo GS09 Plus, da Leica. Os dados foram processados utilizando-se o programa Leica Geoffice 8.0, utilizando a estação fixa do Instituto Brasileiro de Geografia e Estatística-IBGE presente em Vitória/ES para transporte da base para a área da bacia.

As coordenadas dos pontos de observação foram registradas utilizando-se o GPS de navegação Garmin modelo 76 Csx. Foram amostradas todas as cinco classes de solo presentes no local de estudo e a classe com maior número de amostra foi a dos Cambissolos Háplicos, por ser a mais comum na bacia (Tabela 8).

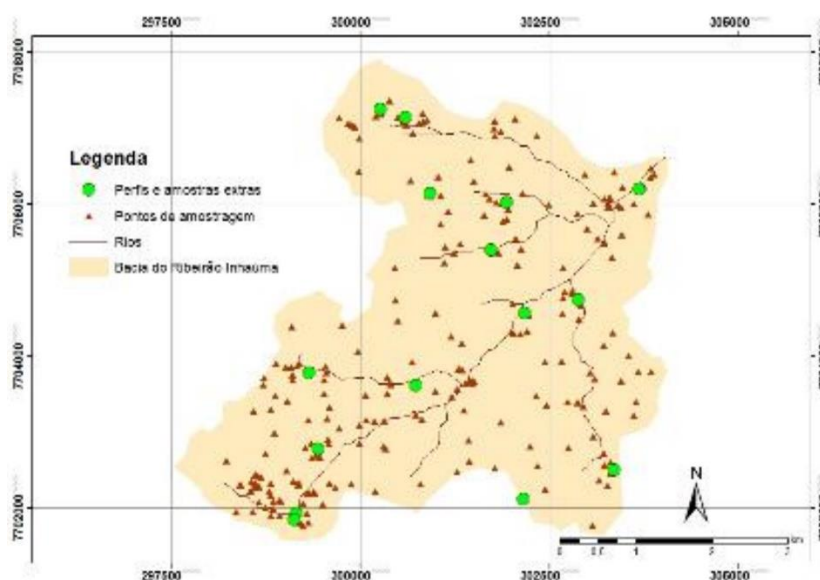


Figura 24- Localização dos perfis e pontos de observação na bacia do Ribeirão Inhaúma, Iconha/ES.

Tabela 8- Distribuição dos perfis, amostras extras e pontos de observação por classes de solos na área de estudo

Classe de solo	Número de amostras	Representatividade (%)
RL	37	12,3
CX	172	57,3
GX	29	9,7
AR	47	15,7
LVA	15	5
Total	300	100

RL- Neossolo Litólico, CX- Cambissolo Háplico, GX-GleissoloHáplico, AR- Afloramento rochoso, LVA-Latossolo Vermelho-Amarelo

Primeiramente as amostras foram secas ao ar, destorroadas e passadas em peneira com malha de 2 mm, sendo separada entre maior e menor que 2 mm, e procedida a análise granulométrica e química segundo EMBRAPA (2011). A análise química constou de determinação de Ca e Mg trocáveis, Alumínio trocável (Al^{+3}), Acidez extraível (H+Al), K e Na trocáveis, P extraível, Carbono orgânico, Valor S (soma de bases trocáveis), valor T (CTC), valor V% (porcentagem de saturação por bases) e pH.

Para predição das classes de solos foram usadas diferentes variáveis independentes. A partir do modelo digital de elevação (MDE) obtido a partir de dados do satélite ALOS (*Advanced Land Observing Satellite*) com resolução espacial de 12,5 m foram geradas 38 covariáveis morfométricas (Tabela 9), utilizando para isso script desenvolvido no programa R para aplicar as ferramentas de análise morfométricas do terreno contidas no software livre SAGA (BOHNER et al., 2006).

Tabela 9- Descrição das variáveis morfométricas utilizadas no mapeamento digital de solos da bacia do ribeirão Inhaúma, Iconha/ES (Continua)

Variáveis morfométricas	Descrição
Aquecimento anisotrópico diurno (Diurnal Anisotropic)	Representa a assimetria no aquecimento comparando as faces leste e oeste locais, baseado em horas de radiação.
Área real da superfície (Real Surface Area)	Considera a superfície real ao contrário da projetada
Classificação de curvatura (Curvature Classification)	Divide as curvaturas em 8 classes
Convexidade da superfície terrestre (Terrain Surface Convexity)	Razão entre o número de células que tem curvatura positiva (células convexas) para o número de todas as células válidas dentro de um raio de pesquisa específico
Curvatura da linha de fluxo (Flow Line Curvature)	Curvatura da linha de fluxo
Curvatura da seção transversal (Cross-sectional Curvature)	Curvatura relativa ao plano vertical das diferenças de nível
Curvatura de perfil (Profile Curvature)	Descreve o segundo mecanismo de acumulação
Curvatura Geral (General Curvature)	Curvatura geral
Curvatura Longitudinal (Longitudinal Curvature)	Curvatura longitudinal
Curvatura máxima (Maximal curvature)	Referente a seção normal local
Curvatura mínima (Minimal curvature)	Referente a seção normal local
Curvatura plana (Plan curvature)	Refere-se a curvatura plana do local
Curvatura tangencial (Tangential Curvature)	Descreve o primeiro mecanismo de acumulação
Curvatura total (Total Curvature)	Curvatura média menos a diferença de curvatura
Declividade (Slope)	Representa a declividade local que pode ser angular ou em porcentagem
Duração da radiação 1 (Duration of Insolation 1)	Duração da radiação solar no mês de janeiro
Duração da radiação 2 (Duration of Insolation 2)	Duração da radiação solar no mês de junho
Elevação da declividade (Slope Height)	Altura total do declive
Elevação normalizada (Normalized Height)	É uma medida de altura relativa de um ponto ao invés de seu valor elevação propriamente dito. Atribui o valor 1 para o ponto mais alto e o valor 0 para a posição mais baixa dentro de uma respectiva área de busca
Elevação Padronizada (Standardized Height)	Produto da altura normalizada multiplicada pela altura absoluta
Formas de relevo (Landforms)	Representa as formas de relevo da área
Gradiente (Gradient)	Corresponde ao gradiente hidráulico
Índice de Balanço de Massa (Mass Balance)	Índice que representa o balanço de massa em cada pixel
Índice de convergência (Convergence index)	Calcula o índice de convergência/divergência em relação ao escoamento superficial
Índice de declividade (Slope index)	
Índice dos vales (Valley index)	Refere-se as depressões do terreno

Tabela 9- Continuação...

Índice de umidade topográfica (Topographic Wetness Index)	Tendência de cada célula em acumular água.
Modelo Digital de Elevação	Representa a elevação de cada célula do modelo
Posição de Meia Encosta (Mid Slope Position)	Representa a distância em relação ao topo e vale, variando entre 0 e 1 cobrindo as regiões de maior declividade
Profundidade dos vales (Valley Depth)	Inverte a elevação, deriva as redes de drenagem e calcula a distância vertical até os mesmos
Radiação solar difusa 1 (Diffuse Insolation 1)	Radiação solar difusa incidente no mês de janeiro
Radiação solar difusa 2 (Diffuse Insolation 2)	Radiação solar difusa incidente no mês de junho
Radiação solar direta 1 (Direct to Diffuse Ratio 1)	Radiação solar direta incidente no mês de janeiro
Radiação solar direta 2 (Direct to Diffuse Ratio 2)	Radiação solar indireta incidente no mês de junho
Radiação solar total 1 (Total Insolation 1)	Radiação solar total (direta mais difusa) incidente no mês de janeiro
Radiação solar total 2 (Total Insolation 2)	Radiação solar total (direta mais difusa) incidente no mês de junho
Superfície de pontos específicos (Surface Specific Points)	Diferença entre pontos específicos de mudança da superfície, resultando em valores discretos positivos e negativos
Textura da superfície terrestre (Terrain Surface Texture)	Divide a textura da superfície em 8, 12 ou 16 classes

Fonte: Adaptado de Abreu (2017)

Ainda foram utilizadas como variáveis independentes as bandas do satélite Landsat 8, o NDVI (*normalized difference vegetation index*) gerado através da equação proposta por Rouse et al. (1973), onde é expresso pela razão entre a diferença da medida da reflectância nos canais do infravermelho próximo e vermelho e a soma desses canais, sendo então:

$$NDVI = \frac{\rho_{IV} - \rho_V}{\rho_{IV} + \rho_V} \quad (3)$$

Onde: NDVI é o Índice de vegetação da diferença normalizada, ρ_{IV} é refletividade no infravermelho próximo e ρ_V é a refletividade no vermelho.

Foram ainda empregados dados gamaespectrométricos e magnométricos obtidos através da Companhia de Pesquisa de Recursos Minerais -CPRM (Tabela 10).

Tabela 10- Relação dos dados gamaespectrométricos e magnométricos, utilizados como parte das variáveis independentes para predição de classes de solos

Dados gamaespectrométricos e magnométricos
Campo Magnético Total
Contagem Total
Potássio
Primeira Derivada do Campo Magnético Total
Razão entre Tório e Potássio
Razão entre Urânio e Potássio
Razão entre Urânio e Tório
Sinal Analítico do Campo Magnético Total
Taxa de Exposição da Contagem Total
Tório
Urânio

Estes dados foram submetidos previamente a processamentos pela própria CPRM (2010), sendo estes processos transcritos no anexo 1.

Por último foi utilizado também o mapa de uso e ocupação do solo, simplificado em 4 classes: Floresta, pastagem, afloramento rochoso e agricultura. Este foi produzido utilizando o software ArcGis 10.2 a partir da observação de elementos básicos de interpretação, como: tonalidade e cor, forma e tamanho, padrão, textura, sombra em imagens ortorretificadas da área, com resolução de 25 cm, cedidas pelo Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural-INCAPER, por meio do Sistema Integrado de Bases Geoespaciais do Estado do Espírito Santo- GEOBASES. Todos os mapas foram padronizados com o mesmo tamanho de células, linhas, colunas e resolução de 30 m.

Foram testados dez classificadores, sendo eles: *Bagged AdaBoost*, *C5.0*, *Conditional Inference Random Forest*, *Extreme Gradient Boosting*, *Multivariate Adaptive Regression Splines*, *Random Forest*, *Ranger*, *Support Vector Machine com Kernel Polynomial*, *Support Vector Machines with Linear Kernel*, *Weighted Subspace Random Forest*. Para a classificação foi utilizado o software R, sendo método e o respectivo pacote R utilizado mostrado na Tabela 11.

Tabela 11- Classificadores utilizados para mapeamento de solos na bacia do Ribeirão Inhaúma, Iconha/ES, com seus respectivos métodos e bibliotecas do R

Classificador	Método	Biblioteca do R
Bagged AdaBoost	AdaBag	adabag, plyr
C5.0	C5.0	C50, plyr
Conditional Inference Random Forest	cforest	Party
Extreme Gradient Boosting	xgbTree	xgboost, plyr
Multivariate Adaptive Regression Splines	earth	Earth
Random Forest	rf	randomForest
Ranger	ranger	e1071, ranger
Support Vector Machine com Kernel Polynomial	svmPoly	Kernlab
Support Vector Machines with Linear Kernel	svmLinear	Kernlab
Weighted Subspace Random Forest	wsrf	randomForest

Inicialmente para a seleção das variáveis independentes mais importantes, foram eliminadas àquelas que possuíam entre si índice de correlação maior que 0,95. Posteriormente cada modelo selecionou as variáveis mais significativas para prever as classes de solo, sendo essas usadas para gerar os mapas finais. O desempenho do modelo foi avaliado pelo método de amostragem *bootstrap* 632, através dos valores índice Kappa.

O Coeficiente Kappa foi obtido segundo a equação:

$$K = \frac{\sum_{i=1}^c \chi_{ii} - \sum_{i=1}^c (\chi_{i+} * \chi_{+i})}{n^2 - \sum_{i=1}^c (\chi_{i+} * \chi_{+i})} \quad (4)$$

em que, K é uma estimativa do coeficiente Kappa; χ_{ii} é o valor na linha i e coluna i; linhas totais χ_{i+} é a soma de linhas i e colunas totais; χ_{+i} é a soma das colunas i da matriz de confusão; n é o número total de amostras e c o número total de classes.

O índice Kappa varia de 0 a 1, sendo que 0 indica a ausência de concordância, ou seja, a estimativa e a referência são estatisticamente independentes (ELNAGGAR, 2007), e 1 indica total concordância. Os valores de Kappa foram interpretados segundo a Tabela 12 (LANDIS e KOCHI (1977)).

Tabela 12- Intervalos de Kappa com respectiva interpretação de desempenho

Kappa	Desempenho
< 0,00	Péssimo
0,00-0,20	Ruim
0,21-0,40	Regular
0,41-0,60	Bom
0,61-0,80	Muito bom
0,81-1,00	Excelente

Fonte: Adaptado de Landis & Kochi (1977).

Os valores de índice Kappa foram submetidos ao teste Z, para verificar se houve diferença estatística significativa entre eles (FOODY, 2004). Segundo Congalton e Mead (1983), a comparação entre classificadores pode ser feita pelo teste Z, sendo considerado estatisticamente significativo a 5% de probabilidade valores de Z superiores a 1,96 e 1,65 a 10 % de probabilidade (valor tabelado de Z a 5 e 10 %, respectivamente). Dessa forma, se o valor Z calculado for superior ao valor Z tabelado, diz-se que o resultado foi significativo e rejeita-se a hipótese nula (Ho: K1=K2), e portanto considera-se que os dois classificadores são estatisticamente diferentes.

O teste Z comparou os coeficientes Kappa entre os classificadores e é obtido pela seguinte equação (CONGALTON e MEAD, 1983):

$$Z = \frac{\hat{k}_1 - \hat{k}_2}{\sqrt{\hat{\sigma}_1 + \hat{\sigma}_2}} \quad (5)$$

Onde: Z = valor Z calculado;

\hat{k}_1 = coeficiente Kappa do classificador 1;

\hat{k}_2 = coeficiente Kappa do classificador 2;

$\hat{\sigma}_1$ = variância de Kappa do classificador 1;

$\hat{\sigma}_2$ = variância de Kappa do classificador 2

A variância do índice Kappa pode ser estimada por:

$$\sigma(\hat{k}) = \frac{1}{n} \left[\frac{\theta_1 (1 - \theta_1)}{(1 - \theta_2)^2} + \frac{2(1 - \theta_1)(2\theta_1\theta_2 - \theta_3)}{(1 - \theta_2)^3} + \frac{(1 - \theta_1)^2(\theta_4 - 4\theta_2^2)}{(1 - \theta_2)^4} \right] \quad (6)$$

Em que:

$$\theta_1 = \frac{1}{n} \sum_{i=1}^c x_{ii}$$

$$\theta_2 = \frac{1}{n^2} \sum_{i=1}^c x_{1+} x_{+1}$$

$$\theta_3 = \frac{1}{n^2} \sum_{i=1}^c x_{ii}(x_{i+} + x_{+i})$$

$$\theta_4 = \frac{1}{n^3} \sum_{i=1}^c \sum_{j=1}^c x_{ij}(x_{j+} + x_{+i})^2$$

Foram também determinadas a acurácia do produtor e do usuário. A acurácia do produtor indica a probabilidade de um pixel de referência ter sido classificado corretamente. É também chamada de erro de omissão, que refere-se a omissão da ocorrência de determinada classe em uma área, quando ela ocorre ali de fato. Enquanto a acurácia do usuário, indica a probabilidade de um pixel classificado no mapa realmente representar essa categoria no campo, ou seja, é também chamada erros de comissão, que referem-se a ocorrência de um tipo de solo em um local onde ele não ocorre verdadeiramente (CONGALTON e GREEN, 2009).

A acurácia do produtor é obtida dividindo-se o número de pixels corretos em dada categoria pelo número total de pixels coluna correspondente, enquanto a acurácia do usuário é determinada dividindo-se o número de pixels corretos em dada categoria pelo número total de pixels linha correspondente.

Os mapas finais apresentados, foram processados utilizando-se a extensão ArcSIE no programa ArcGIS 10.3, para remoção de área mapeadas inferiores a 10 ha, eliminando-se, dessa forma, todos os ruídos presentes. O sistema de Projeção Cartográfica utilizado foi o Universal Transversa de Mercator (UTM), Datum SIRGAS 2000 e Fuso 24 S.

4 RESULTADOS E DISCUSSÃO

Inicialmente foram removidas da análise nove variáveis por apresentarem correlação igual ou superior a 0,95. Sendo elas: as bandas espectrais do Landsat-8 (banda 1, banda 2 e, banda 3), Curvatura da seção transversal, Curvatura Longitudinal, Índice de Balanço de Massa, Radiação solar difusa 2, Radiação solar total 1 e Radiação solar total 2.

Os solos mapeados na área foram: Cambissolo Háplico, Neossolo Litólico, Gleissolo Háplico, Latossolo Vermelho-Amarelo e Afloramento Rochoso. Organossolo Háplico foi encontrado em duas amostras coletadas no campo, porém por serem extremamente pontuais e, portanto, pouco representativas da área, foram tratados como inclusão aos gleissolos.

Os valores de Kappa encontrados (Tabela 13), variaram de 0,54 a 0,65, sendo classificados segundo Landis e Kochi (1977) como desempenhos bons e muito bons. Os altos valores encontrados nesse trabalho podem estar relacionados ao grande número de pontos coletados que conseguiram representar de forma eficiente a área de estudo. Abreu (2017) mostrou em seu estudo que uma maior densidade de pontos de amostragem proporcionou melhor mapeamento de sua área, com índices Kappa superiores.

Os dados de amostra existentes são importantes para o mapeamento digital do solo; diferentes pontos de amostra possuem representatividade diferente e a representatividade das mesmas influencia o mapeamento do solo (AN et al., 2017). Além disso, Samuel-Rosa et al. (2015) apontam que um dos fatores limitantes no mapeamento é muitas vezes o conjunto escasso de dados armazenados do solo que são usados para a calibração do modelo.

Os maiores valores de Kappa foram encontrados para RF (0,65), XGBTREE (0,65), RANGER (0,64) e WSRF (0,63). Zeraatpisheh et al. (2017), também encontraram que RF foi superior aos outros modelos testados em seu trabalho, para mapeamento de solos em nível de grande grupo e subgrupos.

Abreu (2017) trabalhando com mapeamento digital de solos em uma área no sul do estado da Bahia, testou diversos classificadores dentre eles XGBTREE, RANGER, C5.0, RF, CIRF, MARS, ADABAG e SVM, e constatou que XGBTREE foi superior para o delineamento das classes de solo em sua área com Kappa de 0,68.

Brungard et al. (2015) testaram 11 algoritmos para mapear diferentes regiões do semi-árido dos Estados Unidos da América, dentre eles análise discriminante, regressão logística multinomial, redes neurais, métodos baseados em árvores e classificadores de máquinas vetoriais de suporte. O RF usando covariáveis selecionadas por eliminação de características recursivas, foram consistentemente as mais precisas, ou estiveram entre os mais precisas.

Taghizadeh-Mehrjardi et al. (2015) avaliaram seis classificadores, dentre eles RF e SVM, encontrando Kappas de 0,69 e 0,51 para os mesmos, respectivamente. Dias (2015) usando Random Forest para mapeamento digital de solos, encontrou índice Kappa de 0,38. Pahlavan-Rad et al. (2016) também acharam RF como classificador superior aos demais testados em todos os níveis taxonômicos de mapeamento de solos no Irã.

O sucesso de RF está relacionado a várias vantagens em relação a outras estatísticas (CAMERA et al. 2017), como por exemplo: capacidade de modelar relacionamentos não lineares de alta dimensão; tratamento simultâneo de preditores categóricos e contínuos; robustez contra sobreposição; medidas de taxa de erro e importância da variável e relativamente baixa sensibilidade aos valores dos parâmetros (GRIMM et al., 2008).

Tabela 13- Desempenho dos classificadores segundo o índice Kappa

Método	Kappa	Desempenho
ADABAG	0,57	Bom
C50	0,61	Muito bom
CIRF	0,57	Bom
MARS	0,54	Bom
RF	0,65	Muito bom
RANGER	0,64	Muito bom
SVMLINEAR	0,55	Bom
SVMPOLY	0,58	Bom
WSRF	0,63	Muito bom
XGBTREE	0,65	Muito bom

Bagged AdaBoost-ADABAG, C5.0 -C50, Conditional Inference Random Forest-CIRF, Extreme Gradient Boosting-XGBTREE, Multivariate Adaptive Regression Splines-MARS, Random Forest-RF, Suport Vector Machine com Kernel Polynomial-SVMPOLY, Suport Vector Machines with Linear Kernel-SVMLINEAR, Weighted Subspace e Random Forest-WSRF.

As Tabelas 14 a 23 mostram as matrizes de confusão geradas por cada classificador. Será discutido aqui aqueles métodos que tiveram índice Kappa maior ou igual a 0,63 (RF (0,65), XGBTREE (0,65), RANGER (0,64) e WSRF (0,63)).

As classes que obtiveram maior acurácia tanto do produtor quanto do usuário foram Cambissolo e Afloramento rochoso. Em relação a acurácia do produtor (AP), Cambissolo obteve os percentuais de 82, 85, 83 e 82 para WSRF, XGBTREE, RF e RANGER respectivamente, enquanto afloramento rochoso ficou com 85, 79, 83 e 81 para os mesmos classificadores.

Latossolo Vermelho-Amarelo e Gleissolo Háptico obtiveram os menores percentuais de acurácia do produtor, sendo igual a 7, 27, 27 e 20 para LVA em WSRF, XGBTREE, RF e RANGER respectivamente e 17, 21, 28 e 17 para GX nesses mesmos classificadores. A acurácia do produtor representa o número de acertos de uma classe dividida pelo total de amostras dessa classe (CONGALTON, 1991), ou seja, refere-se à probabilidade de uma categoria ser corretamente classificada de acordo com os pontos de referência (ROSENFELD et al., 1986) estando relacionada aos erros de omissão.

Em relação a acurácia do usuário (AU), esta também pode ser considerada a confiabilidade dos testes, e representa a probabilidade de um ponto no mapa representar a realidade encontrada no campo (CONGALTON e REKAS, 1985), estando relacionada aos erros de comissão. O comportamento foi semelhante ao encontrado para acurácia do produtor, sendo as classes melhores preditas os afloramentos rochosos e Cambissolos Hápticos e os menores percentuais encontrados para Gleissolo Háptico e Latossolo Vermelho-Amarelo.

Em WSRF, XGBTREE, RF e RANGER, os percentuais encontrados para afloramento rochoso foram de 89, 84, 83 e 84 e Cambissolos foram 73, 74, 73 e 71 respectivamente. Os menores percentuais foram ligeiramente superiores àqueles determinados na acurácia do produtor, sendo para Latossolo Vermelho-Amarelo igual a 20, 57, 44 e 43 e para Gleissolo Háptico 26, 28, 35 e 26 para os mesmos classificadores.

Os menores valores encontrados para LVA e GX tanto na acurácia do produtor quanto do usuário podem estar relacionados ao menor número de amostras coletadas nessas classes, porém estas representam a realidade local onde a maior parte dos solos encontrados são Cambissolos, perfazendo estes mais da metade da área estudada. Cada ponto de amostra do solo, refere-se ao produto da complexa interação entre os fatores de formação do solo, refletindo a relação *in situ* do solo com o seu nicho ambiental (ZHU et al., 2015).

As matrizes de confusão mostram que os maiores erros na classificação se deram para GX e LVA, que tiveram grande parte de suas amostras classificadas como CX. Supõe-se que isso ocorreu devido a área ser composta em sua maior parte por CX, tendo apenas locais pontuais com outras classes, sendo que GX e LVA são aqueles que possuem menores áreas de distribuição quando comparadas as demais.

RL apresentou pequena confusão com CX, isso pode ser explicado pela ocorrência sempre próxima das duas classes na área de estudo. Os maiores acertos se deram para CX e AR, isso provavelmente se deu pela ampla ocorrência da primeira e pela especificidade da segunda, sendo por isso, de certa forma mais fácil de se classificar.

Tabela 14- Matriz de confusão gerada pela classificação segundo o método Adabag

Classe de solo	RL	CX	GX	AR	LVA	Σ Linhas	AU (%)
RL	20	10	0	2	0	32	63
CX	17	150	24	7	15	212	71
GX	0	6	4	0	0	10	37
AR	0	5	2	38	0	44	86
LVA	0	2	0	0	0	2	0
Σ Colunas	37	172	29	47	15	300	
AP (%)	54	87	14	81	0		

RL- Neossolo Litólico, CX- Cambissolo Háplico, GX-GleissoloHáplico, AR-Afloramento rochoso, LVA-Latossolo Vermelho-Amarelo, AU- Acurácia do usuário, AP-Acurácia do produtor.

Tabela 15- Matriz de confusão gerada pela classificação segundo o método MARS

Classe de solo	RL	CX	GX	AR	LVA	Σ Linhas	AU (%)
RL	22	11	0	2	0	35	62
CX	14	142	21	7	12	195	73
GX	0	10	6	1	0	18	32
AR	0	4	2	37	0	43	86
LVA	1	5	0	0	3	9	30
Σ Colunas	37	172	29	47	15	300	
AP (%)	59	83	21	79	20		

RL- Neossolo Litólico, CX- Cambissolo Háplico, GX-GleissoloHáplico, AR-Afloramento rochoso, LVA-Latossolo Vermelho-Amarelo, AU- Acurácia do usuário, AP-Acurácia do produtor.

Tabela 16- Matriz de confusão gerada pela classificação segundo o método CIRF

Classe de solo	RL	CX	GX	AR	LVA	Σ Linhas	AU (%)
RL	24	10	0	2	0	36	67
CX	13	153	24	6	15	210	73
GX	0	3	3	0	0	6	49
AR	1	5	2	40	0	47	85
LVA	0	1	0	0	0	1	0
Σ Colunas	37	172	29	47	15	300	
AP (%)	65	89	10	85	0		

RL- Neossolo Litólico, CX- Cambissolo Háplico, GX-GleissoloHáplico, AR-Afloramento rochoso, LVA-Latossolo Vermelho-Amarelo, AU- Acurácia do usuário, AP-Acurácia do produtor.

Tabela 17- Matriz de confusão gerada pela classificação segundo o método C5.0

Classe de solo	RL	CX	GX	AR	LVA	Σ	AU
						Linhas	(%)
RL	20	11	0	2	1	34	60
CX	16	139	19	6	11	191	73
GX	0	12	9	1	1	22	39
AR	1	5	2	39	0	46	84
LVA	0	4	0	0	2	7	33
Σ Colunas	37	172	29	47	15	300	
AP (%)	55	81	31	82	13		

RL- Neossolo Litólico, CX- Cambissolo Háplico, GX-GleissoloHáplico, AR- Afloramento rochoso, LVA-Latossolo Vermelho-Amarelo, AU- Acurácia do usuário, AP-Acurácia do produtor.

Tabela 18- Matriz de confusão gerada pela classificação segundo o método SVM Linear

Classe de solo	RL	CX	GX	AR	LVA	Σ	AU
						Linhas	(%)
RL	22	13	0	3	0	38	57
CX	13	140	21	8	10	192	73
GX	0	10	7	0	0	17	39
AR	2	4	1	36	0	43	83
LVA	0	5	0	0	5	10	46
Σ Colunas	37	172	29	47	15	300	
AP (%)	59	81	24	77	33		

RL- Neossolo Litólico, CX- Cambissolo Háplico, GX-GleissoloHáplico, AR- Afloramento rochoso, LVA-Latossolo Vermelho-Amarelo, AU- Acurácia do usuário, AP-Acurácia do produtor.

Tabela 19- Matriz de confusão gerada pela classificação segundo o método SVM Polinomial

Classe de solo	RL	CX	GX	AR	LVA	Σ	AU
						Linhas	(%)
RL	19	11	0	2	0	32	60
CX	16	141	21	9	11	198	71
GX	0	8	7	0	0	16	44
AR	2	6	1	36	0	44	82
LVA	0	6	0	0	4	10	40
Σ Colunas	37	172	29	47	15	300	
AP (%)	51	82	24	77	27		

RL- Neossolo Litólico, CX- Cambissolo Háplico, GX-GleissoloHáplico, AR- Afloramento rochoso, LVA-Latossolo Vermelho-Amarelo, AU- Acurácia do usuário, AP-Acurácia do produtor.

Tabela 20- Matriz de confusão gerada pela classificação segundo o método WSRF

Classe de solo	RL	CX	GX	AR	LVA	Σ	AU
						Linhas	(%)
RL	25	11	0	2	1	39	64
CX	12	141	22	5	12	192	73
GX	0	13	5	0	1	19	26
AR	0	3	2	40	0	45	89
LVA	0	4	0	0	1	5	20
Σ Colunas	37	172	29	47	15	300	
AP (%)	68	82	17	85	7		

RL- Neossolo Litólico, CX- Cambissolo Háplico, GX-GleissoloHáplico, AR-Afloramento rochoso, LVA-Latossolo Vermelho-Amarelo, AU- Acurácia do usuário, AP-Acurácia do produtor.

Tabela 21- Matriz de confusão gerada pela classificação segundo o método Xgbtree

Classe de solo	RL	CX	GX	AR	LVA	Σ	AU
						Linhas	(%)
RL	21	10	1	2	0	34	62
CX	14	147	21	7	10	199	74
GX	0	8	6	0	1	16	28
AR	1	4	1	37	0	44	84
LVA	1	3	0	0	4	7	57
Σ Colunas	37	172	29	47	15	300	
AP (%)	57	85	21	79	27		

RL- Neossolo Litólico, CX- Cambissolo Háplico, GX-GleissoloHáplico, AR-Afloramento rochoso, LVA-Latossolo Vermelho-Amarelo, AU- Acurácia do usuário, AP-Acurácia do produtor.

Tabela 22- Matriz de confusão gerada pela classificação segundo o método Random Forest

Classe de solo	RL	CX	GX	AR	LVA	Σ	AU
						Linhas	(%)
RL	20	11	0	1	1	33	61
CX	17	138	19	7	9	190	73
GX	0	14	8	0	1	23	35
AR	0	4	2	39	0	45	87
LVA	0	5	0	0	4	9	44
Σ Colunas	37	172	29	47	15	300	
AP (%)	54	80	28	83	27		

RL- Neossolo Litólico, CX- Cambissolo Háplico, GX-GleissoloHáplico, AR-Afloramento rochoso, LVA-Latossolo Vermelho-Amarelo, AU- Acurácia do usuário, AP-Acurácia do produtor.

Tabela 23- Matriz de confusão gerada pela classificação segundo o método Ranger

Classe de solo	RL	CX	GX	AR	LVA	Σ	AU
						Linhas	(%)
RL	21	9	0	1	0	31	68
CX	16	141	22	8	11	198	71
GX	0	13	5	0	1	19	26
AR	0	5	2	38	0	45	84
LVA	0	4	0	0	3	7	43
Σ Colunas	37	172	29	47	15	300	
AP (%)	57	82	17	81	20		

RL- Neossolo Litólico, CX- Cambissolo Háplico, GX-GleissoloHáplico, AR- Afloramento rochoso, LVA-Latossolo Vermelho-Amarelo, AU- Acurácia do usuário, AP-Acurácia do produtor.

Quanto a área definida por RF, Ranger, XGBTREE e WSRF (Tabela 24), não nota-se grande disparidade principalmente para a área de RL. WSRF classificou uma área ligeiramente superior para CX, enquanto XGBTREE determinou maiores áreas em relação aos outros métodos para GX, AR e LVA.

As maiores diferenças são encontradas em AR e LVA. AR apresentou área máxima de 431,2 ha (XGBTREE) e mínima de 319,5 ha (WSRF), enquanto LVA exibiu máximo de 55,7 ha (XGBTREE) e mínimo de 20,3 ha (WSRF). Mostrando que esses algoritmos poderiam estar super e subestimando essas áreas, respectivamente.

Tabela 24- Áreas definidas para cada classe pelos classificadores Random forest, Ranger, XGBTREE, WSRF

Classificador	Classe de solo				
	RL	CX	GX	AR	LVA
	(ha)				
Random Forest	424,9	1570,7	28,5	350,0	26,4
Ranger	418,5	1583,2	25,6	331,4	41,9
XGBTREE	421,5	1458,8	33,6	431,2	55,7
WSRF	425,7	1611,4	23,6	319,5	20,3

RL- Neossolo Litólico, CX- Cambissolo Háplico, GX-GleissoloHáplico, AR- Afloramento rochoso, LVA-Latossolo Vermelho-Amarelo

A Tabela 25 mostra as variáveis mais importantes na predição das classes para os métodos RF, XGBTREE e RANGER. O uso do solo mostrou-se como a variável mais importante para os três métodos. Em seguida apareceram bandas do Landsat 8, seguidas por covariáveis do terreno e dados de magnetometria e gamaespectrometria.

As bandas que se destacaram nos três métodos foram b4 e b7, tendo também b6 sido considerada em RF e XGBTREE. b4 corresponde a região do vermelho enquanto b6

e b7 a região do infravermelho de ondas curtas. Cada um dos classificadores identificou como principais variáveis, 11 covariáveis do terreno e cinco referentes aos dados de magnetometria e gamaespectrometria da região. Os três métodos empregados selecionaram as mesmas covariáveis como mais importantes, tendo RF e XGBTREE determinado 22 variáveis principais e Ranger 20.

Já é conhecido na literatura a capacidade dos atributos do terreno derivados do MDE predizerem propriedades e classes do solo (CARVALHO JÚNIOR et al., 2011; CHAGAS et al., 2010; HANSEN et al., 2009), já que o relevo é um dos fatores de formação do solo mais importantes (JENNY, 1941). Behrens et al. (2010) utilizaram 19 atributos do terreno para mapeamento de classes de solo em uma região da Alemanha, enquanto Ten Caten et al. (2011) utilizaram nove covariáveis do terreno para mapear solos de um município no Rio Grande do Sul.

Chagas et al. (2011) encontraram que a declividade foi a covariável mais importante para o delineamento dos dados em sua área de estudo, tendo maior contribuição na melhora do desempenho da classificação.

Estudos ambientais e pedológicos têm utilizado-se de dados de magnetometria e gamaespectrometria com sucesso (SILVA et al., 2010; BECEGATO e FERREIRA, 2005; DICKSON e SCOTT, 1997), sendo essas variáveis a cada dia mais difundidas nas duas últimas décadas (COSTA, 2013).

Costa (2013) cita que estes dados são uma ferramenta muito útil para o estudo dos solos, pois os raios gama são emitidos nas camadas mais superficiais da crosta terrestre, ou seja, nos primeiros 30 cm onde ocorrem de forma mais intensa os processos pedogenéticos, compreendendo assim os horizontes A e B do solo.

Tabela 25- Ordem de importância das variáveis utilizadas pelos classificadores *Randon Forest*, *Extreme Gradient Boosting* e *Ranger* na predição das classes de solo no Ribeirão Inhaúma, Iconha/ES

Variáveis		
RF	XGBTREE	RANGER
Uso do solo (classe 3)	Uso do solo (classe 3)	Uso do solo
b4	b4	b4
b7	Convergence index	b7
Slope Height	b6	Slope Height
Convergence index	Real Surface Area	Razão entre Urânio e Potássio
Uso do solo (classe 2)	Razão entre Urânio e Potássio	Real Surface Area
b6	Modelo digital de elevação	Slope
Sinal analítico do Campo Magnético Total	Sinal analítico do Campo Magnético Total	b6
Razão entre Urânio e Potássio	Taxa de Exposição da Contagem Total	Potássio
Slope	Direct to Diffuse Ratio 1	Standardized Height
Razão entre Tório e Potássio	Topographic Wetness Index	Modelo digital de elevação
Real Surface Area	Razão entre Tório e Potássio	Razão entre Tório e Potássio
Taxa de Exposição da Contagem Total	Slope Height	Normalized Height
Diffuse Insolation 1	Standardized Height	Taxa de Exposição da Contagem Total
Standardized Height	Potássio	Sinal analítico do Campo Magnético Total
Topographic Wetness Index	Diffuse Insolation 1	Convergence index
Potássio	b7	Topographic Wetness Index
Modelo digital de elevação	Direct to Diffuse Ratio 2	Direct to Diffuse Ratio 2
Normalized Height	Uso do solo (classe 2)	Direct to Diffuse Ratio 1
Direct to Diffuse Ratio 1	Normalized Height	Diffuse Insolation 1
Direct to Diffuse Ratio 2	Slope	
Uso do solo (classe 4)	Uso do solo (classe 4)	

O teste Z (Tabela 27) mostra a diferença estatística entre os classificadores. Ela demonstra que RF e XGBTREE foram superiores em relação a MARS, porém mostraram-se igualmente eficientes aos demais. No entanto, é importante uma avaliação completa, levando-se em consideração, por exemplo, o índice Kappa, já que estes apresentaram valores variando entre 0,55 e 0,65, excetuando-se o valor obtido por MARS de 0,54.

Tabela 26- Teste Z para os coeficientes Kappa de 10 métodos de classificação usados no mapeamento digital de solos do Ribeirão Inhaúma, Iconha/ES

	ADABAG	C5.0	CIRF	MARS	RF	RANGER	SVM LINEAR	SVM POLY	WSRF	XGBTREE
ADABAG	--	0,64	0,00	0,48	1,28	1,11	0,32	0,15	0,96	1,28
C5.0		--	0,64	1,12	0,64	0,48	0,96	0,48	0,32	0,64
CIRF			--	0,48	1,29	1,12	0,32	0,16	0,97	1,29
MARS				--	1,76*	1,59	0,16	0,64	1,45	1,77*
RF					--	0,16	1,60	1,58	0,32	0,00
RANGER						--	1,43	1,34	0,16	0,16
SVM LINEAR							--	0,48	1,29	1,61
SVM POLY								--	0,80	1,60
WSRF									--	0,32
XGBTREE										--

Bagged AdaBoost-ADABAG, C5.0 -C50, Conditional Inference Random Forest-CIRF, Extreme Gradient Boosting-XGBTREE, Multivariate Adaptive Regression Splines-MARS, Random Forest-RF, Suport Vector Machine com Kernel Polynomial-SVMPOLY, Suport Vector Machines with Linear Kernel-SVMLINEAR, Weighted Subspace e Random Forest-WSRF.

*Significativo a 10 % pelo teste Z.

Os mapas produzidos pelos métodos RF, Ranger, XGBTREE e WSFR são mostrados nas figuras 25, 26, 27 e 28.

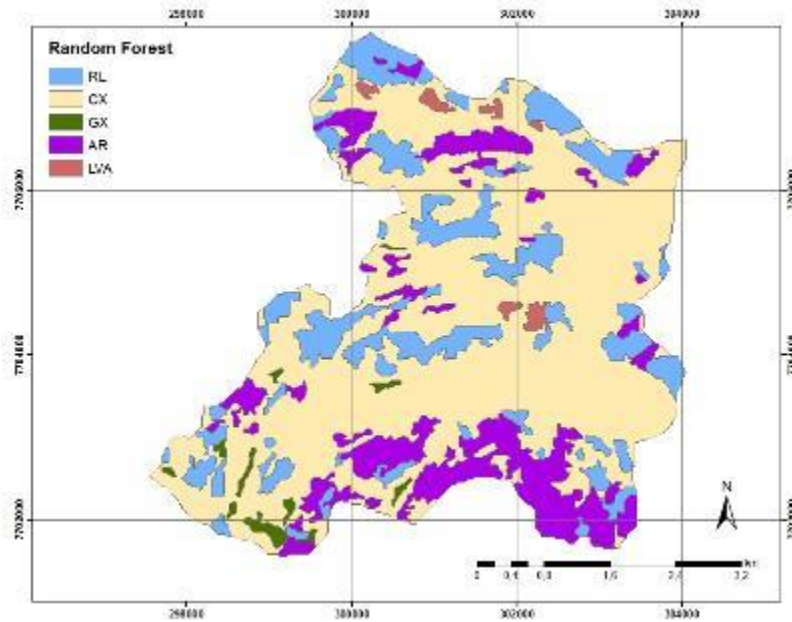


Figura 25- Classificação do solo segundo o método *Random Forest*.

RL- Neossolo Litólico, CX- Cambissolo Háplico, GX-GleissoloHáplico, AR- Afloramento rochoso, LVA-Latossolo Vermelho-Amarelo

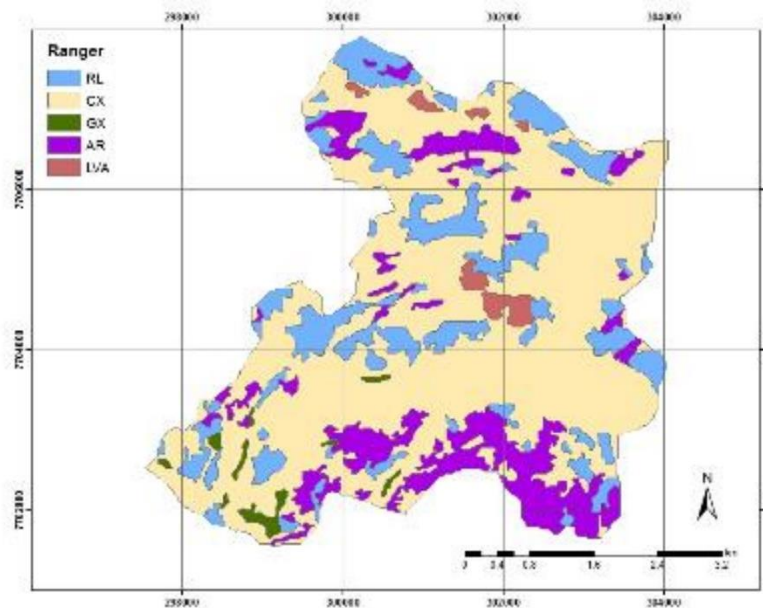


Figura 26- Classificação do solo segundo o método *Ranger*.

RL- Neossolo Litólico, CX- Cambissolo Háplico, GX-GleissoloHáplico, AR- Afloramento rochoso, LVA-Latossolo Vermelho-Amarelo

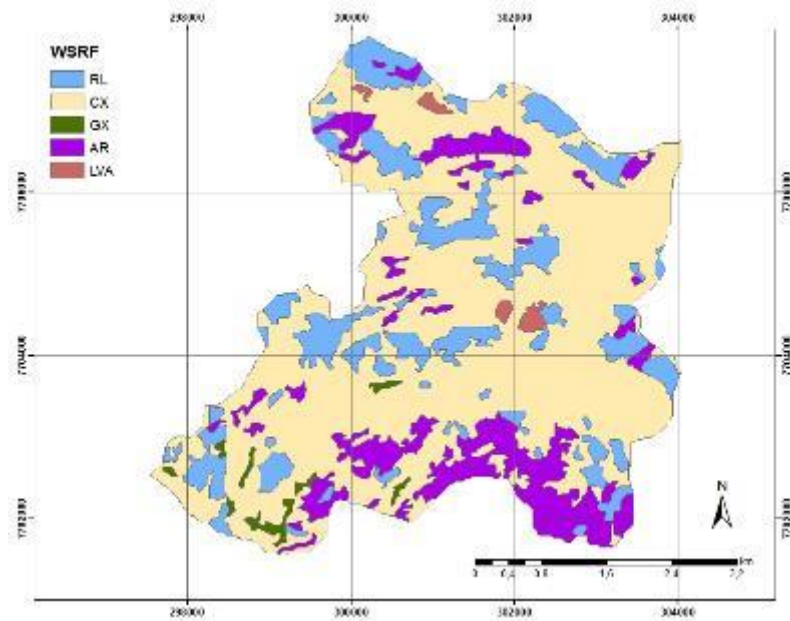


Figura 27- Classificação do solo segundo o método WSRF.

RL- Neossolo Litólico, CX- Cambissolo Háplico, GX-GleissoloHáplico, AR- Afloramento rochoso, LVA-Latossolo Vermelho-Amarelo

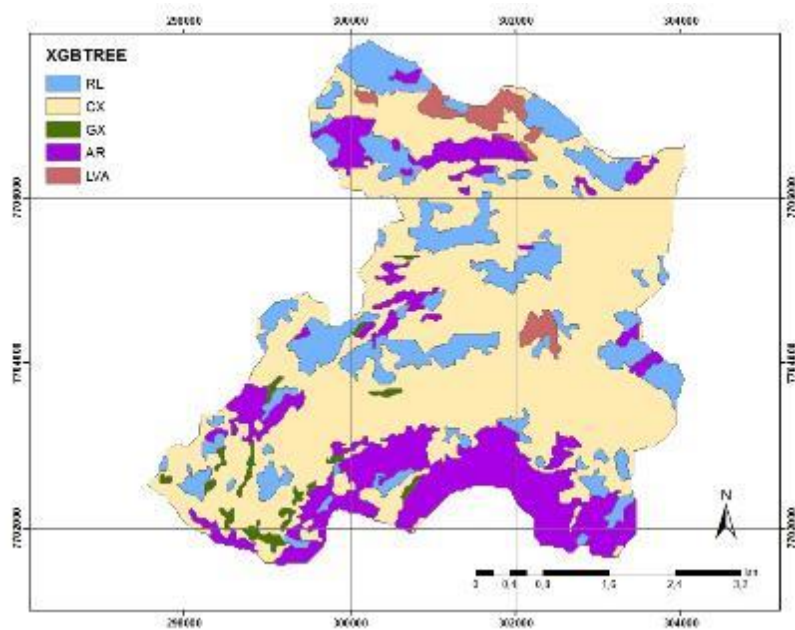


Figura 28- Classificação do solo segundo o método XGBTREE.

RL- Neossolo Litólico, CX- Cambissolo Háplico, GX-GleissoloHáplico, AR- Afloramento rochoso, LVA-Latossolo Vermelho-Amarelo

5 CONCLUSÕES

1. É possível o mapeamento digital de solos do Ribeirão Inhaúma em Iconha/ES, utilizando-se as técnicas de RF, Ranger, XGBTREE e WSRF, com Kappa superior a 0,63.
2. Áreas com maior número de amostragens, foram melhor treinadas pelo classificador, obtendo as maiores acurácias do produtor e do usuário.
3. Comprovou-se a eficiência de covariáveis do terreno e dados magnetométricos e gamespectrométricos como variáveis preditoras de classes de solo na região estudada.

6 REFERÊNCIAS BIBLIOGRÁFICAS

ABREU, VS. Uso de georadar e mapeamento digital de solos no estudo de mussunungas em área do Grupo Barreiras.[Dissertação].Universidade Federal de Viçosa-UFV, Viçosa, 166 p. Fev de 2017.

BALLANTINE, J.A.C., OKIN, G.S., PRENTISS, D.E., ROBERTS, D.A. Mapping North African landforms using continental scale unmixing of MODIS imagery. *Remote Sens. Environ.* 97 (4), 470–483.2005.

BECEGATO, V. A.; FERREIRA, F. J. F. Gamaespectrometria, resistividade elétrica e suscetibilidade magnética de solos agrícolas no noroeste do Estado do Paraná. *Revista Brasileira de Geofísica*, 23:371-405, 2005.

BEHRENS T.; Zhu A-X; Schmidt K; Scholten T. Multi-scale digital terrain analysis and feature selection for digital soil mapping. *Geoderma*, [s.l.], v. 155, n. 3-4, p.175-185, mar. 2010. Elsevier BV. <http://dx.doi.org/10.1016/j.geoderma.2009.07.010>.

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

BREIMAN, L. (1996): "Bagging predictors". *Machine Learning*, Vol 24, 2, pp.123–140.

BREIMAN, L. (1998): "Arcing classifiers". *The Annals of Statistics*, 26(3), 801–849.

BREIMAN, L. (2001). *Random forests*. *Machine learning* 45:5-32.

BRUNGARD COLBY W; BOETTINGER JANIS L. ; DUNIWAY MICHAEL C.; WILLS SKYE A. ; EDWARDS JR. THOMAS C. Machine learning for predicting soil classes in three semi-arid landscapes. *Geoderma*, [s.l.], v. 239-240, p.68-83, fev. 2015. Elsevier BV. <http://dx.doi.org/10.1016/j.geoderma.2014.09.019>.

CAMERA C, ZOMENI Z, NOLLER JS, ZISSIMOS AM, CHRISTOFOROU IC, BRUGGEMAN A. A high resolution map of soil types and physical properties for Cyprus: A digital soil mapping optimization. *Geoderma*, [s.l.], v. 285, p.35-49, jan. 2017. Elsevier BV. <http://dx.doi.org/10.1016/j.geoderma.2016.09.019>.

CARVALHO JUNIOR, W. de; CHAGAS, C. da S, FERNANDES FILHO, E.I; VIEIRA, C. A; SCHAEFER, C. E. G. R; BHERING, S. B; FRANCELINO, M. R. Digital soilscape mapping of tropical hillslope areas by neural networks. *Scientia Agricola*, v.68, n.6, p.691-696, 2011.

CHAGAS C DA S; CARVALHO JUNIOR W DE; BHERING SB. Integração de dados do Quickbird e atributos do terreno no mapeamento digital de solos por redes neurais artificiais. *R. Bras. Ci. Solo*, 35:693-704, 2011.

CHAGAS, C. S; FERNANDES FILHO, E. I.; VIEIRA, C. A. O.; SCHAEFER, C. E. G. R.; CARVALHO JÚNIOR, W. Atributos topográficos e dados do Landsat7 no mapeamento digital de solos com redes neurais. *Pesquisa Agropecuária Brasileira*, 45:497-507, 2010.

CHAGAS, César da Silva; CARVALHO JÚNIOR, Waldir de; BHERING, Silvio Barge. Integração de dados do quickbird e atributos do terreno no mapeamento digital de solos por redes neurais artificiais. *Revista Brasileira de Ciência do Solo*, [s.l.], v. 35, n. 3, p.693-704, jun. 2011. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0100-06832011000300004>.

CHAGAS, CESAR DA SILVA; FERNANDES FILHO, ELPÍDIO INÁCIO; BHERING, SILVIO BARGE. Relação entre atributos do terreno, material de origem e solos em uma área no noroeste do estado do Rio de Janeiro. *Sociedade & Natureza*, [s.l.], v. 25, n. 1, p.147-162, abr. 2013. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s1982-45132013000100012>.

CHEN, T. AND GUESTRIN, C. (2016). XGBoost: A scalable tree boosting system. *CoRR*, abs/1603.02754. arXiv:1603.02754v3 [cs.LG]. 10 Jun 2016.

CHEN, T.; HE, T. Xgboost: eXtreme Gradient Boosting.R package version 0.4-4. 2017.

CONGALTON, R.; A. REKAS. COMPAR: A computerized technique for the in depth comparison of remotely sensed data. *Proceedings of the Fifty First Annual Meeting of the American Society of Photogrammetry*, Washington, DC. 1985. p. 98-106.

CONGALTON, R.G. & MEAD R.A. A quantitative method to test for consistency and correctness in photointerpretation. *Photogrammetric Engineering and Remote Sensing*, v.49, n.1, p.69-74, 1983.

CONGALTON, R.G.; GREEN, K. *Assessing the accuracy of remotely sensed data: principles and practices*. New York: Lewis Publishers, 2009.

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

COSTA, M DA GA. *Gamaespectrometria e magnetometria em mapeamento de solos no domínio de rochas pelíticas do Grupo Bambuí, MG*. [Dissertação]. 83 p. UFV-Universidade Federal de Viçosa. Viçosa-MG. 2013.

CPRM – Serviço Geológico do Brasil. Projeto aerogeofísico Espírito Santo. Relatório final do levantamento e processamento dos dados magnetométricos e gamaespectrométricos. Ministério de Minas e Energia. *Prospectores Aerolevantamentos e sistemas LTDA*. v. 1. 206p. 2010.

DIAS LM da S. predição de classes de solo por atributos do meio físico e de sensoriamento remoto em área da bacia sedimentar do São Francisco. [Dissertação]. Instituto Agonômico de Campinas. Campinas/SP. 127 p. 2015.

DICKSON, B. L.; SCOTT, K. M. Interpretation of aerial gamma-ray surveys – adding the geochemical factors. AGSO Journal of Australian Geology & Geophysics, 17:187-200, 1997.

DISTRIBUTED (DEEP) MACHINE LEARNING COMMUNITY (2017). Scalable, portable and distributed gradient boosting library, Disponível em : <<https://github.com/dmlc/xgboost>>. Acesso em Maio, 2017.

EMBRAPA – CENTRO NACIONAL DE PESQUISA DE SOLO. **Manual de Métodos de Análise de Solo**. 2ed. Rio de Janeiro, RJ: EMBRAPA Solos, 230p. - Documentos/Embrapa Solos, ISSN 1517-2627:132, 2011. EMBRAPA

EMBRAPA-Centro Nacional de Pesquisa de Solos. Sistema Brasileiro de classificação de solos. Rio de Janeiro: EMBRAPA_SPI . 2ed. 306 p. 2006.

ESCABIAS, C.B. : Tree Boosting Data Competitions with XGBoost. (Dissertação de mestrado).Universitat Politècnica de Catalunya - Universitat de Barcelona. Departament d'Estadística I Investigació Operativa. 52 p. 2017.

FARIA, M.M.; FERNANDES FILHO, E.I. Avaliação dos algoritmos SVM e Maxver para a classificação de sistemas florestais monodominantes de candeia (*Eremanthus* sp.). In: Simpósio Brasileiro de Sensoriamento Remoto (SBSR), 16, 2013, Foz do Iguaçu. Anais...São José dos Campos:INPE, 2013. Artigos, p.3222-3229. CD-ROM, On-line. ISBN: 978-85-17-00066-9. Disponível em:< <http://www.dsr.inpe.br/sbsr2013/files/p0882.pdf>>. Acesso em Maio, 2017.

FERNANDES, E. R. Q.; CARVALHO, A. P. L. de ;COELHO, A. L. V. Ensembles de Classificadores para Bases de Dados Desbalanceadas: Uma Abordagem baseada em Amostragem Evolucionária. Symposium on Knowledge Discovery, Mining and Learning, 2th, 2014, São Carlos. Universidade de São Paulo. Biblioteca Digital da Produção Intelectual – BDPI. Departamento de Ciências de Computação - ICMC/SCC.

FOODY, G. M.; MATHUR, A. A Relative Evaluation of Multiclass Image Classification by Support Vector Machines. IEEE Transactions on Geoscience and Remote Sensing, v.42, n.6, 2004.

FOODY, G.M. Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. Photogrammetric Engineering & Remote Sensing, v.70, n.5, p.627-633, 2004.

FREUND, Y. and Schapire, R.E. “Experiments with a new boosting algorithm”. In Proceedings of the Thirteenth International Conference on Machine Learning, 148–156, Morgan Kaufmann. 1996.

FREUND, Y.; SCHAPIRE, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, New York, v. 55, n. 1, p. 119-139, Aug. 1997.

FRIEDMAN J, HASTIE T, TIBSHIRANI R. (2000). \Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors).\" The annals of statistics, 28(2) 337-407.

- FRIEDMAN JH (2001). "Greedy function approximation: a gradient boosting machine." *Annals of Statistics*, pp. 1189-1232.
- GIASSON, E.; CLARKE, R. T.; INDA JUNIOR, A. V.; MERTEN, G. H.; TORNQUIST, C.G. Digital soil mapping using multiple logistic regression on terrain parameters in southern Brazil. *Scientia Agricola*, v.63, n.3, p.262-268, 2006.
- GRIMM, R., BEHRENS, T., MÄRKER, M., ELSENBEER, H., 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island — digital soil mapping using Random Forests analysis. *Geoderma* 146, 102–113. <http://dx.doi.org/10.1016/j.geoderma.2008.05.008>.
- GRUNWALD S, RAMASUNDARAM V., COMERFORD N.B., BLISS C.M. Are Current Scientific Visualization and Virtual Reality Techniques Capable to Represent Real Soil-Landscapes? Cap 43. In: P. Lagacherie, A.B. McBratney and M. Voltz (eds.). *Digital Soil Mapping - An Introductory Perspective*. Elsevier Publisher - Developments in Soil Science Series. 2006.
- HANSEN, M.K.; BROWN, D.J.; DENNISON, P.E. Inductively mapping expert-derived soil-landscape units within dambo wetland catenae using multispectral and topographic data. *Geoderma*, v.150, p.72-84, 2009.
- HOTHORN T., HORNIK K. AND ZEILEIS A. Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics* 15(3), 651-674. 2006.
- HUANG, C.; DAVIS, L. S.; TOWNSHEND, J. R. G. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, v. 23, n. 4, p. 725-749, 2002.
- HUDSON BD. The soil survey as paradigm-based science. *Soil Sci. Soc. Am. J.*, v. 56, p.836–841. 1992.
- IBM- International Business Machines. Nó C5. IBM Knowledge Center. Disponível em: < https://www.ibm.com/support/knowledgecenter/pt-br/SS3RA7_17.1.0/modeler_mainhelp_client_ddita/clementine/c50node_general.html> Acesso em Maio, 2017.
- JENNY, H. *Factors of Soil Formation*. New York, McGraw-Hill. 281. 1941.
- KUHN, M.; JOHNSON, K. *Applied Predictive Modeling*. Springer, New York, 2013.
- LAGACHERIE, P.; MCBRATNEY, A. B. Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping. In: LAGACHERIE, P. et al. *Digital soil mapping: an introductory perspective*. Amsterdam: Elsevier, 2007, Cap. 1, p. 3-22.
- LIMA LA DE S; NEUMANN MRB; REATTO A; ROIG HL. *Mapeamento de Solos: do tradicional ao digital*. Série Documentos:316. Planaltina, DF: Embrapa Cerrados, 2013.
- LOPES, K.M. de O. Modelos baseados em data mining para classificação multitemporal de culturas no Mato Grosso utilizando dados de NDVI/MODIS. Dissertação (mestrado). Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica. 2013.

- LORENA, A.C. ; CARVALHO, A. C. P. L. F. de . Uma Introdução às Support Vector Machines RITA, v. XIV , n. 2 .2007.
- MATTERA, D., PALMIERI, F., AND HAYKIN, S. (1999). An explicit algorithm for training support vector machines. IEEE Signal Processing Letters, 6(9):243–245.
- McBRATNEY, A. B. & LAGACHERIE P.; Voltz, M. Digital Soil Mapping, Na Introductory Perspective. Developments in Soil Science. Volume 31. Amsterdam: Elsevier, 2007. 600p.
- McBRATNEY, A. B. A thoeretical framework for land evaluation. Geoderma, Amsterdam, v. 72, p. 198-199, 1996.
- McKENSIE, N.J., GESSLER, P.E., RYAN, P.J. AND O’CONNELL, D. The role of terrain analysis in soil mapping. In Wilson, J. and Gallant, J.,editors, Terrain analysis: principles and applications. New York City, NY: John Wiley and Sons, 65-245. 2000.
- MELGANI, F.; BRUZZONE, L. Classification of Hyperspectral Remote Sensing Images with Support Vector Machines. IEEE Transactions on Geoscience and Remote Sensing, vol. 42, No. 8, August 2004.
- MENG, Q , ZHAO, HE, WILLIAMS, G.J., LV, J. , XU, B. XU , HUANG, J.Z.. Disponível em <<https://cran.r-project.org/web/packages/wsrp/wsrp.pdf>>. Acesso em Maio, 2017.
- NASCIMENTO, R.F.F.; ALCÂNTARA, E.H. de; KAMPEL, M.; STECH, J.L.; MORAES, E.M.L.de; FONSECA, L.M.G. O algoritmo Support Vector Machines (SVM): avaliação da separação ótima de classes em imagens CCD-CBERS-2 In: Simpósio brasileiro de Sensoriamento Remoto (SBSR), 14, 2009, Natal. Anais...São José dos Campos: INPE, 2009. Artigos, p. 2079-2086. CD-ROM, On-line. ISBN: 978-85-17-00044-7. Disponível em: <<http://marte.sid.inpe.br/col/dpi.inpe.br/sbsr@80/2008/10.20.10.59/doc/2079-2086.pdf>> Acesso em Maio, 2017.
- OLIVEIRA, L.M. de. Classificação de dados sensoriais de cafés especiais com resposta multiclasse via Algoritmo Boosting e Bagging (Dissertação de mestrado) – Lavras : UFLA, 2016. 85 p.Universidade Federal de Lavras, 2016.
- OSHIRO, T.M. Uma abordagem para construção de uma árvore a partir de uma random Forest para classificação de bases de expressão gênica. (Dissertação de mestrado). USP- Universidade de São Paulo. Ribeirão Preto, 2013. 101 p.
- PACKAGE ADABAG. 2015. Applies Multiclass AdaBoost.M1, SAMME and Bagging.
- PACKAGE C50. C5.0 Decision Trees and Rule-Based Models . 2015. Max Kuhn, Steve Weston, Nathan Coulter, Mark Culp.
- PACKAGE PARTY. A Laboratory for Recursive Partytioning. 2017.
- PACKAGE WSRF. 2017. Wsrp: um pacote R para subescala ponderada escalável florestas aleatórias. Disponível em< <https://cran.r-project.org/web/packages/wsrp/README.html>>. Acesso em Maio, 2017.
- PAHLAVAN-RAD MOHAMMAD REZA; KHORMALI FARHAD ; TOOMANIAN NORAIR ; BRUNGARD COLBY W. ; KIANI FARSHAD ; KOMAKI CHOOGHI BAYRAM ; BOGAERT PATRICK. Legacy soil maps as a covariate in digital soil

- mapping: A case study from Northern Iran. *Geoderma*, [s.l.], v. 279, p.141-148, out. 2016. Elsevier BV. <http://dx.doi.org/10.1016/j.geoderma.2016.05.014>.
- PAL, M.; MATHER, P.M. Support vector machines for classification in remote sensing. *International Journal of Remote Sensing*, v. 26, n. 5, p. 1007-1011. 2005.
- PATIL, N.; LATHI, R.; CHITRE, V. Comparison of C5.0 & CART Classification algorithms using pruning technique. *International Journal of Engineering Research & Technology*. V.1, n.4, 2012.
- QUINLAN, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.
- ROSENFELD, G. H; FITZPATRICK-LINS, K. & LING, S. H. Sampling for Thematic Map Accuracy Testing. *Photogrammetric Engineering and Remote Sensing*, Vol. 48 (1), p. 131-137, 1982.
- RUBESAM, A. Estimação não paramétrica aplicada a problemas de classificação via Bagging e Boosting. (Dissertação de mestrado). Universidade Estadual de Campinas-UNICAMP, Instituto de Matemática, Estatística e Computação Científica. 127p. Campinas. 2004.
- RUIZ, L.F.C. Uma abordagem de classificação da cobertura da terra em imagens obtidas por veículo aéreo não tripulado. (Dissertação de mestrado). Porto Alegre: UFRS-Universidade Federal do Rio Grande do Sul. 105 p. 2015.
- SAMUEL-ROSA, A., HEUVELINK, G.B.M., VASQUES, G.M., ANJOS, L.H.C. Do more detailed environmental covariates deliver more accurate soil maps? *Geoderma*, [s.l.], v. 243-244, p.214-227, abr. 2015. Elsevier BV. <http://dx.doi.org/10.1016/j.geoderma.2014.12.017>.
- SANTOS, E.M. dos. Teoria e Aplicação de Support Vector Machines à Aprendizagem e Reconhecimento de Objetos Baseado na Aparência. (Dissertação de mestrado). Universidade Federal da Paraíba. Campina Grande. 2002.
- SANTOS, V.N.P. dos. Modelo de data mining para detecção de tumores em exames de rastreio. (Dissertação de mestrado). INSTITUTO SUPERIOR DE ENGENHARIA DE LISBOA-ISEL. 2013
- SCHAPIRE, R. E.; FREUND, Y. Boosting: foundations and algorithms. Cambridge: MIT, 2012.
- SCHEER, Maurício Bergamini; CURCIO, Gustavo Ribas; RODERJAN, Carlos Vellozo. Funcionalidades ambientais de solos altomontanos na Serra da Igreja, Paraná. *Revista Brasileira de Ciência do Solo*, [s.l.], v. 35, n. 4, p.1013-1026, ago. 2011. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0100-06832011000400005>.
- SCULL P; FRANKLIN J.; CHADWICK OA; MCARTHUR D . Predictive soil mapping: a review. *Progress In Physical Geography*, [s.l.], v. 27, n. 2, p.171-197, jun. 2003.
- SILVA, S. M. P.; CROSTA, A. P.; FERREIRA, F. J. F.; BEURLIN, H.; SILVA, A. M.; SANTOS, L. F. Espectrometria de raios gama de granitos pegmatíticos da Província pegmatítica da Borborema (PPB), Nordeste do Brasil. *Revista Brasileira de Geofísica*, 28:673-690, 2010.

SIRTOLI, A. E. Mapeamento de solos com auxílio de atributos do terreno, índices espectrais e geologia integrados por redes neurais artificiais. [Tese]. Universidade Federal do Paraná, Setor de Ciências da Terra. Curitiba.2008.

SOUZA, P.H.T.de O. Classificação utilizando técnicas de aprendizado estatístico: Estudo de casos. (Trabalho de conclusão de curso). Universidade de Brasília. Departamento de Estatística. Brasília. 2016.

TAGHIZADEH-MEHRJARDIR; NABIOLLAHI K.; MINASNY B.; TRIANTAFILISD J. Comparing data mining classifiers to predict spatial distribution of USDA-family soil groups in Baneh region, Iran. *Geoderma*, [s.l.], v. 253-254, p.67-77, set. 2015. Elsevier BV. <http://dx.doi.org/10.1016/j.geoderma.2015.04.008>.

TAN, P.-N., STEINBACH, M., and KUMAR, V. Introdução ao data mining: mineração de dados. Rio de Janeiro: Ciência Moderna, 2009.

TEN CATEN, A.; DALMOLIN RSD; PEDRON F DE A; SANTOS M DE LM. Componentes principais como preditores no mapeamento digital de classes de solos. *Ciência Rural*, [s.l.], v. 41, n. 7, p.1170-1176, jul. 2011. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0103-84782011000700011>.

TEN CATEN, A.; DALMOLIN, R. S. D.; RUIZ, L. F. C.; SEBEM, E.; PEREIRA, R. S. P. Mapeamento digital de solos através da aplicação de componentes principais em modelos logísticos. In: Simpósio Brasileiro de Sensoriamento Remoto, Natal, RN, 2009. Anais XIV. INPE., 2009. p. 7677-7684.

TSO, B.; MATHER, P. Classification methods for remotely sensed data. Danvers: Taylor & Francis Group, 2009, 376 p. 2ed. ISBN 9781420090727

VAPNIK, V. N. (1999). The Nature of Statistical Learning Theory. Springer Verlag, New York, 2nd edition. Very High-Dimensional Data with Random Forests Built from Small Subspaces. *International Journal of Data Warehousing and Mining (IJDWM)*, 8(2), 44-63.

XU B., HUANG J. Z., WILLIAMS G., WANG Q., YE Y. Classifying very high-dimensional data with random forests built from small subspaces. *International Journal of Data Warehousing and Mining*. 2012;8(2):44–63. doi: 10.4018/jdwm.2012040103.

WRIGHT, M.N; E ZIEGLER, A. Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. arXiv:1508.04409v1 [stat.ML] 18 Aug 2015.

WRIGHT, M;N. A Fast Implementation of Random. ForestsPackage Ranger. p. 1-19 . 2017

ZADROZNY, B. Aprendizado de Máquina. 2010. Disponível em:< <http://www2.ic.uff.br/~bianca/aa/aulas/AA-Aula4.pdf>>. Acesso em Maio, 2017.

ZERAATPISHEH MOJTABA; AYOUBI SHAMSOLLAH, JAFARI AZAM , FINKE PETER . Comparing the efficiency of digital and conventional soil mapping to predict soil types in a semi-arid region in Iran. *Geomorphology*, [s.l.], v. 285, p.186-204, maio 2017. Elsevier BV. <http://dx.doi.org/10.1016/j.geomorph.2017.02.015>.

ZHU, A. X; HUDSON, B; BURT, J; LUBICH, K; SIMONSON, D. Soil mapping using GIS, expert knowledge, and fuzzy logic. *Soil Science Society of American Journal*, Madison,v.65, p. 1463-1472, 2001.

CAPÍTULO III

MAPEAMENTO DIGITAL DE ATRIBUTOS DO SOLO UTILIZANDO APRENDIZADO DE MÁQUINAS

RESUMO

O mapeamento de atributos químicos do solo em larga escala podem acarretar vários ganhos no planejamento de uso e ocupação do mesmo. Atualmente existem várias técnicas que estão sendo testadas para tal fim, porém não foi encontrado na literatura o mapeamento digital de uma larga escala de elementos presentes no solo. Existem várias técnicas promissoras pelo seu caráter robusto e/ou flexível, como *Random Forest-RF*, *Ridge Regression-RIDGE*; *Cubist*, *Partial Least Squares-PLS*, *Regressão por componentes principais-PCR*, *Adaptive Forward-Backward Greedy Algorithm-FOBA*; *Generalized Boosted Regression Models-GBM*; *Gradient Boosting with Component-wise Linear Models-GLMBOOST*. Dessa forma objetivou-se mapear uma gama de atributos químicos do solo em Iconha/ES, comparando diferentes métodos para predição. Como variáveis independentes foram usadas 40 covariáveis geradas a partir do modelo digital de elevação-MDE, além de dados geofísicos, imagens do satélite Landsat 8 e o NDVI. Foram ainda calculadas todas as combinações possíveis entre as bandas, gerando 28 novas variáveis, e entre Th, U e K, gerando 3 novas variáveis. As últimas variáveis calculadas foram as distâncias entre os quatro pontos das extremidades da bacia (d1, d2, d3 e d4). As variáveis dependentes do modelo foram Al, Ca, Fe, K, Mg, Na, Si, Ti, Cr, Cu, Mn, Ni, P, Pb, V, Zn, Zr, S, Cl. Foram coletados 200 amostras na profundidade 0-10 e 10-30 cm na área da bacia, e foi feita análise de elementos totais em analisador de raios-X. O algoritmo Random Forest mostrou-se superior aos demais modelos para predizer os atributos químicos do solo nas duas profundidades, principalmente em 10-30 cm, sendo, portanto, indicado para predição nessa região; Variáveis espaciais mostraram-se altamente prescindíveis para predição, devendo ser consideradas nas modelagens dos elementos químicos do solo; É possível a predição de Al, Ca, Cr, Fe, K, Mn, P, Pb, S, Ti, V, Zn e Zr com R² variando de 0,32 a 0,62 pelos métodos testados.

Palavras-chave: mapeamento de atributos do solo, variáveis espaciais, *Random Forest*

ABSTRACT

The mapping of soil chemical attributes on a large scale can lead to several gains in the planning of soil use and occupation. Currently there are several techniques that are being tested for this purpose, but the digital mapping of a large scale of elements present in the soil has not been found in the literature. There are several promising techniques for their robust and / or flexible character, such as *Random Forest-RF*, *Ridge Regression-RIDGE*; *Cubist*, *Partial Least Squares-PLS*, *Principal Regression-PCR*, *Adaptive Forward-Backward Greedy Algorithm-FOBA*; *Generalized Boosted Regression Models-GBM*; *Gradient Boosting with Component-wise Linear Models-GLMBOOST*. In this way, the objective was to map a range of soil chemical attributes in Iconha / ES, comparing different methods for prediction. As independent variables, 40 covariates were generated from the digital elevation-MDE model, as well as geophysical data, Landsat 8 satellite images and NDVI. All possible combinations between the bands were calculated, generating 28 new variables, and between Th, U and K, generating 3 new variables. The last calculated variables were the distances between the four points of the ends of the basin (d1, d2, d3 and d4). The dependent variables of the model were Al, Ca, Fe, K, Mg,

Na, Si, Ti, Cr, Cu, Mn, Ni, P, Pb, V, Zn, Zr, S, Cl. 200 samples were collected at depth 0-10 and 10-30 cm in the basin area, and analysis of total elements was performed in x-ray analyzer. The Random Forest algorithm was superior to the other models to predict the chemical attributes of the soil in the two depths, mainly in 10-30 cm, being, therefore, indicated for prediction in this region; Spatial variables were highly expendable for prediction and should be considered in the modeling of soil chemical elements; It is possible to predict Al, Ca, Cr, Fe, K, Mn, P, Pb, S, Ti, V, Zn and Zr with R² ranging from 0.32 to 0.62 by the methods tested.

Keywords: mapping of soil attributes, spatial variables, Random Forest

1 INTRODUÇÃO

Novas técnicas computacionais tem se apresentado como uma ferramenta para mapeamento de classes e atributos de solos, proporcionando maior rapidez e diminuição do custo em relação aos métodos convencionais (PINHEIRO, et al. 2012). O estudo e espacialização de atributos dos solos se faz importante para o planejamento do uso e ocupação do mesmo (BHERING, et al. 2015) e, através do mapeamento digital, é possível obter essas informações de maneira rápida para uma grande área com a possibilidade de verificação dos erros obtidos e repetição do método.

O mapeamento convencional de solos é feito com base no modelo CIORPT, onde considera-se que a interação dos cinco fatores (*climate* - Cl, *organisms* - O, *relief* - R, *parent material* - P e *time* - T) é responsável pela formação do solo (DOKUCHAEV, 1949). McBratney et al. (2000), propuseram o modelo SCORPAN, onde uma determinada classe de solo seria função dos cinco fatores do modelo CIORPT, somado aos fatores solo (s) e posição espacial ou geográfica (n). Com esse novo modelo não somente poderia mapear as classes, como também os atributos dos solos.

Desde o final da década de 1960, houve uma ênfase no que poderia ser chamado de abordagens geográficas ou puramente espaciais, ou seja, os atributos do solo poderiam ser preditos a partir da posição espacial em grande parte pela interpolação entre locais de observação do solo (McBRATNEY et al. 2003). Geralmente, pode-se considerar o solo em algum local dependente das coordenadas geográficas x , y e do solo em locais vizinhos $(x + u, y + v)$; $S(x, y) = f((x, y), s(x + u, y + v))$ (McBRATNEY et al., 2003).

Dessa forma, alguns trabalhos começaram a testar o uso da espacialidade para predição de atributos (DAVIES e GAMM, 1969; KISS et al., 1988) e surgiram então métodos de interpolação como a krigagem, entre outros. Porém, o uso apenas da ferramenta espacial levou ao desenvolvimento de modelos cujas superfícies de tendência são representações bastante simplificadas e "não naturais", e os padrões espaciais mais complexos geralmente precisam ser modelados (McBRATNEY et al. 2003).

Com isso, começou-se a adotar modelos híbridos que unem diferentes métodos, como *Random Forest* e a krigagem. Essa junção é feita, com intuito de melhorar as técnicas, não dispensando o uso do componente espacial e esse fator é potencialmente uma fonte valiosa e econômica de informações ambientais, e nunca deve ser desconsiderado (McBRATNEY et al. 2003).

Diferentes fatores derivados do modelo digital de elevação (MDE) têm sido testados como variáveis preditoras (ABREU, 2017; MEIER, 2016; YANG et al. 2016; MENEZES et al., 2014; OLIVEIRA et al. 2012; PINHEIRO et al., 2012; RYAN et al., 2000) para geração de mapas digitais em diferentes métodos. Para o mapeamento de

atributos têm sido avaliadas várias técnicas, dentre elas a *Random Forest* (RF); *Cubist*, *Partial Least Squares* (PLS), dentre outros. Porém não foram encontrados na literatura, trabalhos que mapeassem um número elevado de elementos químicos presentes no solo, utilizando além das covariáveis ambientais, variáveis espaciais como preditoras.

Dessa forma, este trabalho objetivou mapear os atributos químicos do solo Al, Ca, Fe, K, Mg, Na, Si, Ti, Cr, Cu, Mn, Ni, P, Pb, V, Zn, Zr, S, Cl; em Iconha/ES, em duas profundidades, utilizando covariáveis ambientais e espaciais como preditoras e comparando os métodos *Random Forest*, *Ridge regression*, *Cubist*, *Partial Least Squares*, Regressão por componentes principais (PCR), *Adaptive Forward-Backward Greedy Algorithm* - FOBA, *Generalized Boosted Regression Models*- GBM e *Gradient Boosting with Component-wise Linear Models* – GLMBOOST para predição.

2 REVISÃO BIBLIOGRÁFICA

Primeiramente serão sucintamente apresentados os Classificadores que foram testados neste trabalho.

2.1 Classificadores Utilizados Para Predição De Atributos

2.1.1 *Random Forest*

Random Forest é um dos métodos desenvolvidos para classificadores tipo árvore de decisão (LOPES, 2013). Esse modelo combina previsões feitas por múltiplas árvores de decisão, onde cada uma delas é gerada a partir dos valores de um vetor aleatório. Estes vetores são gerados a partir de uma distribuição de probabilidade fixa (TAN et al., 2009).

Dessa forma, RF é uma combinação de preditores de árvore de modo que cada árvore depende dos valores de um vetor aleatório amostrado de forma independente e com a mesma distribuição de todas as árvores (BREIMAN, 2001). Cada uma dessas árvores de decisão gera um voto indicando sua decisão sobre a qual classe pertencerá um determinado objeto; assim o objeto em questão pertencerá a classe que obtiver mais votos. Este tem sido método estatístico cada vez mais usado para modelagem via classificação e regressão (GENUER et al., 2010).

2.1.2 *Ridge Regression*

Este pacote, também conhecido como método de encolhimento (*Shrinkage Method*) (CLÉSIO, 2015), contém funções para montagem de modelos de regressão lineares e logísticos, especialmente para pequenos conjuntos (MORITZ e CULE, 2017), sendo que o parâmetro de regressão pode ser escolhido automaticamente usando o método proposto por Cule e Iorio (2012).

Ele é usado para estimar coeficientes para múltiplos modelos quando as variáveis independentes são altamente correlacionadas (HILT E SEEGRIST, 1977) e quando o

número de variáveis explicativas excede o número de observações na presença de grandes bases de dados, sendo considerado um modelo de regressão penalizada, onde é reduzido os efeitos de variáveis preditivas que não contribuem para modelagem da variável resposta, excluindo-as do modelo, o que evita o *overfitting*, essa escolha é feita baseada na soma dos quadrados dos coeficientes. Uma revisão acerca do método é encontrada em Reynaldo (1997).

Dessa forma, segundo Clésio (2015):

A Ridge Regression é um método de regularização do modelo que tem como principal objetivo suavizar atributos que sejam relacionados uns aos outros e que aumentam o ruído no modelo (A.K.A multicolinearidade). Com isso com a retirada de determinados atributos do modelo, o mesmo converge para um resultado muito mais estável em que com a redução desses atributos, a redução em termos de acúcia do modelo se mantém inalterada. O mecanismo algorítmico que faz isso é através de um mecanismo de penalização que coloca um viés e que vai reduzindo os valores os betas até não zero. Com isso os atributos que contribuem menos para o poder preditivo do modelo são levados para a irrelevância usando esse mecanismo de penalização do viés.

2.1.3 Cubist

De acordo com Kuhn et al. (2016) esta função se ajusta ao modelo baseado em regras descrito em Quinlan (1992) (também conhecido como M5) com correções adicionais baseadas nos vizinhos mais próximos do conjunto de treinamento, conforme descrito em Quinlan (1993).

Inicialmente cria uma estrutura de árvore, gerando um modelo que contém uma ou mais regras, onde cada regra reúne um conjunto de condições associadas (DOZUALDO E REZENDE, 2003). Um modelo de regressão é adequado para cada regra com base no subconjunto de dados definido por essas. Esse conjunto é podado e possivelmente combinado.

2.1.4 Partial Least Squares- PLS

Método desenvolvido por Wold nos anos 60 (MORELLATO, 2010). O PLS (Regressão por quadrados mínimos parciais) é um método estatístico que encontra uma regressão linear capaz de projetar as variáveis preditas e observáveis para um novo espaço (VINZI et al., 2010). O PLS é fácil de usar e frequentemente usado em diferentes campos de aplicação (MARTENS E NAES, 1984).

A regressão PLS é semelhante à regressão dos componentes principais, pois ambos empregam rotações estatísticas para superar os problemas de alta dimensionalidade e multicolinearidade (VOHLAND et al., 2014). PLS é eficiente para tratar dados com alto nível de ruído, dados faltantes e colinearidade (ANDERSEN e BRO, 2010).

Segundo Morellato (2010), algumas vantagens e desvantagens deste método são:

Capacidade de modelar regressões com múltiplas variáveis resposta; não é afetado por multicolinearidade e alto poder de predição; Dificuldade na interpretação das cargas dos fatores; os estimadores dos coeficientes de regressão não possuem distribuições conhecidas e faltam estatísticas de teste para o modelo.

PLS é uma análise multivariada, na qual as variáveis resposta são transformadas em um pequeno número de combinações lineares (ZIMMER e ANZANELLO, 2014). É criado uma série de modelos, cada modelo conseguinte ao anterior contém um

componente adicional, estes componentes são calculados um por vez, iniciando com a matriz X e Y padronizadas.

2.1.5 Regressão por componentes principais- PCR

A regressão por componentes principais é realizada em duas fases, e veio como forma de resolver os problemas apresentados pela regressão linear múltipla, como ruído e colinearidade (Panero, 2010). Assim como PLS, é apropriado para modelagens cujo o número de covariáveis preditoras é maior que o número de observações (AZEVEDO et al., 2013).

Segundo Parreira (2003) em PCR, é utilizado apenas a matrix X para determinar a combinação linear das variáveis (primeira fase), enquanto as concentrações são utilizadas somente quando é feita a estimativa dos coeficientes da regressão, sendo posteriormente os resultados correlacionados com as variáveis independentes y (segunda fase).

Garthwaite (1994) aponta que PLS e PCR apresentam similaridades, diferenciando-se principalmente no fato de PCR usar somente as variáveis explicativas para constituir os componentes, e PLS levar em consideração também a variável dependente.

2.1.6 Adaptive Forward-Backward Greedy Algorithm - FOBA

Adaptive Forward-Backward Greedy Algorithm (FOBA) para regressão de mínimos quadrados é descrito em Zhang (2008). Esse método usa a adição e exclusão de variáveis a medida que gera o modelo, de modo a encontrar o melhor ajuste do mesmo.

Inicialmente, parte-se do pressuposto que há apenas o intercepto, não tendo variáveis adicionadas ao modelo. Assim, é adicionado uma variável por vez, de modo que a primeira a ser acrescentada será aquela com maior correlação com a resposta. Sendo a variável escolhida, é feita a adição de uma próxima variável, sendo essa a que tiver maior correlação com a resposta obtida após a inserção da primeira variável, e assim por diante até que não seja incluída mais nenhuma variável.

A outra maneira possível, é partir inicialmente com todas as variáveis no modelo, e posteriormente são retiradas uma a uma aquelas que não contribuem para o mesmo.

De acordo com Zhang (2015), no método FOBA quando o aumento do risco penalizado é inferior a x vezes a redução do risco na curva correspondente, o mesmo toma um passo atrás.

Para o método "foba.conservative" o passo para trás é dado, quando o aumento de risco é menor que x vezes a menor redução de risco em todas as etapas avançadas anteriormente, enquanto o método "foba.aggressive" toma um passo para trás quando o risco cumulativo no passo para trás é inferior a x vezes as mudanças nos passos avançados Zhang (2015).

2.1.7 Generalized Boosted Regression Models- GBM

Este pacote implementa extensões para o algoritmo *AdaBoost* de Freund e Schapire (1997), e para o *gradient boosting machine* de J. Friedman (2001). Inclui métodos de regressão para mínimos quadrados, perda absoluta, logística, Poisson, modelo de risco proporcional da Cox, Multinomial, T-distribuição, entre outros (RIDGEWAY, 2017).

De acordo com Zampieri e Colombari (2015):

O modelo gradiente *boosted* é de regressão aditiva, na qual os termos são árvores decisórias obtidas após simples partição recursiva. Após a construção de uma primeira árvore decisória, com base nos dados, é organizada outra árvore decisória, para os resíduos da primeira. Esse processo é continuado até que se atinjam limites preestabelecidos, produzindo centenas ou milhares de árvores decisórias que são, então, incluídas por meio de um algoritmo que, eventualmente, produz um modelo final.

Dessa forma, apresenta como vantagens conseguir trabalhar com dados perdidos e *outliers*, se adequar a modelos não lineares e não exigir transformação dos dados (ELITH et al., 2008).

2.1.8 Gradient Boosting with Component-wise Linear Models - GLMBOOST

Nesse modelo, podem ser trabalhados dados de grande dimensão, onde aumenta-se o gradiente para otimizar as funções de perda arbitrária, e os modelos lineares de componente são utilizados como base de aprendizado (HOTHORN et al., 2017). Assim, um modelo linear (generalizado) é ajustado usando um algoritmo de aumento baseado em modelos lineares univariados de componentes (HOTHORN et al., 2017)). Os modelos lineares generalizados (GLMs) estendem os modelos de regressão linear para acomodar com as distribuições de resposta não normal (Hastie e Pregibon, 1992).

O ajuste, isto é, os coeficientes de regressão, podem ser interpretados da maneira usual (HOTHORN et al., 2017).

A metodologia é descrita em Buehlmann e Yu (2003), Buehlmann (2006), e Buehlmann e Hothorn (2007). Exemplos e detalhes adicionais são dados em Hofner et al (2012).

3 MATERIAL E MÉTODOS

O trabalho foi conduzido na bacia hidrográfica do Ribeirão Inhaúma (Figura 29), com área de 2.403,9 ha, com coordenada central de 21°10'58.82"S e 41°00'08.87"O, localizada quase que totalmente no município de Iconha, no sul do estado do Espírito Santo.

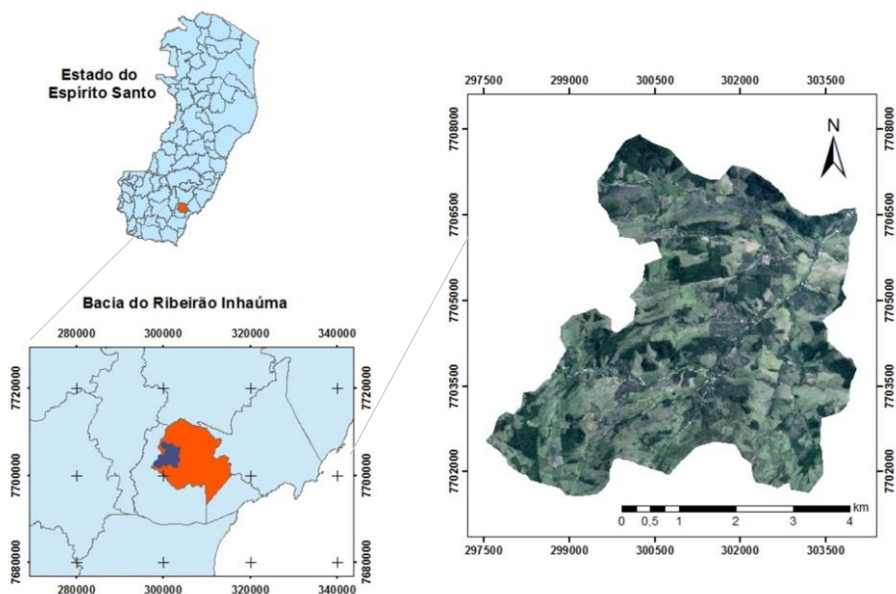


Figura 29- Bacia do Ribeirão Inhaúma, localizada em Iconha/ES.

Para predição dos atributos químicos do solo foi necessário determinar as variáveis independentes que seriam testadas. Foi utilizado o modelo digital de elevação - MDE (Figura 30) obtido a partir de dados do satélite ALOS (*Advanced Land Observing Satellite*) com resolução espacial de 12,5 m para gerar 37 covariáveis (Tabela 28), utilizando para isso script desenvolvido no programa R para aplicar as ferramentas de análise morfométricas do terreno contidas no software livre SAGA (BOHNER et al., 2006) e o mapa de uso e ocupação do solo.

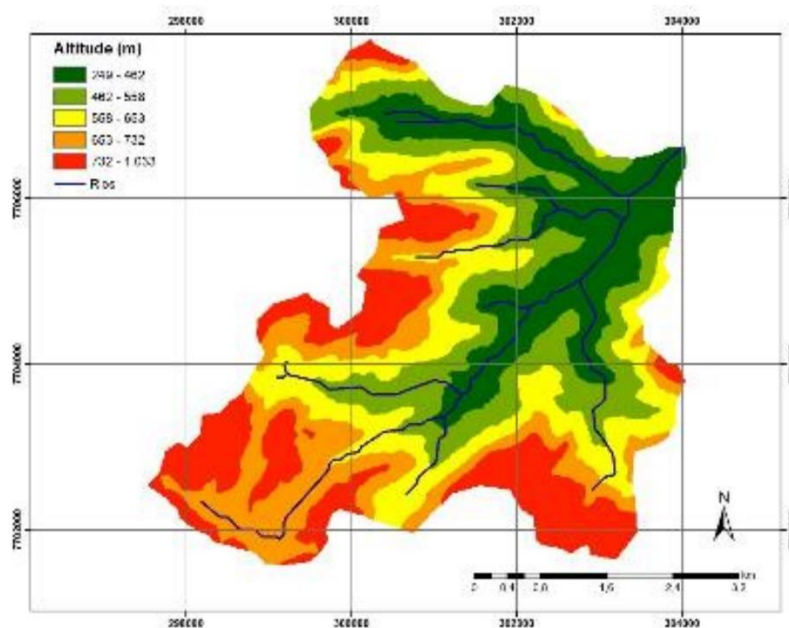


Figura 30- Modelo digital de elevação.

Também foram utilizadas como variáveis independentes as bandas do satélite Landsat 8, o NDVI (*Normalized Difference Vegetation Index*) gerado através destas, calculado, segundo a equação proposta por Rouse et al. (1973), onde é expresso pela razão entre a diferença da medida da reflectância nos canais do infravermelho próximo e vermelho e a soma desses canais, sendo então:

$$\text{NDVI} = (\rho_{\text{IV}} - \rho_{\text{V}}) / (\rho_{\text{IV}} + \rho_{\text{V}}) \quad (1)$$

Onde: NDVI é o Índice de vegetação da diferença normalizada, ρ_{IV} é refletividade no infravermelho próximo e ρ_{V} é a refletividade no vermelho.

Foram também utilizados dados aerogeofísicos obtidos através da Companhia de Pesquisa de Recursos Minerais-CPRM, como a gamaespectrometria e magnetometria (Tabela 27).

Foram calculadas todas as combinações possíveis entre as bandas, gerando 28 novas covariáveis espectrais. Essas foram obtidas, através da relação:

$$B_{xy} = \frac{b_x - b_y}{b_x + b_y} \quad (2)$$

Onde B_{xy} é o resultado da razão entre banda x e banda y, b_x e b_y representam as bandas do Landsat 8.

Para as variáveis provenientes de gamaespectrometria, foi usada a mesma razão acima, onde as combinações foram feitas entre os valores estimados de tório (Th), urânio (U) e potássio (K), gerando três novas variáveis. Foram calculadas as distâncias entre os quatro pontos das extremidades da bacia (d1, d2, d3 e d4), segundo a relação:

$$d_{AB} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (3)$$

Onde: d_{AB} é a distância entre dois pontos A e B, x_2 e y_2 são as coordenadas do ponto A e x_1 e y_1 são as coordenadas do ponto B.

Considerou-se também como variável as coordenadas x e y, ou seja, latitude e longitude. Todos os mapas foram padronizados com o mesmo tamanho de células, linhas, colunas e resolução de 30 m.

Tabela 27- Variáveis independentes utilizadas para predição de atributos químicos do solo

Derivadas do MDE		
Convergence index	Mid Slope Position	Total Curvature
Cross-Sectional Curvature	Minimal Curvature	Total Insolation1
Curvature Classification	Modelo digital de elevação	Total Insolation2
Diffuse Insolation1	Normalized Height	Valley Depth
Diffuse Insolation2	Plan Curvature	Valley Index
Direct to Diffuse Ratio1	Profile Curvature	
Direct to Diffuse Ratio2	Real Surface Area	
Diurnal Anisotropic	Slope	
Duration of Insolation1	Slope Height	
Duration of Insolation2	Slope index	
Flow Line Curvature	Standardized Height	
General Curvature	Surface Specific Points	
Landforms	Tangencial Curvature	
Longitudinal Curvature	Terrain Surface Texture	
Mass Balance	Terrain Surface Convexity	
Maximal Curvature	Topographic Wetness Index	
Dados geofísicos		Outros
Campo Magnético Total		Bandas do Landsat8
Contagem Total		NDVI
Potássio		Mapa de uso e ocupação
Primeira Derivada do Campo Magnético Total		Razão entre bandas
Razão entre Tório e Potássio		Razão entre Th, U e K
Razão entre Urânio e Potássio		Distâncias entre pontos
Razão entre Urânio e Tório		Coordenadas x,y
Sinal Analítico do Campo Magnético Total		
Taxa de Exposição da Contagem Total		
Tório		
Urânio		

As variáveis dependentes do modelo foram: Al, Ca, Fe, K, Mg, Na, Si, Ti, Cr, Cu, Mn, Ni, P, Pb, V, Zn, Zr e S. Para se obter o valor observado foram coletadas no campo, 200 amostras nas profundidades de 0-10 cm e 10-30 cm. Para se estabelecer a malha de pontos, foi utilizado o método do hiper-cubo latino condicionado, devido à dificuldade de acesso, pelo relevo extremamente montanhoso da área. As coordenadas de cada ponto de coleta foram registradas com auxílio de receptor GNSS modelo GS09 Plus, da Leica. Os dados foram processados utilizando-se o programa Leica Geoffice 8.0, utilizando a estação fixa do Instituto Brasileiro de Geografia e Estatística-IBGE presente em Vitória/ES para transporte da base para a área da bacia.

As amostras foram secas ao ar, destorroadas e passadas em peneira com malha de 2 mm. No laboratório foram maceradas em almofariz de ágata e novamente peneiradas em malha 1 mm, posteriormente colocadas em molde padrão e prensadas manualmente, a fim de se formarem pastilhas.

Essas pastilhas foram usadas para leitura no analisador de fluorescência de raios-X do Laboratório de Solos, da Universidade Federal de Viçosa-UFV, fornecendo

os teores totais dos elementos em cada amostra. Foi utilizado para tal o modelo Micro-EDX-1300 da Shimadzu, foram usadas as voltagens padrão do equipamento.

Os resultados que apresentaram valores atípicos (*outliers*) foram identificados e tiveram seus dados substituídos por valores estimados através da técnica de imputação por regressão, onde o novo valor é calculado por intermédio da regressão dos demais valores dessa variável. Dessa forma evitou-se a diminuição de dados na análise.

Foi realizado teste de normalidade para as variáveis resposta nas duas profundidades, utilizando o teste de Kolmogorov Smirnov (K-S) ($p < 0,05$).

Os mapas de atributos foram gerados utilizando-se o software R (*R Development Core Team, 2009*). Inicialmente para a seleção das variáveis independentes mais importantes, foram eliminadas àquelas que possuíam índice de correlação maior que 0,99. Posteriormente cada modelo selecionou as variáveis mais significativas para prever cada um dos atributos, sendo essas usadas para gerar os mapas finais. Buscou-se por modelos mais simplificados, ou seja, que usassem o menor número possível de variáveis com bons valores de R^2 , por isso permitiu-se a perda de até 5% nos valores de R^2 em favor de um modelo mais parcimonioso.

Os valores dos pontos para cada uma das variáveis independentes foram extraídos utilizando-se o software ArcGis, através do comando *Extract value to points*. Para se ter maior representatividade da área, considerou-se como valor típico do local amostrado a média do ponto central e de seus vizinhos.

O desempenho dos modelos foi avaliado utilizando-se o procedimento de validação cruzada com 10 repetições onde foram feitas comparações entre os valores observados e preditos de cada variável dependente. Esses valores foram expressões através do coeficiente de determinação R^2 . Esse índice é calculado conforme a equação:

$$R^2 = \frac{\sum_{i=1}^n (P_i - \bar{o})^2}{\sum_{i=1}^n (O_i - \bar{o})^2} \quad (4)$$

Onde P_i e O_i , são os valores preditos e observados no local i respectivamente; n é o número de amostras.

Também foi calculado o valor de RMSE (raiz quadrada do erro médio).

$$RMSE = \sqrt{\frac{1}{l} \sum_{j=1}^l [\hat{z}(S_j) - z^*(S_j)]^2} \quad (5)$$

em que, RMSE é a raiz quadrada do erro médio, e l é o número de pontos destinados a validação.

4 RESULTADOS E DISCUSSÃO

Na Tabela 28 é mostrado o teste de normalidade segundo Kolmogorov-Smirnov para as camadas 0-10 e 10-30 cm. Na camada 0-10 cm os elementos que mostraram distribuição normal foram Al, Fe, K, Mg, Na, Mn, Pb, Zn e Zr, enquanto na camada 10-30 cm foram Al, K, Cu, Pb, Zn e S. Devido à natureza de distribuição dos atributos nos

solos, nem sempre, estes exibem distribuição normal (BELLON-MAUREL et al., 2010), este fato não prejudica os testes realizados, já que estes não exigem tal distribuição.

Tabela 28- Teste de normalidade para os elementos Al, Ca, Fe, K, Mg, Na, Si, Ti, Cr, Cu, Mn, Ni, P, Pb, V, Zn, Zr e S nas camadas 0-10 e 10-30 cm

Kolmogorov-Smirnov							
		Sig.				Sig.	
Elemento	Nº de amostras em cada prof.	0-10 cm	10-30 cm	Elemento	Nº de amostras em cada prof.	0-10 cm	10-30 cm
Al	100	0,200*	0,200*	Cu	100	0,003	0,200*
Ca	100	0,000	0,014	Mn	100	0,200*	0,172
Fe	100	0,200*	0,031	Ni	100	0,001	0,003
K	100	0,200*	0,200*	P	100	0,014	0,000
Mg	100	0,200*	0,040	Pb	100	0,200*	0,200*
Na	100	0,200*	0,000	V	100	0,052	0,182
Si	100	0,039	0,002	Zn	100	0,200*	0,200*
Ti	100	0,028	0,109	Zr	100	0,200*	0,019
Cr	100	0,014	0,100	S	100	0,005	0,200*

*Distribuição normal pelo teste de Kolmogorov-Smirnov a 5 %

Em relação a predição pelos modelos testados, são apresentados apenas os elementos cujo valor de R^2 foi superior a 0,30 em pelo menos um método. Dessa forma, os métodos não se mostraram eficientes para predição de Ca, Mg, Na, Si, Cr, Cu, Ni e Zr na profundidade 0-10 cm e para Mg, Na, Si, Cu, Mn, Ni e Zn para 10-30 cm.

A Tabela 29 mostra a estatística descritiva dos elementos modelados nas duas profundidades, sendo Al, Fe, K, Mn, P, Pb, S, Ti, V e Zn em 0-10 cm e Al, Ca, Cr, Fe, K, P, Pb, S, Ti, V e Zr em 10-30 cm. Para o conjunto de dados da primeira profundidade, tem-se distribuição assimétrica negativa para Al, K e Zn, e assimétrica positiva para os demais, enquanto em 10-30 cm a distribuição assimétrica negativa ocorre apenas para Al e Fe. Quando se tem assimetria positiva, geralmente, os valores da média são superiores ao da mediana, indicando alta frequência de valores abaixo da média (LIMA et al., 2010), esse comportamento pôde ser observado para P, S e Zn em 0-10 cm e para todos os elementos em 10-30 cm exceto com o Fe.

A assimetria negativa indica a concentração de dados (alongamento da cauda) a esquerda da média, e a assimetria positiva dados a direita da média. Valores mais próximos a zero indicam maior simetria, e com isso distribuição normal dos dados (GROENEVELD e MEEDEN, 1984).

A curtose indica o grau de achatamento dos dados, podendo esses serem classificados em distribuição leptocúrtica (Se valor de curtose for $< 0,263$), mesocúrtica ($= 0,263$) ou platicúrtica ($> 0,263$). Nas duas profundidades os dados possuem distribuição leptocúrtica, sendo caracterizada por apresentar uma curva mais afunilada, com pico mais alto em relação a normal (mesocúrtica), exceto para P e S (0-10 cm) e Ca (10-30 cm), que possuem distribuição platicúrtica, sendo essa curva mais achatada que a mesocúrtica e a leptocúrtica.

Tabela 29- Análise descritiva dos atributos químicos do solo nas profundidades 0-10 cm e 10-30 cm

Variável*	Média	Mediana	Desvio Padrão	Mínimo	Máximo	Assimetria	Curtose
Al	24,88	24,94	4,44	13,76	35,45	-0,33	-0,02
Fe	10,37	10,65	2,76	3,51	17,73	0,003	-0,15
K	1,22	1,25	0,48	0,14	2,22	-0,26	-0,54
Mn	601,40	600,68	315,16	53,50	1314,77	0,24	-0,69
P	2359,64	2082,35	1394,97	0,00	5867,08	0,46	-0,52
Pb	198,65	200,39	69,78	57,27	395,50	0,20	-0,24
S	0,12	0,11	0,05	0,03	0,30	1,02	1,13
Ti	2,12	2,13	0,67	0,97	3,45	0,14	-1,06
V	417,78	424,47	133,77	210,05	719,71	0,15	-1,08
Zn	119,01	118,53	42,78	28,68	232,19	-0,02	-0,41
Profundidade 10-30 cm							
Al	25,02	24,75	4,93	13,44	33,86	-0,15	-0,67
Ca	0,20	0,19	0,13	0,01	0,61	0,93	1,17
Cr	289,28	272,04	109,92	18,12	541,42	0,23	-0,32
Fe	10,36	10,43	2,80	3,31	17,56	-0,13	0,09
K	1,27	1,27	0,31	0,10	2,57	0,15	-0,20
P	2133,07	1736,18	1340,61	0,00	5981,57	0,79	0,18
Pb	202,47	199,63	78,71	61,00	429,81	0,52	0,03
S	0,11	0,10	0,02	0,01	0,22	0,49	-0,37
Ti	2,16	2,14	0,70	0,93	3,80	0,20	-0,83
V	427,82	423,37	139,62	183,15	745,18	0,20	-0,95
Zr	727,76	709,67	223,03	249,38	1315,95	0,50	-0,05

*Al, Ca, Fe, K, S, Ti em dag/kg e Cr, Mn, P, Pb, V, Zn e Zr em ppm.

Em relação as variáveis predictoras, foram eliminadas aquelas que apresentaram correlação igual ou superior a 0,99 entre si. Sendo essas: *Diffuse Insolation2*, *Total Insolation2*, *Valley Index*, *Bandas 2 e 4 do Landsat 8*.

Para camada 0-10 cm, os valores de R^2 (Tabela 30), calculados com base na regressão entre o que foi observado e o predito pelos modelos, mostram que o RF foi superior para três das variáveis (Fe, Mn, V), enquanto RIDGE para duas (S e Zn) e os demais (excepto PLS) para uma. PLS e PCR foram igualmente eficientes para predição de P ($R^2 = 0,42$), enquanto RIDGE, Cubist, PLS, PCR e FOBA para Ti ($R^2 = 0,50$). Os valores de R^2 para esta camada variaram de 0,23 a 0,54.

Quanto aos menores valores de R^2 , FOBA se destacou não apresentando nenhum dos menores coeficientes de correlação. Para Al, Mn, P e Pb houve mais de um método apresentando igualmente os menores valores. De modo geral, PLS, PCR, GBM e GLMBOOST foram aqueles que apresentaram menores valores de R^2 na predição.

Em relação aos elementos os maiores valores de R^2 foram encontrados para Ti (0,50) e V (0,54). Os minerais de Ti são muito resistentes a intempéries, aparentemente, sob condições redutoras, os íons Fe^{2+} são adsorvidos em superfícies de minerais de Ti, podendo o Ti entrar na estrutura de alguns silicatos e provavelmente ser adsorvido na superfície de concreções de Fe-Mn, na solução do solo esta presente em pequenas

quantidades (em torno de 0,03 mg/L), por ser um dos elementos mais estáveis (KABATA-PENDIAS 2010).

Já o V é distribuído em perfis de solo de forma bastante uniforme e a variação no seu conteúdo no solo é herdada dos materiais originais (KABATA-PENDIAS 2010). Assim, as concentrações mais altas de V (até 500 mg/kg) são relatadas para Cambissolos (KABATA-PENDIAS 2010), teores esses que provavelmente facilitam seu mapeamento, já que este é o tipo de solo distribuído amplamente na área de estudo. O V nos solos parece estar associado especialmente aos óxidos hidratados de Fe e a matéria orgânica; Em alguns solos, os minerais de argila também podem controlar a mobilidade desse elemento (KABATA-PENDIAS 2010).

Tabela 30- Valores de R^2 gerados pelos três métodos de predição dos atributos químicos do solo na profundidade 0-10 cm

Elementos	Métodos							
	RF	RIDGE	CUBIST	PLS	PCR	FOBA	GBM	GLMBOOST
Al	0,39	0,37	0,35	0,33	0,33	0,38	0,33	0,40
Fe	0,49	0,41	0,44	0,40	0,41	0,43	0,45	0,41
K	0,40	0,34	0,36	0,34	0,23	0,35	0,41	0,33
Mn	0,38	0,31	0,32	0,31	0,32	0,33	0,36	0,31
P	0,37	0,39	0,39	0,42	0,42	0,40	0,37	0,40
Pb	0,39	0,36	0,42	0,37	0,37	0,38	0,40	0,36
S	0,36	0,38	0,33	0,37	0,37	0,38	0,35	0,37
Ti	0,48	0,50	0,50	0,50	0,50	0,50	0,46	0,48
V	0,54	0,48	0,52	0,50	0,50	0,48	0,52	0,45
Zn	0,37	0,39	0,30	0,27	0,24	0,37	0,36	0,38

Random Forest-RF, Ridge regression-RIDGE, Partial Least Squares-PLS, Regressão por componentes principais-PCR, Adaptive Forward-Backward Greedy Algorithm-FOBA, Generalized Boosted Regression Models- GBM e Gradient Boosting with Component-wise Linear Models – GLMBOOST

Os valores de RMSE (Tabela 31) tenderam a ser menores quanto maior o valor de R^2 , como esperado. Para predição de P, cujo R^2 foi o mesmo tanto para o PLS como para o PCR, é possível inferir que PCR foi superior devido ao menor valor de RMSE, porém o mesmo não foi possível para S e Ti cujos valores de RMSE se mantiveram iguais para os melhores preditores.

Tabela 31- Valores de RMSE encontrados para Al, Fe, Mn, P, Pb, S, Ti, V e Zn na camada 0-10 cm, nos diferentes método testados

Elementos*	Métodos							
	RF	RIDGE	CUBIST	PLS	PCR	FOBA	GBM	GLMBOOST
Al	3,79	3,99	3,96	3,88	3,87	3,82	3,92	3,77
Fe	2,04	2,21	2,18	2,21	2,21	2,16	2,13	2,18
K	0,38	0,42	0,40	0,42	0,44	0,40	0,38	0,41
Mn	257,28	272,11	269,69	270,89	269,45	266,48	260,67	270,95
P	1098,40	1058,68	1082,61	1037,59	1035,29	1053,91	1079,42	1054,49
Pb	62,42	62,78	61,76	62,24	62,23	61,54	62,01	62,23
S	0,04	0,04	0,04	0,04	0,04	0,04	0,04	0,04
Ti	0,50	0,48	0,48	0,48	0,48	0,48	0,51	0,50
V	92,17	98,36	95,23	96,66	96,65	98,24	95,13	102,51
Zn	35,06	35,58	37,69	38,99	39,18	36,08	35,68	35,06

*Al, Fe, K, S, Ti em dag/kg e Mn, P, Pb, V e Zn em ppm. *Random Forest*-RF, *Ridge regression*-RIDGE, *Partial Least Squares*-PLS, *Regressão por componentes principais*-PCR, *Adaptive Forward-Backward Greedy Algorithm*-FOBA, *Generalized Boosted Regression Models*- GBM e *Gradient Boosting with Component-wise Linear Models* – GLMBOOST

Para a camada 10-30 cm obteve-se maior disparidade quanto ao melhor preditor, em relação aos valores de R^2 (Tabela 32). Os atributos do solo, em geral, variam continuamente com a profundidade no perfil do solo (RUSSELL E MOORE, 1968), o que pode gerar diferença na capacidade de predição de cada modelo. RF se destacou em 6 variáveis (Al, Ca, Fe, K, P, Pb) e empatou com FOBA para a variável Zr com $R^2= 0,34$. FOBA foi superior para Cr, GBM para Ti e V, e PLS e PCR obtiveram mesmo valor de R^2 para S (0,37). Os valores de R^2 variaram de 0,24 a 0,62.

Como na camada 0-10 cm, o V e o Ti foram os elementos melhor preditos com R^2 de 0,62 e 0,56 respectivamente.

RIDGE, *Cubist* e GLMBOOST não se mostraram preditores eficientes na camada 0-30 cm, não obtendo nenhum dos maiores valores de R^2 , e estando entre os menores valores de coeficiente de correlação encontrados. FOBA assim como na camada 0-10 cm não obteve nenhum valor inferior de R^2 , bem como RF nesta camada.

Tabela 32- Valores de R² gerados pelos três métodos de predição dos atributos químicos do solo na profundidade 10-30 cm

Elementos	Métodos							
	RF	RIDGE	CUBIST	PLS	PCR	FOBA	GBM	GLMBOOST
Al	0,37	0,28	0,31	0,29	0,30	0,32	0,36	0,33
Ca	0,32	0,29	0,26	0,24	0,24	0,31	0,28	0,29
Cr	0,26	0,32	0,24	0,33	0,32	0,35	0,28	0,32
Fe	0,43	0,33	0,35	0,33	0,33	0,34	0,39	0,33
K	0,38	0,35	0,32	0,36	0,36	0,36	0,36	0,36
P	0,45	0,36	0,37	0,42	0,43	0,40	0,40	0,39
Pb	0,32	0,21	0,25	0,29	0,30	0,25	0,30	0,23
S	0,32	0,32	0,32	0,37	0,37	0,32	0,30	0,30
Ti	0,52	0,42	0,44	0,44	0,44	0,44	0,56	0,43
V	0,55	0,52	0,52	0,50	0,50	0,55	0,62	0,52
Zr	0,34	0,28	0,27	0,28	0,28	0,34	0,28	0,31

Random Forest-RF, Ridge regression-RIDGE, Partial Least Squares-PLS, Regressão por componentes principais-PCR, Adaptive Forward-Backward Greedy Algorithm-FOBA, Generalized Boosted Regression Models- GBM e Gradient Boosting with Component-wise Linear Models – GLMBOOST

O comportamento do RMSE na camada 10-30 cm foi semelhante a primeira, ou seja, os menores valores foram, em geral, encontrados para as maiores estimativas de R² (Tabela 33). A única exceção foi para o P, onde o menor valor de RMSE foi obtido com RIDGE. Com menores valores de RMSE, é possível inferir que RF foi superior a FOBA para predição de Zr, cujo R² havia sido semelhante. Em relação a S, que também apresentou seu melhor valor de R² em mais de um método, teve seus valores de RMSE também iguais.

Tabela 33- Valores de RMSE encontrados para Al, Ca, Cr, Fe, K,P, Pb, S, Ti, V e Zr na camada 10-30 cm, nos diferentes método testados

Elementos*	Métodos							
	RF	RIDGE	CUBIST	PLS	PCR	FOBA	GBM	GLMBOOST
Al	4,10	4,61	4,38	4,50	4,47	4,34	4,15	4,25
Ca	0,23	0,24	0,24	0,24	0,24	0,23	0,23	0,23
Cr	100,42	94,74	102,97	94,90	94,99	92,03	97,81	94,32
Fe	2,35	2,55	2,52	2,54	2,54	2,52	2,46	2,55
K	0,45	0,47	0,47	0,46	0,46	0,46	0,45	0,46
P	1063,33	943,15	1031,79	972,34	966,61	991,67	991,74	1005,35
Pb	71,13	82,70	76,68	73,21	72,99	76,79	72,37	76,53
S	0,04	0,04	0,04	0,04	0,04	0,04	0,04	0,04
Ti	0,50	0,55	0,54	0,54	0,54	0,54	0,48	0,54
V	95,76	100,57	99,59	101,79	101,79	97,11	89,07	99,27
Zr	215,96	234,96	235,68	229,85	229,46	219,65	229,07	226,21

*Al, Ca, Fe, K, S e Ti em dag/kg e Cr, P, Pb, V e Zr em ppm. *Random Forest-RF, Ridge regression-RIDGE, Partial Least Squares-PLS, Regressão por componentes principais-PCR, Adaptive Forward-Backward Greedy Algorithm-FOBA, Generalized Boosted Regression Models- GBM e Gradient Boosting with Component-wise Linear Models – GLMBOOST*

Dentro os preditores testados, RF é o mais comumente utilizado nos trabalhos encontrados na literatura, sendo usado para prever diversas propriedades do solo (CAMERA et al., 2017). Essas propriedades, no entanto, são na maior parte das vezes relacionadas a textura (areia, silte, argila) (MA et al., 2017; VAYSSE e LAGACHERIE, 2015; HEUVELINK et al., 2016; CHAGAS et al., 2016; LAGACHERIE et al., 2013; BISHOP et al., 1999), ou a química com foco no pH (MA et al., 2017; DHARUMARAJAN et al., 2017; HEUVELINK et al., 2016; TOR et al., 2016; VAYSSE E LAGACHERIE, 2015; MALONE et al., 2014; BISHOP et al., 1999), carbono orgânico (RAMIFEHIARIVO et al., 2017; MA et al., 2017; DHARUMARAJAN et al., 2017; AKPA et al., 2016; SREENIVAS et al., 2016; HEUVELINK et al., 2016; TOR et al., 2016; GUO et al., 2015; ADHIKARI et al., 2014; POGGIO et al., 2013; WIESMEIER et al., 2011; GOMEZ et al., 2008; GRIMM et al., 2008; BISHOP et al., 1999), condutividade elétrica (DHARUMARAJAN et al., 2017; VAYSSE E LAGACHERIE, 2015; BISHOP et al., 1999) e capacidade de troca catiônica (TOR et al., 2016; LAGACHERIE et al., 2013; CHAGAS et al., 2016).

Hengl et al. (2015) encontraram valores negativos de R^2 ao mapear sódio trocável usando regressão linear, estes observaram ganho no R^2 ao comparar o mapeamento de bases trocáveis, especialmente Ca e Mg, feito com RF em relação à regressão. Não foram encontradas na literatura, predições realizadas pelos métodos testados nesse trabalho em relação aos elementos obtidos por fluorescência de raios-x.

Nas Tabelas 34 e 35 são apresentadas as oito principais variáveis usadas nos modelos de predição, nas camadas 0-10 e 10-30 cm, respectivamente. Para a camada 0-10 cm, metade dos elementos mapeados (Al, Fe, K, Pb e S), consideraram os dados geofísicos como variáveis importantes. Podendo ser citadas: razão entre Urânio e Potássio (Al, Pb); Razão entre Tório e Potássio (Al, S); Urânio (Pb); Potássio (Al, S); Tório (K); Sinal analítico do campo magnético total (Al, Fe, Pb).

E para a camada 10-30 cm, a exceção de Fe, todos os elementos consideraram os dados geofísicos importantes. Podendo-se citar: Razão entre Urânio e Potássio (Al, Ca); Razão entre Tório e Potássio (Ca, S); Razão entre Urânio e Tório (K); Urânio (Zr); Potássio (Al, Cr, S); Tório (K, Ti, V); Campo magnético Total (Cr); Sinal analítico do campo magnético total (Cr); Taxa de exposição do campo magnético total (P).

O campo magnético surge devido ao fluxo magnético, e os minerais apresentam diferentes comportamentos em relação a ele, desde uma fraca intensidade de magnetização como quartzo e calcita; até elevada susceptibilidade magnética como magnetita e óxidos de ferro (GONÇALVES, 2008). E o sinal analítico consiste na amplitude dos gradientes do campo magnético.

A contagem total de urânio, potássio e tório são dadas pela emissão de raios gama por esses elementos capturadas por gamaespectrômetros com alteração apenas da faixa energética captada pelo equipamento para cada canal. A taxa de contagem total é usada para assinalar uma escala de intensidade de cada elemento, e é a medida de todos os raios gama dentro da janela energética de 0,41 a 2,81 MeV (RIBEIRO et al., 2013). As razões, U/Th, U/K e Th/K, são bastante empregadas nas análises do ambiente por representarem melhor que os radioisótopos as mudanças nos tipos de rocha e outras alterações (GONÇALVES, 2008).

Em relação as variáveis criadas pelas relações de distância, bandas e dados geofísicos, essas foram amplamente apontadas por todos os elementos como covariáveis importantes na predição. As razões entre as bandas foram amplamente utilizadas (K, Mn, V e Zn em 0-10 cm e Al, Cr, K, Ti e V em 10-30 cm), assim como as razões entre K, U e Th (Al, K, Pb e S em 0-10 cm e Al, K e S em 10-30 cm).

Exceto Al e K na camada 0-10 cm e K e Pb em 10-30 cm, as relações de distância (x, y, d1, d2, d3, d4), foram relacionadas como principais, sendo que para Ti e P foram as únicas escolhidas na primeira camada. De acordo com McBratney et al. (2003), desde o final da década de 1960, houve uma ênfase no que poderia ser chamado de abordagens geográficas ou puramente espaciais, ou seja, os atributos do solo poderiam ser preditos a partir da sua posição espacial, principalmente pela interpolação entre eles.

Essas abordagens puramente espaciais são quase que inteiramente baseadas em geoestatística, como os métodos de krigagem e co-krigagem, sendo um dos seus problemas os limites não naturais que estes estabelecem no mapa (McBRATNEY et al., 2003). Esses limites pouco naturais foram observados nesse trabalho, principalmente para aqueles elementos, cuja as variáveis espaciais foram as mais relevantes, aparecendo nas primeiras colocações ou exclusivamente, como Ti e P na camada 0-10 cm (Figuras 31 e 32) e S na camada 10-30 cm (Figura 35).

Quando há inserção de outras variáveis preditoras não espaciais, suavizam-se esses limites, como no caso de V e Fe em 0-10 cm (Figuras 31 e 32) e Ti em 10-30 cm (Figura 35). Os valores de R^2 satisfatórios, variando de 0,38 a 0,54 (0-10 cm) e de 0,32 a 0,62 (10-30 cm), mostram a relevância de se considerar essas relações e associá-las aos atributos do terreno para predição.

Nesse trabalho, os atributos do terreno considerados como variáveis preditoras mais relevantes (Tabelas 35 e 36), foram: *Terrain Surface Convexity* (Al em 0-10 cm); *Valley Depth* (Al, Mn e Zn em 0-10 cm e Ca e P em 10-30 cm); *Real surface Area* (Al em 0-10 cm e Ca, Fe e Pb em 10-30 cm); *Terrain Surface Texture* (K, mn e V em 0-10 cm); *Diurnal Anisotropic* (K e S em 0-10 cm); *Convergence Index* (K em 0-10 cm e Fe, Cr e K em 10-30 cm); *Standardized Height* (Mn em 0-10 cm); *Mid Slope Positon* (Mn e Zn em 0-10 cm e Ca e K em 10-30 cm); *Topographic Wetness Index* (Mn e Zn em 0-10 cm e Ca, K, Pb e Zr em 10-30 cm); *Slope Height* (Pb em 0-10 cm); *Diffuse Insolation I* (Al e S em 0-10 cm e Al, Cr e P em 10-30 cm); *Direct to diffuse ratio1* (S em 0-10 cm); *Surface Specific Points* (Cr em 10-30 cm); *Slope* (Fe, Pb em 10-30 cm); MDE (S em 10-30 cm).

A Convexidade da superfície terrestre (*Terrain Surface Convexity*) é a razão entre o número de células que tem curvatura positiva (células convexas) e o número de todas as células válidas dentro de um raio de pesquisa específico; Profundidade dos vales (*Valley Depth*) converte a elevação, deriva as redes de drenagem e calcula a distância vertical até os mesmos; Área real da superfície (*Real Surface Area*) considera a superfície real ao contrário da projetada.

Textura da superfície terrestre (*Terrain Surface Texture*) divide a textura da superfície em 8, 12 ou 16 classes; Aquecimento anisotrópico diurno (*Diurnal Anisotropic*) representa a assimetria no aquecimento comparando as faces leste e oeste locais, baseado em horas de radiação.; Índice de convergência (*Convergence index*) calcula o índice de convergência/divergência em relação ao escoamento superficial.

Elevação Padronizada (*Standardized Height*) é o produto da altura normalizada multiplicada pela altura absoluta; Posição de Meia Encosta (*Mid Slope Positon*) representa a distância em relação ao topo e vale, variando entre 0 e 1 cobrindo as regiões de maior declividade; Índice de umidade topográfica (*Topographic Wetness Index*) mostra a tendência de cada célula em acumular água; Elevação da declividade (*Slope Height*) é a altura total do declive.

Radiação solar difusa 1 (*Diffuse Insolation I*) consiste na radiação solar difusa incidente no mês de janeiro; Radiação solar direta 1 (*Direct to Diffuse Ratio I*) na radiação solar direta incidente no mês de janeiro; Superfície de pontos específicos (*Surface Specific Points*) é a diferença entre pontos específicos de mudança da superfície,

resultando em valores discretos positivos e negativos; Declividade (*Slope*) representa a declividade local que pode ser angular ou em porcentagem e o Modelo Digital de Elevação (MDE) representa a elevação de cada célula do modelo.

De acordo com Grimm et al. (2008) os atributos do terreno têm potencial para explicar grande parte da variação do carbono orgânico. Hengl et al., 2015 mapeando bases trocáveis do solo, encontraram que a variação nessas foram melhor preditas pelos mapas globais dessas mesmas variáveis, entretanto covariáveis locais sobre relevo e vegetação ajudaram a melhorar a precisão. Encontraram, por exemplo, que as covariáveis elevação, declividade e Índice de Umidade Topográfica foram importantes para prever Mg trocável no solo; tendo também os tipos de solo se mostrado bons preditores de bases trocáveis.

Fe, P e Ti na camada 0-10 cm e Ti e V em 10-30 cm não usaram para predição nenhum atributo do terreno. Modelos não parcimoniosos foram encontrados para Al e Mn na camada 0-10 cm e para K em 10-30, independente da camada estes usaram mais de 30 variáveis na predição, sendo 50 para Al, 40 para Mn e 60 para K. Os modelos mais simplificados foram obtidos em 0-10 cm para as variáveis que selecionaram apenas atributos espaciais (P e Ti), na camada 10-30 cm foi para Zr que além de variáveis espaciais usou a covariável do terreno Índice de Umidade Topográfica (*Topographic Wetness Index*) e o Urânio proveniente da magnetometria.

Tabela 34- Principais variáveis independentes e número total de variáveis usadas pelos modelos de predição de atributos químicos do solo na profundidade 0-10 cm

Variável predita	Método	Nº de variáveis	Variáveis
Al	GLMBOOST	50	Razão entre Urânio e Potássio +SA do Campo Magnético Total + Razão entre Urânio e Potássio ¹ + Terrain Surface Convexity+ Potássio + Valley Depth + Razão entre Tório e Potássio + Real Surface Area
Fe	RF	6	d3 + y + d2 + d4 + Urânio + SA do Campo Magnético Total
K	GBM	12	Razão entre B2 e B6 Landsat 8 + Razão entre B4 e B6 Landsat 8 +Razão entre Tório e Urânio ¹ + Razão entre B3 e B6 Landsat 8 + Tório + Terrain Surface Texture + Diurnal Anisotropic + Convergence index
Mn	RF	40	Terrain Surface Texture + Valley Depth + Razão entre B3 e B7 Landsat 8 + Standardized Height + Mid Slope Position + y + Razão entre B1 e B3 Landsat 8 + Topographic Wetness Index
P	PLS / PCR	5	d3 + y + d2 + d4 + d1
Pb	CUBIST	10	y + Razão entre Urânio e Potássio + d4 + SA do Campo Magnético Total + Razão entre Urânio e Potássio ¹ + d3 + Urânio + Slope Height
S	RIDGE	7	d3 + Razão entre Tório e Potássio + Diurnal Anisotropic + Diffuse Insolation ¹ + Direct to Diffuse Ratio ¹ + Razão entre Tório e Potássio ¹ + Potássio
Ti	RIDGE/CUBIST/ PLS/PCR/ FOBA	5	y + d4 + d3 + d1 + d2
V	RF	7	y + d4 + d3 + d2 + d1 + Terrain Surface Texture + Razão entre B2 e B3 Landsat 8
Zn	RIDGE	7	Valley Depth + Topographic Wetness Index + Mid Slope Position + Razão entre B4 e B6 Landsat 8 + Razão entre B1 e B6 Landsat 8 + Razão entre B2 e B6 Landsat 8 + d3

¹Variável obtida pela razão entre as diferenças nos teores de K, Th e U conforme indicado em material e métodos. *Random Forest*-RF, *Ridge regression*-RIDGE, *Partial Least Squares*-PLS, *Regressão por componentes principais*-PCR, *Adaptive Forward-Backward Greedy Algorithm*-FOBA, *Generalized Boosted Regression Models*- GBM e *Gradient Boosting with Component-wise Linear Models* – GLMBOOST

Tabela 35- Principais variáveis independentes e número total de variáveis usadas pelos modelos de predição de atributos químicos do solo na profundidade 10-30 cm

Variável predita	Método	Nº de variáveis	Variáveis
Al	RF	15	Razão entre B3 e B4 Landsat 8 + Razão entre B2 e B4 Landsat 8 + Diffuse Insolation1 + Potássio + Razão entre B1 e B4 Landsat 8 + Razão entre Urânio e Potássio + y + Razão entre Urânio e Potássio ¹
Ca	RF	20	Real Surface Area + Valley Depth + d3 + Topographic Wetness Index + Slope + Mid Slope Position + Razão entre Tório e Potássio + Razão entre Urânio e Potássio
Cr	FOBA	10	Convergence index + Surface Specific Points + Campo Magnético Total + 1ª Derivada do Campo Magnético + Diffuse Insolation1 + d3 + Potássio + Razão entre B1 e B9 Landsat 8
Fe	RF	20	d3 + y + Real Surface Area + Slope + d2 + d4 + Convergence Index + d1
K	RF	60	Razão entre Tório e Potássio ¹ + Razão entre Urânio e Tório + Razão entre Urânio e Potássio ¹ + Convergence index + Tório + Topographic Wetness Index + Razão entre B3 e B6 Landsat 8 + Mid Slope Position
P	RF	12	d3 + y + d4 + d2 + Taxa de Exposição da Contagem Total + Valley Depth + d1 + Duration of Insolation1
Pb	RF	25	d3 + d2 + Real Surface Area + Convergence Index + y + Slope + Topographic Wetness Index
S	PLS / PCR	20	y + Razão entre Tório e Potássio + MDE + d3 + Potássio + Razão entre Tório e Potássio ¹ + x + d4
Ti	GBM	8	y + d4 + d3 + d1 + d2 + Razão entre B1 e B3 Landsat 8 + Razão entre B2 e B3 Landsat 8 + Tório
V	GBM	10	y + d4 + d2 + d1 + Razão entre B1 e B3 Landsat 8 + Tório + Razão entre B2 e B3 Landsat 8
Zr	RF	5	Topographic Wetness Index + y + d4 + Urânio + d1

¹Variável obtida pela razão entre as diferenças nos teores de K, Th e U conforme indicado em material e métodos. *Random Forest*-RF, *Ridge regression*-RIDGE, *Partial Least Squares*-PLS, Regressão por componentes principais-PCR, *Adaptive Forward-Backward Greedy Algorithm*-FOBA, *Generalized Boosted Regression Models*- GBM e *Gradient Boosting with Component-wise Linear Models* – GLMBOOST

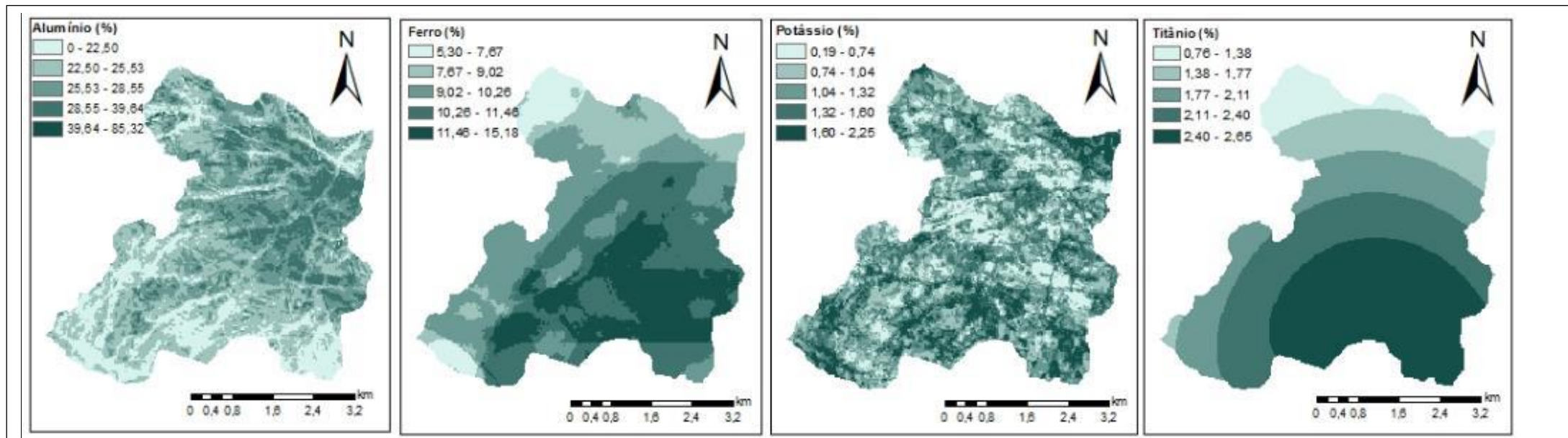


Figura 31- Teores totais de Alumínio (Al_2O_3), Ferro (Fe_2O_3), Potássio (K_2O) e Titânio (TiO_2) usando os regressores GLMBOOST (Al), RF (Fe), GMB (K) e RIDGE (Ti), em 0-10 cm. *Random Forest-RF, Ridge regression-RIDGE, Generalized Boosted Regression Models- GBM e Gradient Boosting with Component-wise Linear Models – GLMBOOST.*

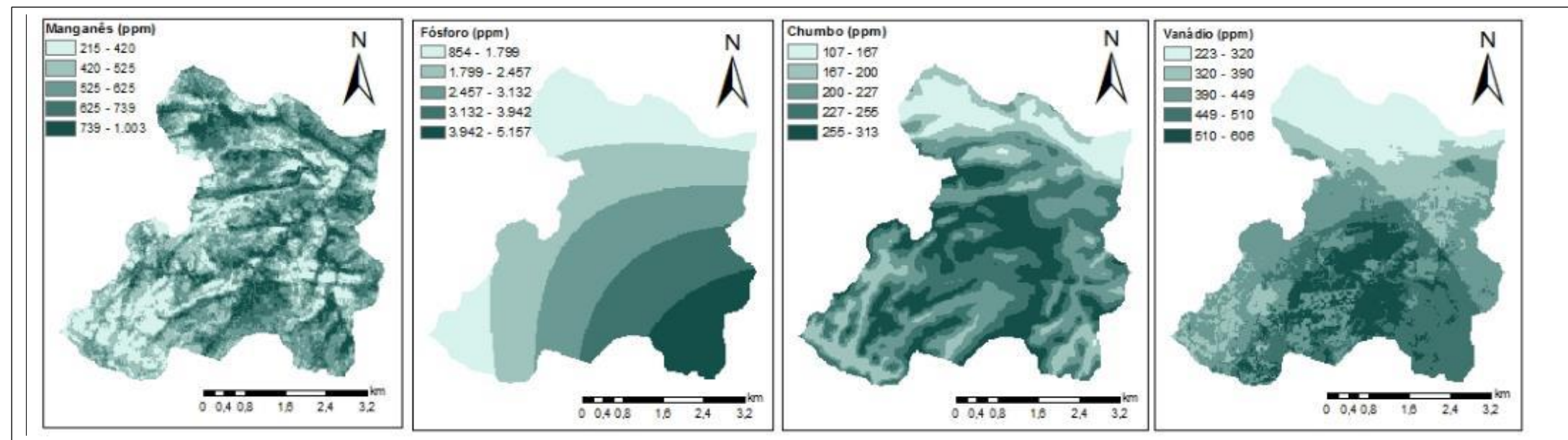


Figura 32- Teores totais de Manganês (MnO), Fósforo (P_2O_5), Chumbo (PbO) e Vanádio (V_2O_5) usando regressores RF (Mn e V), Cubist (Pb), PLS 101 (P) em 0-10 cm. *Random Forest-RF, Partial Least Squares-PLS.*

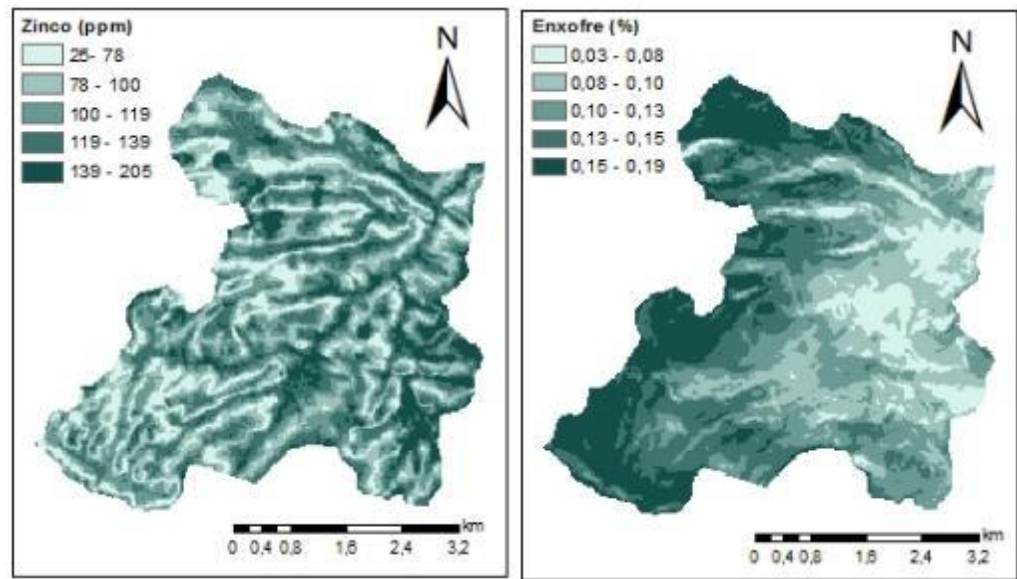


Figura 33- Teores totais de Zinco (ZnO), Enxofre (SO₃) usando o regressor RIDGE em 0-10 cm. *Ridge regression-RIDGE*.

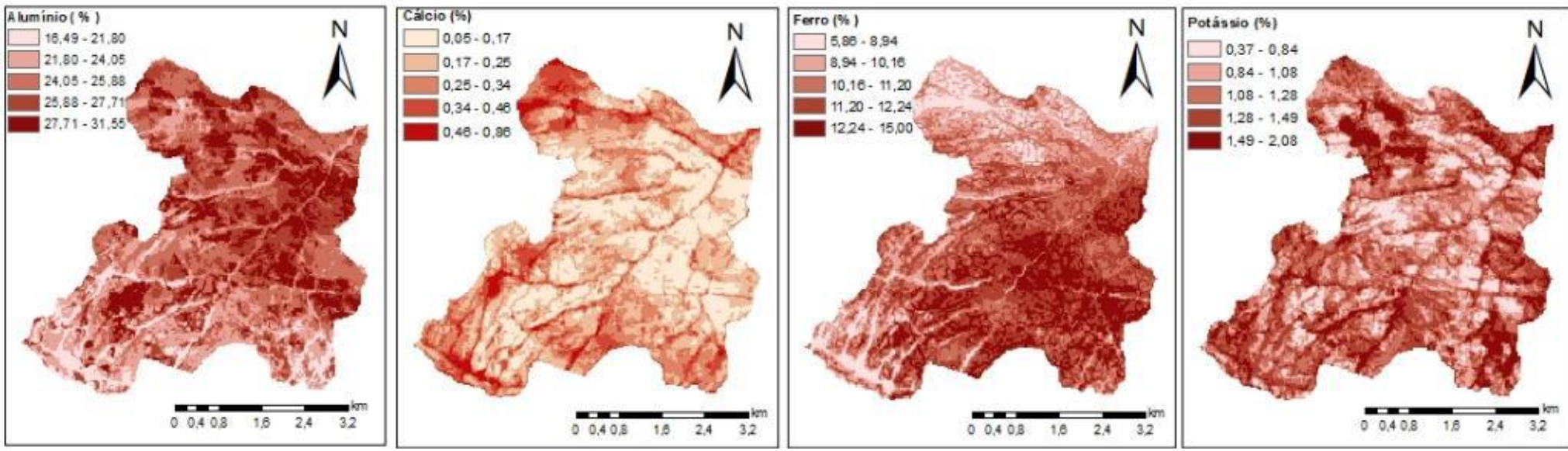


Figura 34- Teores totais de Alumínio (Al₂O₃), Cálcio (CaO), Ferro(Fe₂O₃) e Potássio (K₂O) usando o regressor Random Forest em 10-30 cm.

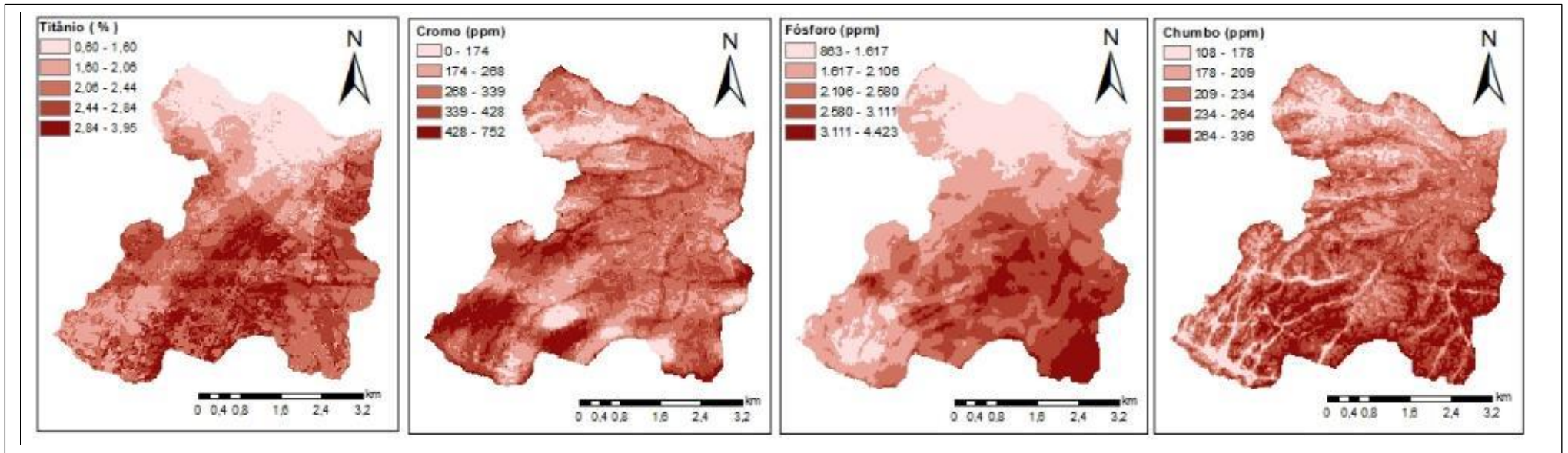


Figura 35- Teores totais de Titânio (TiO_2), Cromo (Cr_2O_3), Fósforo (P_2O_5) e Chumbo (PbO) usando os regressores GBM (Ti), FOBA (Cr), RF (P e Pb) em 10-30 cm. *Random Forest-RF, Adaptive Forward-Backward Greedy Algorithm-FOBA, Generalized Boosted Regression Models- GBM.*

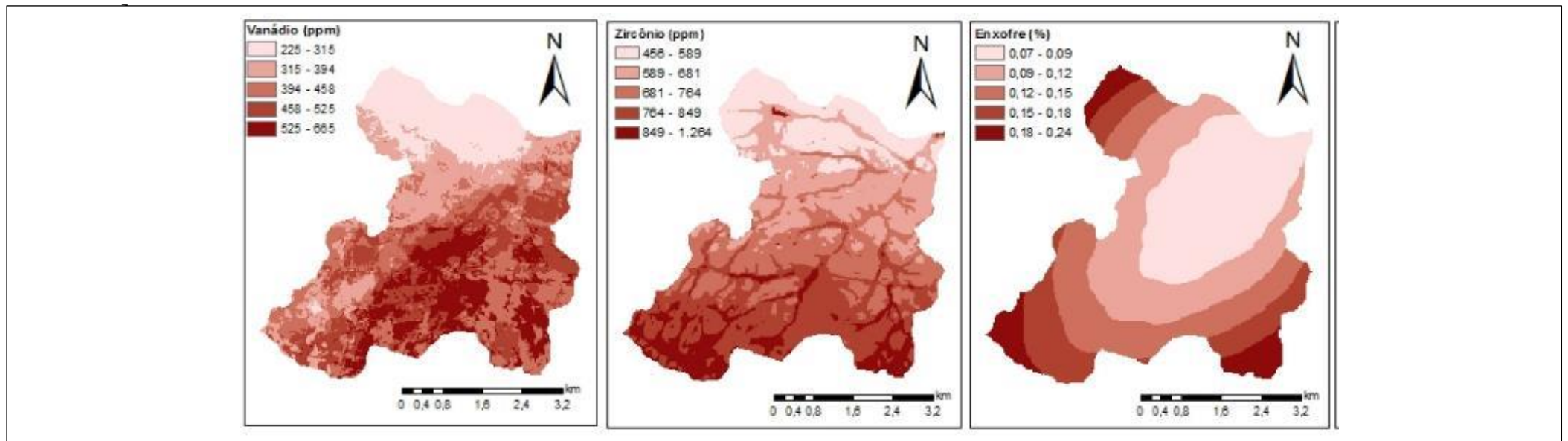


Figura 36- Teores totais de Vanádio (V_2O_5), Zircônio (ZrO_2) e Enxofre (SO_3), usando os regressores GBM (V), RF (Zr) e PLS (S) em 0-30 cm. *Random Forest-RF, Partial Least Squares-PLS, Generalized Boosted Regression Models- GBM.* 103

Dentre os valores máximos e mínimos estabelecidos pelos melhores preditores de cada elemento (Tabela 36) pode-se perceber que a maioria se manteve próximo ao encontrado pelas análises da fluorescência. Na camada 0-10 cm o Al apresentou valores mínimos subestimados enquanto o Mn, o P e o Pb foram superestimados. Quanto aos valores máximos, o Al foi superestimado, enquanto Mn subestimado.

Na profundidade 10-30 cm (Tabela 37), pode-se observar que a maior parte dos elementos também se manteve dentro de uma faixa esperada, apresentando apenas Cr e P com valores de mínimo abaixo do encontrado nas amostras e valores de máximo acima.

Valores de K total nas duas profundidades avaliadas foram baixos considerando-se que trata-se de um macro-nutriente encontrado nos solos com teores podendo variar de 300 (SOUZA et al., 1979) a 39.500 mg/kg (NACHTIGALL e VAHL, 1991), dependendo do material de origem, além do grau de intemperização e composição mineralógica (ERNANI et al., 2007). A geologia local, pode explicar esse fato, visto que a região estudada é formada basicamente por metatonalitos e, segundo Pinho et al. (2005), essas são rochas de filiação cálcio-alcalina de baixo teor de potássio, o que pode influenciar nos baixos teores encontrados nas análises dos solos e na predição.

Tabela 36- Valores máximos e mínimos encontrados para Al, Fe, K, Mn, P, Pb, S, Ti, V e Zn na análise de raios-x e pelos modelos de predição, na profundidade 0-10 cm

Elemento	Raios-x		Predição		Método de predição
	Valor mínimo	Valor máximo	Valor mínimo	Valor máximo	
Al	12,70	35,45	0,00	85,32	GLMBOOST
Fe	3,51	17,73	5,30	15,18	RF
K	0,14	3,21	0,19	2,25	GBM
Mn	53,50	2576,98	215,00	1003,00	RF
P	0,00	5867,08	854,00	5157,00	PLS / PCR
Pb	57,27	440,54	107,00	313,00	CUBIST
S	0,03	0,30	0,03	0,19	RIDGE
Ti	0,97	3,45	0,76	2,65	RIDGE/CUBIST/PLS/PCR/ FOBA
V	210,05	719,71	223,00	606,00	RF
Zn	28,68	232,19	25,00	205,00	RIDGE

*Al, Fe, K, S, Ti em dag/kg e Mn, P, Pb, V e Zn em ppm. *Random Forest*-RF, *Ridge regression*-RIDGE, *Partial Least Squares*-PLS, *Regressão por componentes principais*-PCR, *Adaptive Forward-Backward Greedy Algorithm*-FOBA, *Generalized Boosted Regression Models*- GBM e *Gradient Boosting with Component-wise Linear Models* – GLMBOOST

Tabela 37- Valores máximos e mínimos encontrados para Al, Ca, Cr, Fe, K, P, Pb, S, Ti, V e Zr na análise de raios-x e pelos modelos de predição na profundidade 10-30 cm

Elemento	Raios-X		Predição		Método de predição
	Valor mínimo	Valor máximo	Valor mínimo	Valor máximo	
Al	13,44	33,86	16,49	31,55	RF
Ca	0,01	1,33	0,05	0,86	RF
Cr	18,12	541,42	0,00	752,00	FOBA
Fe	3,31	18,63	5,86	15,00	RF
K	0,10	3,08	0,37	2,08	RF
P	0,00	6850,07	863,00	4423,00	RF
Pb	60,99	1101,56	108,00	336,00	RF
S	0,01	0,24	0,07	0,24	PLS / PCR
Ti	0,93	3,79	0,60	3,95	GBM
V	183,15	745,18	225,00	665,00	GBM
Zr	249,38	1844,85	456,00	1264,00	RF

*Al, Ca, Fe, K, S e Ti em dag/kg e Cr, P, Pb, V e Zr em ppm. *Random Forest*-RF, *Ridge regression*-RIDGE, *Partial Least Squares*-PLS, Regressão por componentes principais-PCR, *Adaptive Forward-Backward Greedy Algorithm*-FOBA, *Generalized Boosted Regression Models*- GBM e *Gradient Boosting with Component-wise Linear Models* – GLMBOOST

5 CONCLUSÕES

1. O algoritmo Random Forest mostrou-se superior aos demais modelos para predizer os atributos químicos do solo nas duas profundidades, principalmente em 10-30 cm, com superioridade nos valores de R^2 em sete elementos, sendo, portanto, indicado para predição nessa região;
2. Variáveis espaciais mostraram-se indispensáveis para predição de atributos do solo, devendo ser consideradas nas modelagens de seus elementos químicos;
3. É possível a predição de Al, Ca, Cr, Fe, K, Mn, P, Pb, S, Ti, V, Zn e Zr com R^2 variando de 0,32 a 0,62 pelos métodos testados.

6 REFERÊNCIAS BIBLIOGRÁFICAS

ABREU, VS. Uso de georadar e mapeamento digital de solos no estudo de mussunungas em área do Grupo Barreiras. [Dissertação]. Universidade Federal de Viçosa-UFV, Viçosa, 166 p. Fev de 2017.

ALVES SOBRINHO, T.; PERTUSSATTI, C. A.; REBUCCI, L.C. S.; OLIVEIRA, P. T. S. Estimativa da erosividade local das chuvas, utilizando redes neurais artificiais. **Revista Ambiente e Água**, Taubaté, v. 6, n. 2, p.246-254, 2011.

ANDERSEN, C.; BRO, R. Variable selection in regression- a tutorial. *Journal of Chemometrics*. v.24, p. 728-737. 2010.

ANDRADE, F.; SANTOS, J.; CARNEIRO, L.; PAULO, R. Notas de aula. Disponível em: <http://paginas.fe.up.pt/~eol/IA/IA0708/APONTAMENTOS/IAlogica_conjuntos_di_fusos.pdf> . Acesso em : Ago, 2014.

ANDRIOLO, M.V. Análise de Métodos não Lineares para Previsão de Vazões Médias Mensais. Dissertação (mestrado). Universidade Federal do Paraná. UFP. 248 p. 2006.

ARRUDA, G. O; DEMATTÊ, J. A. M. CHAGAS, C. S. Mapeamento digital de solos por redes neurais artificiais com base na relação solo-paisagem. **Revista Brasileira Ciência do Solo**, v. 37, n.2, p.327-338, 2013.

ASHTEKAR, J.M.; OWENS, P.R. Remembering Knowledge: Na expert knowledge based approach to digital soil mapping. **Soil Horizons**. v. 54, n 5, p. 1-6, 2013.

AZEVEDO, C.F.; RESENDE, M.D.V. DE; SILVA, F.F. E; LOPES, P.S; GUIMARÃES, S.E.F. Regressão via componentes independentes aplicada à seleção genômica para características de carcaça em suínos. *Pesquisa Agropecuária Brasileira*, [s.l.], v. 48, n. 6, p.619-626, jun. 2013.

BARIN, A.; CANHA, L.N. MAGNAGO, K.F.; ABAIDE, A.R. Seleção de fontes alternativas de geração distribuída utilizando uma análise multicriterial baseada no método ahp e na lógica fuzzy. **Revista Controle & Automação**. v.21 n.5, p.477-486 2010.

BARRETO-NETO, A. A.; SOUZA FILHO, C. R. de. Application Of Fuzzy Logic To The Evaluation Of Runoff In A Tropical Watershed. **Environmental Modeling & Software**, v23, n1, p.244-253. 2008.

BENINI, L.C. Modelagem neuro-fuzzy para precisão da densidade do solo. Anais do XXXIV Congresso Nacional de Matemática aplicada e computacional. p.856-857. 17 a 21 de Set, 2012.

BHERING, S.B.; CHAGAS, C. da S.; CARVALHO JUNIOR, W. de; PEREIRA, N.R.; CALDERANO FILHO, B.. Influência do modelo digital de elevação no mapeamento digital de atributos do solo por random forest (RF). Congresso brasileiro de Ciência do solo. 2015. Disponível em < <http://www.cbcs2015.com.br/anais/index.php#menuanais> > Acesso em mai, 2016.

BÖHNER, J., MCCLOY, K.R., STROBL, J. [Eds.]: SAGA – Analysis and Modelling Applications. *Göttinger Geographische Abhandlungen*, v.115, 130p. 2006.

BONISCH, S.; LOPES ASSAD, M.L.; CÂMARA, G.; MONTEIRO, A.M.V. Representação e propagação de incertezas em dados solos. **Revista Brasileira de Ciência do Solo**. v.28, n.1, p.21-32, 2004.

BRIAND, L. C.; FREIMUT, B.; VOLLEI, F. Using multiple adaptive regression splines to support decision making in code inspections. *The Journal of Systems and Software*. v.73, p. 205-217. 2004.

- BUEHLMANN, P. Boosting for high-dimensional linear models. *The Annals of Statistics*, 34(2), 559–583.2006.
- BUEHLMANN, P.; HOTHORN, T. Boosting algorithms: regularization, prediction and model fitting. *Statistical Science*, 22(4), 477–505.2007.
- BUEHLMANN, P.; YU, B. Boosting with the L2 loss: regression and classification. *Journal of the American Statistical Association*, 98, 324–339. 2003.
- BURROUGH, P. A. Fuzzy mathematical methods for soil survey and land evaluation. **Journal of Soil Science**, v. 40, n.3,p. 477-482, 1989
- CAMERA C, ZOMENI Z, NOLLER JS, ZISSIMOS AM, CHRISTOFOROU IC, BRUGGEMAN A. A high resolution map of soil types and physical properties for Cyprus: A digital soil mapping optimization. *Geoderma*, [s.l.], v. 285, p.35-49, jan. 2017. Elsevier BV. <http://dx.doi.org/10.1016/j.geoderma.2016.09.019>.
- CARRÉ, F.; GIRARD, M. C. Quantitative mapping of soil types based on regression kriging of taxonomic distances with landform and cover attributes. *Geoderma*, Amsterdam. v. 110, p. 241-263, 2002.
- CARVALHO JÚNIOR, W. de; CHAGAS, C. da 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**, v.68, n.6, p.691-696, 2011.
- CARVALHO, A. de P.; MORAES NETO, J. M. de; LIMA, V. L. A. de; & SILVA, D. G.K. C. Estudo da degradação ambiental do açude Bodocongó. **Engenharia Ambiental - Espírito Santo do Pinhal**, v.6,n.2, p. 293-305, 2009.
- CATEN, A. T . **Mapeamento digital de solos: metodologias para atender a demanda por informação espacial em solos**. 106 p . 2011a. (Tese de Doutorado). Universidade Federal de Santa Maria.
- CATEN, A.T.; DALMOLIN, R.S.D; PEDRON, F.A.; SANTOS, M.L.M. Extrapolação das relações solo-paisagem a partir de uma área de referência. **Ciência Rural**, v.41, n.5,p. 812-816, 2011b.
- CHAGAS, C. da S , CARVALHO JUNIOR, W; BHERING, S.B; CALDERANO FILHO, B. Spatial prediction of soil surface texture in a semiarid region using random forest and multiple linear regressions. *Catena*. v.139, p.232–240, 2016.
- CHAGAS,C. DA S. **Mapeamento digital de solos por correlação ambiental e redes neurais em uma bacia hidrográfica de domínio de mar de morros**. 223 p. 2006 (Tese de Doutorado).Universidade Federal de Viçosa.
- CHATTERJEE, S. & HADI, A.S. *Regression analysis by example*. 4.ed. New York, John Willey & Sons, 375p. 2006.
- CLÉSIO, F. Qual a diferença entre lasso e ridge regression?.2015. Disponível em< <https://mineracaodedados.wordpress.com/2015/06/20/qual-a-diferenca-entre-lasso-e-ridge-regression/>> Acesso em Agp, 2017.

CONGALTON, R.G.; GREEN, K. Assessing the accuracy of remotely sensed data: Principles and practices. **International Journal of Applied Earth Observation and Geoinformation**, v.11, n.2, p.448-449, 1999.

COSTA, A. M.; CURI, N.; MENEZES, M. D.; ARAÚJO, E. F.; MARQUES, J. J. Levantamento detalhado de solos da microbacia hidrográfica do Horto Florestal Terra Dura (RS) e considerações sobre escala de mapeamento. **Ciência e Agrotecnologia**, Lavras, v. 33, n. 5, p. 1272-1279, set./out., 2009

CULE, E.; DE IORIO, M. A semi-automatic method to guide the choice of ridge parameter in ridge regression. 2012. Disponível em < <https://arxiv.org/abs/1205.0686>> Acesso em Ago, 2017.

DAVIES, B.E., GAMM, S.A. Trend surface analysis applied to soil reaction values from Kent, England. *Geoderma* 3, 223– 231. 1969.

DEBELLA-GILO, M. & ETZELMÜLLER, B. Spatial prediction of soil classes using digital terrain analysis and multinomial logistic regression modeling integrated in GIS: Examples from Vestfold County, Norway. *Catena*. v. 77, p.8-18, 2009.

DEMATTÊ, J. A. M.; GENÚ, A. M.; FIORIO, P. R.; ORTIZ, J. L.; MAZZA, J. A.; LEONARDO, H. C. L. Comparação entre mapas de solos obtidos por sensoriamento remoto espectral e pelo método convencional. *Pesquisa Agropecuária Brasileira*, Brasília, v. 39, n. 12, p.1219-1229, 2004.

DOKUCHAEV, V. V. Russian chernozem: complete set of works. Moscow: Academy of Sciences Publisher. p. 404. v. 3. 1949.

DOZUALDO, D.G; REZENDE, S.O. Análise da Precisão de Métodos de Regressão. Relatórios técnicos. Universidade de São Paulo-USP. Instituto de Ciências Matemáticas e de computação. n 197.45 p. 2003.

ELITH, J.; LEATHWICK, J.R.; HASTIE, T. A working guide to boosted regression trees. *J Anim Ecol*. 77(4):802-13. 2008.

ELNAGGAR, A. A. **Development of Predictive Mapping Techniques for Soil Survey and Salinity Mapping**. 148 p. 2007. Thesis (PhD) – Oregon State University.

EMBRAPA. EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA -Centro Nacional de Pesquisa de Solos. **Sistema brasileiro de classificação de solos**. 2.ed. Rio de Janeiro, Embrapa Solos, 2006. 316p.

ERNANI PR. ; ALMEIDA JA DE; SANTOS FG DOS. Potássio. In: NOVAIS, R. F.; ALVAREZ V.. V. H.; Barros, N.F.; Fontes, R.L.F.; Cantarutti, R.B.; Neves, J.C. (Ed.) *Fertilidade do Solo*. Viçosa. Sociedade Brasileira de Ciência do Solo. v. 1, p.551-594, 2007.

FIGUEIREDO, S.R.; GIASSON, E. TORNQUIST, C.G.; NASCIMENTO, P.C. Uso de regressões logísticas múltiplas para mapeamento digital de solos no planalto médio do RS. **Revista Brasileira de Ciência do Solo**. v. 32, n. Spe, p.2779-2785, 2008.

FREUND, Y.; SCHAPIRE, R.E. A decision-theoretic generalization of on-line learning and na application to boosting. *Journal of Computer and System Sciences*, 55(1):119-139. 1997.

FRIEDMAN, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics* 29(5):1189-1232.2001.

FUJIMOTO, R.Y. **Diagnóstico automático dde defeitos em rolamentos baseado em Lógica Fuzzy**. 158 p. 2005. (Dissertação de Mestrado). Escola Politécnica, Universidade de São Paulo.

GARTHWAITE, P.H. An interpretation of partial least squares. *Journal of the American Statistical Association*, v.89, p.122- 127, 1994.

GENÚ, A.M. Sensoriamento remoto e geoprocessamento na caracterização e avaliação pontual e espacial de solos e seus atributos. Tese (Doutorado). Escola Superior de Agricultura Luiz de Queiróz-ESALQ/USP. Piracicaba. 120p. 2006.

GENUER, R., POGGI, J.M, MALOT, C.T.. Variable selection using random forests. *Pattern Recognition Letters*. v.31, p. 2225–2236, 2010.

GONÇALVES, B.F. Modelagem de dados geofísicos, magnéticos e radiométricos aéreos e terrestres aplicados à interpretação geológica. Trabalho de conclusão de curso. Graduação em Geofísica. Universidade Federal da Bahia. Salvador. 142p. 2008.

GRIMM, T. BEHRENS, M. MÄRKER, H. ELSENBEER. Soil organic carbon concentrations and stocks on Barro Colorado Island—Digital soil mapping using Random Forests analysis *R. Geoderma*. v. 146, p.102–113, 2008.

HENGL, T.; HEUVELINK, G.B.M, KEMPEN, B.; LEENAARS, J.G.B.; WALSH, M.G., SHEPHERD, K.D., SILA, A.; MacMILLAN, R.A.; JESUS, J. M. de; TAMENE, L.; TONDOH, J.E.. Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions . *PLOS ONE* . v. 25, p. 1-26, 2015.

HILT, D. E.; SEEGRIST, D. W. A computer program for calculating ridge regression estimates. Forest Service, United States, Dept. of agriculture, University of Minnesota, Northeastern Forest Experiment Station, 1-7, 1977.

HOFNER, B.; MAYR, A.; ROBINZONOV, N.; SCHMID, M. Model-based boosting in R: a hands-on tutorial using the R package mboost. *Computational Statistics*, [s.l.], v. 29, n. 1-2, p.3-35, 22 dez. 2012.

HOTHORN, T.; BUEHLMANN, P.; KNEIB, T.; SCHMID, M.; HOFNER, B.; SOBOTKA, F.; SCHEIPL, F.; MAYR, A. Model-Based Boosting. Package mboost. *Glmboost*. 2017. Disponível em <<https://cran.r-project.org/web/packages/mboost/mboost.pdf>>. Acesso em: Ago, 2017.

HUDSON, B.D. Concepts of soil mapping and interpretation. **Soil Survey Horizons**. v. 31,n.3 p.63-72, 1990.

INSTITUTO CAPIXABA DE PESQUISA E EXTENSÃO RURAL- INCAPER. Município de Iconha. Disponível em <<http://www.incaper.es.gov.br/institucional/downloads/iconha.pdf>>. Acesso em Maio, 2014.

KABATA-PENDIAS, A. Trace elements in soils and plants / Alina Kabata-Pendias. - 4th ed. 533 p.2010. CRC Press:EUA.

KISS, J.J., DE JONG, E., MARTZ, L.W. The distribution of fallout Cesium-137 in southern Saskatchewan, Canada. *Journal of Environmental Quality* 17, 445– 452.1988.

KUHN,M; WESTON, S.; KEEFER, C.; COULTER, N. C code for Cubist by Ross Quinlan. Package Cubist. 13p. 2016.

LAGACHERIE, P. Digital Soil Mapping: A State of the Art. In: HARTEMINK, A. E.; McBRATNEY, A.B., MENDONÇA-SANTOS M. L. (Eds) **Digital Soil Mapping with Limited Data**. New York: Springer. Ed. 1, cap.1, p. 3-14, 2008.

LANDIS, J. R. & KOCH, G. G. The measurement of observer agreement for categorical data.**Biometrics**, v.33, n.1, p.159-174, 1977.

LIAO, W.H. Region Description Using Extended local Ternary Patterns. **International Conference on Pattern Recognition**. p. 1003-1006, 2010.

LOBÃO, J. S. B.; ROCHA, W. de J. S. da F; SILVA, A. de B. Utilização De Lógica Fuzzy Na Modelagem De Vulnerabilidade À Erosão No Município de Morro Chapéu – BA. In: III SIMPÓSIO REGIONAL DE GEOPROCESSAMENTO E SENSORIAMENTO REMOTO, Anais, Aracaju –SE. 2006.

LOPES, K.M. de O. Modelos baseados em data mining para classificação multitemporal de culturas no Mato Grosso utilizando dados de NDVI/MODIS. Dissertação (mestrado). Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica. 2013.

MacMILLAN, R. A. Experiences with Applied DSM: Protocol, Availability, Quality and Capacity Building. In.: HARTEMINK, A.E.; McBRATNEY, A.; MENDONÇA-SANTOS,M. DE L. (Ed.), **Digital Soil Mapping with Limited Data**. New York: Springer. Ed 1, cap. 10. p.113-135, 2008.

MEIER, M.M. Zoneamento Pedoambiental em Assentamento da Reforma Agrária: Caso de Estudo – P.A. Denis Gonçalves, Zona da Mata Mineira. [Dissertação]. Universidade Federal de Viçosa, Viçosa, 122p. Jul de 2016.

Zoneamento Pedoambiental em Assentamento da Reforma Agrária: Caso de Estudo – P.A. Denis Gonçalves, Zona da Mata Mineira.

McBRATNEY, A. B.; MENDONÇA-SANTOS, M. L.; MINASNY, B.. On digital soil mapping. **Geoderma**. v.117, n.2.p.3-52. 2003.

McBRATNEY, A. B.; WEBSTER, R. Spatial dependence and classification of soil along a transect in northeast Scotland. **Geoderma**, v. 26,n.1, p. 63-82, 1981.

MCBRATNEY, A.b; SANTOS, M.l Mendonça; MINASNY, B. On digital soil mapping. *Geoderma*, [s.l.], v. 117, n. 1-2, p.3-52, nov. 2003.

MEDEIROS, S.; MELLO, R.; CAMPOS FILHO, P. Análise de projetos para unidades de conservação usando lógica fuzzy. **Produção**. v.17, n.2, p.317-329, 2007.

MENEZES, M. D.; SILVA, S. H. G.; OWENS, P. R.; CURI, N. Digital soil mapping approach based on fuzzy logic and field expert knowledge.**Ciência e Agrotecnologia**. 2013, vol.37, n.4, pp. 287-298.

- MENEZES, M.D.; SILVA, S.H.G., MELLO, C.R. de, OWENS, P.R., CURI, N. Solum depth spatial prediction comparing conventional with knowledge-based digital . *Sci. Agric.* v.71, n.4, p.316-323, 2014.
- MONSERUD, R. A. & LEEMANS, R. Comparing global vegetation maps with the Kappa statistic. **Ecological Modelling**, v.62,n.4, p.275-293, 1992.
- MOORE, I.D., GESSLER, P.E., NIELSEN, G.A., PETERSON, G.A., 1993. Soil attribute prediction using terrain analysis. *Soil Science Society of America Journal* 57, 443–452.
- MORAES, C.M.M. Fuzzycom- Componente de lógica fuzzy. Puc-Departamento de Engenharia Elétrica, p.1-15. Disponível em: < http://www.puc-rio.br/pibic/relatorio_resumo2007/relatorios/ele/ele_claudio_magno_martins_moraes.pdf >. Acesso em: Ago, 2014.
- MORELLATO, S. A. Modelos de regressão PLS com erros heteroscedásticos (Dissertação de mestrado). São Carlos : UFSCar-Universidade Federal de São Carlos, 49 p. 2010.
- MORITZ, S.; CULE, E. R Package Ridge. Disponível em: < <https://cran.r-project.org/web/packages/ridge/ridge.pdf> >. Acesso em Ago, 2017.
- NACHTIGALL GR E VAHL LC. Capacidade de suprimento de potássio dos solos da região sul do Rio Grande do Sul. *Revista Brasileira de Ciência do Solo*. 15:43-47, 1991.
- NEUMANN, M. R. B.. **Mapeamento digital de solos, no Distrito Federal**. 123 p. 2012. (Tese de Doutorado). Universidade de Brasília. Brasília.
- OLIVEIRA, S., OEHLER, F., AYANZ, J.S.M., CAMIA, A., PEREIRA, J.M.C. Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. *Forest Ecology and Management*. v. 275, p.117–129, 2012.
- PANDORFI, H.; SILVA, I.J.O; GUISELINI, C.; PIEDADE, S.M.S. Uso da lógica fuzzy na caracterização do ambiente produtivo para matrizes gestantes. **Engenharia Agrícola**. v.27, n.1, p.83-92, 2007.
- Panero, F. dos Santos. Quimiometria -Regressão por Componentes Principais. 2010. Disponível em < <http://quimiometria.pbworks.com/w/page/23745622/Quimiometria> >. Acesso em Ago, 2017.
- PARREIRA, T.F. Utilização de métodos quimiométricos em dados de natureza multivariada. (Dissertação de mestrado). Universidade Estadual de Campinas-Unicamp. Instituto de Química. Campinas-SP. 106 p. 2003.
- PINHEIRO, H.S.K.; CHAGAS, C.S.; CARVALHO JÚNIOR, W. ANJOS, L.H.C. Modelos de elevação para obtenção de atributos topográficos utilizados em mapeamento digital de solos. **Pesquisa Agropecuária Brasileira**. v.47, n.9, p.1384-1394, 2012.
- POLETO, L. J. C.; SILVA, C. H. C.; R. JÚNIOR S.; VIEIRA, C. A. O., SANTOS, A. P.. Validação de modelo digital de elevação hidrologicamente consistente (MDEHC). **II Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação**. Recife - PE, 8-11 de setembro de 2008.

PROJETO RADAMBRASIL. Folha SH 22, SH 21 e SI 22. Rio de Janeiro, IBGE. **Levantamento dos Recursos Naturais**, Série III, nº33. p.303-404, 1986.

QI, F.; ZHU, A.X.; HARROWER, M.; BURT, J.E. Fuzzy soil mapping based on prototype category theory. **Geoderma**. v. 136, n 1, p. 774-787, 2006.

QI, F.; ZHU, A.X. Knowledge Discovery from soil maps using inductive learning. **International Journal of Geographical Information Science**. v. 17, n 8, p. 771-795, 2003.

QUINLAN. C4.5: Programs For Machine Learning. Morgan Kaufmann Publishers Inc. San Francisco, CA.1993.

QUINLAN. Learning with continuous classes. Proceedings of the 5th Australian Joint Conference On Artificial Intelligence. p. 343-348.1992

R DEVELOPMENT CORE TEAM (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

REN-MIN YANG, GAN-LIN ZHANG, FENG LIU, YUAN-YUAN LU, FAN YANG, FEI YANG, MIN YANG, YU-GUO ZHAO, DE-CHENG LI. Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem. *Ecological Indicators*. v. 60, p. 870–878, 2016.

REYANALDO, C. Regressão "Ridge": Um Método Alternativo para o Mal Condicionamento da Matriz das Regressoras (Dissertação de mestrado). - Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. Campinas. 132 p. 1997.

RIBEIRO, V.B.; MANTOVANI, M.S.M.; LOURO, V.H.A.. Aerogamaespectrometria e suas aplicações no mapeamento geológico. *TERRA E DIDÁTICA*, v.10, p.29-51, 2013.

RIDGEWAY, G. Generalized Boosted Regression Models. Package gbm in R. 2017. Disponível em < <https://cran.r-project.org/web/packages/gbm/gbm.pdf>>. Acesso em Ago, 2017.

ROUSE J. W.; HAAS Jr, R. H.; DEERING D. W.; SCHELL J. A., Monitoring the vernal advancement and retrogradation (Green wave effect) of natural vegetation, Texas A&M Univ., College Station. p. 1-69, 1973.

ROVEDA, R.G.A.; COSTA, L.M.; SHAEFER, C.E.G.R.; FERNANDES FILHO, E.J.; GAGGERO, M.R. Análise digital do terreno ferramenta na identificação de pedoformas em microbacia na região de mar de morros (MG). **Revista Brasileira de Ciência do Solo**. v.29,n.1.p.269-276, 2005.

ROVEDA, S.R.M.M.; ROVEDA, J.A.F.; LOUREÇO, R.W. Aplicação da lógica fuzzy para estudo de permeabilidade de solos de região impactada da Baixada santista. **Holos Environment**. v. 11, n.2, p.180-187, 2011.

RYAN, P.J., MCKENZIE, N.J., O'CONNELL, D., LOUGHHEAD, A.N., LEPPERT, P.M., JACQUIER, D., ASHTON, L.. Integrating forest soils information across scales:

spatial prediction of soil properties under Australian forests. *Forest Ecology and Management*. v. 138, p. 139-157. 2000.

SALIBY, E. Descriptive Sampling: An Improvement Over Latin Hypercube Sampling. In: **Winter Simulation Conference**, v. 1, n 1, p. 1-6, 1997.

SANCHEZ, P.A.; AHAMED, S.; CARRÉ, F.; HARTEMINK, A.E.; HEMPEL, J.; HUISING, J.; LAGACHERIE, P.; MCBRATNEY, A.B.; MCKENZIE, N.J.; MENDONCA-SANTOS, M. de L.; MINASNY, B.; MONTANARELLA, L.; OKOTH, P.; PALM, C.A.; SACHS, J.D.; SHEPHERD, K.D.; VÅGEN, T.-G.; VANLAUWE, B.; WALSH, M.G.; WINOWIECKI, L.A.; ZHANG, G.-L. Digital soil map of the world. *Science*, v.325, n 5941, p.680-681, 2009.

SANDRI, S.; CORREA, C. *Lógica Nebulosa*. São José dos Campos: ITA, 1999. Disponível em: <<http://www.ele.ita.br/cnrm/minicursos-5ern/log-neb.pdf>>. Acesso em: 22 jul.2014

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. Viçosa, MG, Sociedade Brasileira de Ciência do Solo/Rio de Janeiro, Embrapa, 2005. 100p.

SANTOS, H.G. Importância e evolução dos levantamentos de solos no Brasil. **Boletim Informativo**, Viçosa, v.32.n.1, p.21-26. 2007.

SANTOS, M.L.M E SANTOS, H.G. Mapeamento digital de classes e atributos de solos: métodos, paradigmas e novas técnicas. **Série Documentos Embrapa Solos**. v.1 ,n 55. 19 p. 2003.

SARMENTO, E. C. **Comparação entre quatro algoritmos de aprendizagem de máquina no mapeamento digital de solos no Vale dos Vinhedos, RS, Brasil**. 109 p. 2010. (Tese de Doutorado). Universidade Federal do Rio Grande do Sul.

SHI, A.X; LONG, R.; DEKETT, R.; PHILIPPE, J. Integrating different types of knowledge for digital soil mapping. **Soil Science Society of America Journal**. v. 73, n 5, p. 1682-1692, 2009.

SILVA, S. H. G. Cambisol (Inceptisol) Solum thickness digital mapping based on expert knowledge with limited data from a watershed in Minas Gerais, Brazil. 41 p. Monografia – Universidade Federal de Lavras, Lavras, 2013.

SILVA, Cristiano Cassiano. **Mapeamento digital de classes de solo: aplicação de metodologia na folha Botucatu (SF-22-Z-B-IV-3) e validação de campo**. 117 p. 2012. (Dissertação de Mestrado). Instituto Agronômico de Campinas.

SILVA, L.M.; GONÇALVES, R.M.; LIRA, M.M.S.; FERREIRA, P.S. Modelagem fuzzy aplicada na detecção da vulnerabilidade à erosão costeira. **Boletim de Ciências Geodésicas**, v. 19, n 4, p.746-764, 2013.

SILVA, S.A e LIMA, J.S. Lógica fuzzy no mapeamento de variáveis indicadoras de fertilidade do solo. **Idesia(Chile)**. v.27, n.3,p.41-46, 2009.

- SILVA,S.A.; LIMA,J.S. de; SOUZA, G.S.; XAVIER, A.C. Distribuição e incerteza da acidez de um Latossolo vermelho-amarelo húmico sob o cultivo do café. **Revista Brasileira de Ciência do Solo**. v. 33; n.1; p.1053-1060, 2009.
- SOUZA, DMG; RITCHEY, KD; LOBATO E; GOEDERT, WJ. Potássio em solo do Cerrado: II. Balanço no solo. *Revista Brasileira de Ciência do Solo*, 3:33-36, 1979.
- STOLLE,L; LINGNAU,C.; ARCE, J.E.; BOGNOLA, I.A. Aplicação de um sistema baseado em conhecimento e lógica fuzzy para o mapeamento da fragilidade ambiental em relação às atividades florestais. *Anais do XIV Simpósio Brasileiro de Sensoriamento Remoto*.p. 3103-3110. 25 a 30 de Abr, 2009.
- SUAREZ,A.F. E CANDEIAS, A.L.B. Avaliação da acurácia da classificação de dados de sensoriamento remoto para o município de Maragogipe. **IV Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação**. Recife/PE.p 1-10. 2012.
- TAN, P.-N., STEINBACH, M., and KUMAR, V. Introdução ao data mining: mineração de dados. Rio de Janeiro: Ciência Moderna, 2009.
- TANSCHKEIT, R. Sistemas fuzzy. PUC-Rio, Rio de Janeiro, 2014. Disponível em <<http://www2.ica.ele.puc-rio.br/Downloads/41/LN-Sistemas%20Fuzzy.pdf>> Acesso em 21 jul, 2014.
- TEN CATEN, A., DALMOLIN, R.S.D, PEDRON, F.A., SANTOS, M.de L.M. Regressões logísticas múltiplas: fatores que influenciam sua aplicação na predição de classes de solos.. *R. Bras. Ci. Solo*, v. 35, p. 53-62, 2011.
- WERE, K., BUI, D.T., DICK, O.B., SINGH, B.R. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. *Ecological Indicators*. v.52, p.394–403, 2015.
- WIESMEIER, M.; BARTHOLD, F., BLANK, B; KÖGEL-KNABNER,I . Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. *Plant Soil*. v. 340, p.7–24, 2011.
- ZADEH,L.A. Fuzzy sets. **Information and control**. n. 8, p.338-353. 1965.
- ZHANG, T. "Adaptive Forward-Backward Greedy Algorithm for Learning Sparse Representations". Rutgers Technical Report. Departamento de Estatística, Rutgers University, New Jersey, EUA. 2008.
- ZHANG, T. Greedy variable selection. Package R Foba. 2015. Disponível em <<https://cran.r-project.org/web/packages/foba/foba.pdf>>. Acesso em Ago, 2017.
- ZHIU, A.; FENG, Q.; MOORE, A. BURT, J.E. Prediction of soil properties using fuzzy membership values. **Geoderma**. v. 158, n 3, p.199-206, 2010.
- ZHU, A.X.; YANG, L.; LI,B.; QIN, C. ENGLISH, E.; BURT, J.; ZHOU, C. Purposive Sampling for digital soil mapping for areas with limited data. In: HARTEMINK, A.E.; McBRATNEY,A.; MENDONÇA-SANTOS, M.L. (Ed.). **Digital Soil Mapping With limited Data**. Springer, p. 233-245, 2008.

ZHU, Q.; LIN, H. S. Comparing ordinary kriging and regression kriging for soil properties in contrasting landscapes. **Pedosphere**, v. 20, n. 5, p. 594-606, 2010.

ZHU, A.X.; BAND, L.; VERTESSY, R.; DUTTON, B. Derivation of soil properties using a soil land inference model (SoLIM). **Soil Science Society of American Journal**. v. 61, n 2, p. 523-533, 1997.

ZIMMER, J.; ANZANELLO, M.J. Um novo método para seleção de variáveis preditivas com base em índices de importância. *Production*. v. 24, n.1, p. 84-93, 2014.

CAPÍTULO IV
PREDIÇÃO DE ATRIBUTOS DO SOLO UTILIZANDO
DADOS ESPECTRAIS

RESUMO

A quantificação dos atributos do solo é essencial para maior entendimento dos ambientes e melhor manejo do solo. Porém, as tradicionais análises de laboratório que permitem a obtenção dessas informações são, por vezes, onerosas e impactantes para o meio ambiente. Dessa forma, é necessário o desenvolvimento de novas técnicas de análise, como a espectroscopia, que apresenta-se cada dia mais como uma forma viável para estimar uma ampla gama de propriedades do solo. Este trabalho objetivou prever os teores de carbono orgânico (CO), argila e P extraível, a partir dos espectros MIR, Vis-NIR e Vis-NIR-MIR, utilizando diferentes métodos de pré-processamento (*Continuum removal-CR*; *Absorbance-ABS*; *Savitzky Golay Derivative-SGD*; *Satandard Normal Variate-SNV* e *Multiplicative Scatter Correction-MS-C*) combinados a cinco modelos de predição (*Partial Least Squares Regression-PLSR*; *Artificial Neural Network-ANN*; *Random Forest-RF*; *Guassian Process Regression-GPR* e *Support Vector Machine-SVM*). O trabalho foi realizado em Iconha/ES, na bacia do Ribeirão Inhaúma. Foram utilizadas 184 amostras, coletadas em 92 pontos em duas profundidades (0-10 e 10-30 cm). As análises físicas, químicas e espectrais foram feitas segundo método utilizado no laboratório de rotina de solos. Das amostras, foram selecionadas aleatoriamente 70 % para treinamento e 30% para validação (conjunto externo de amostras) dos modelos, utilizou-se a validação cruzada. Ao se combinar os métodos de pré-processamento e os algoritmos de regressão multivariada, obteve-se 30 testes. Para análise de desempenho foi calculado o coeficiente de determinação R^2 ; raiz quadrada do erro médio (RMSE), e os índices de padronização do erro de predição, RPD (*Ratio of performance/prediction to deviation*) e RPIQ (*Ratio of performance/prediction to interquatile*). A interpretação dos valores de RPD quanto à qualidade/confiabilidade da predição, foi feita segundo os critérios propostos por Chang e Laird (2002) e Chang et al. (2001). Para argila e CO o melhor valor de R^2 foi encontrado no espectro MIR sendo igual a 0,69 e 0,65, respectivamente e para P de 0,57 em Vis-NIR. Este estudo encontrou modelos satisfatórios (RPIQ e/ou RPD superior a 1,4) e excelentes (RPIQ superior a 2,0) capazes de prever argila, CO e P. Os pré-processamentos MSC, CR e SNV foram os mais eficientes para predição de argila, CO e P, respectivamente, enquanto os métodos PLSR (CO e P) e SVM (argila) se destacaram com as melhores predições, sendo por isso, recomendados para modelagem desses atributos na área estudada. Os modelos encontrados podem ser testados para discriminar solos segundo a um valor de teste crítico para argila, CO e P. Como, por exemplo, separar solos arenosos e argilosos. Acredita-se que a adição de mais amostras no treinamento e validação, pode melhorar os modelos com RPIQ e RPD entre 1,4 e 2,0 tornando-os modelos acurados, com excelente predição. Palavras-chave: análise espectral, Reflectância, quimiometria.

ABSTRACT

The quantification of soil attributes is essential for a better understanding of the environments and better soil management. However, the traditional laboratory analyzes that allow obtaining this information are sometimes costly and impactful to the

environment. Thus, the development of new analytical techniques, such as spectroscopy, is increasingly needed as a viable way of estimating a wide range of soil properties. This work aimed to predict the organic carbon (CO), clay and extractable P contents from the MIR, Vis-NIR and Vis-NIR-MIR spectra using different methods of pre-processing (Continuum removal-CR; Absorbance-ABS; Savitzky Golay Derivative-SGD (GMS) and the Multiplicative Scatter Correction (MSC), combined with five predictive models (Partial Least Squares Regression-PLSR, Artificial Neural Network-ANN, Random Forest-RF, Gaussian Process Regression-GPR and Support Vector Machine-SVM). The work was carried out in Iconha / ES, in the Ribeirão Inhaúma basin. We used 184 samples, collected in 92 points at two depths (0-10 and 10-30 cm). The physical, chemical and spectral analyzes were performed according to the method used in the routine soil laboratory. Of the samples, 70% were randomly selected for training and 30% for validation (external set of samples) of the models, cross-validation was used. By combining the pre-processing methods and the multivariate regression algorithms, we obtained 30 tests. For the performance analysis, the coefficient of determination R^2 was calculated; (RMSE), and prediction error standardization indexes, RPD (Ratio of performance / prediction to deviation) and RPIQ (Ratio of performance / prediction to interquatile). The interpretation of the RPD values regarding the quality / reliability of the prediction was made according to the criteria proposed by Chang and Laird (2002) and Chang et al. (2001). For clay and CO the best value of R^2 was found in the MIR spectrum being equal to 0.69 and 0.65, respectively, and for P of 0.57 in Vis-NIR. This study found satisfactory models (RPIQ and / or RPD superior to 1,4) and excellent (RPIQ superior to 2,0) able to predict clay, CO and P. The MSC, CR and SNV preprocesses were the most efficient for prediction of clay, CO and P, respectively, while the PLSR (CO and P) and SVM (clay) methods stood out with the best predictions and are therefore recommended for modeling these attributes in the studied area. The models found can be tested to discriminate soils according to a critical test value for clay, CO and P. For example, separate sandy and clayey soils. It is believed that the addition of more samples in the training and validation can improve the models with RPIQ and RPD between 1.4 and 2.0 making them accurate models with excellent prediction.

Keywords: Spectral analysis, Reflectance, chemometrics.

1 INTRODUÇÃO

A Pedometria aborda os problemas da ciência do solo utilizando técnicas modernas de análise de dados para quantificar de maneira mais rápida as propriedades do solo ao longo do espaço e do tempo, como por exemplo, a calibração e predição de propriedades do solo a partir de espectros de reflectância infravermelha difusa (MINASNY E McBRATNEY, 2008).

Nanni e Demattê (2006) apontam que a análise de laboratório, usada para quantificação de atributos do solo, apresentam custo elevado (BASHAGALUKE et al., 2015) e possibilidade de geração de impactos ambientais pelo uso de reagentes químicos na análise, além de dispendir muito tempo entre a amostragem e a aquisição dos resultados. Segundo Reeves (2009), um outro desafio enfrentado pelos cientistas do solo,

está no fato das informações existentes da pesquisa de solo geralmente não atenderem aos requisitos de dados para modelagem ou previsão.

Por isso, torna-se necessário o uso de novas técnicas de análise do solo. Nos últimos anos, tem sido dada atenção ao potencial de uso da espectroscopia de reflectância, tanto nos espectros do infravermelho próximo (NIR) e quanto no infravermelho médio (MIR), pois estas são técnicas não destrutivas, rápidas e convenientes para se quantificar os atributos do solo (BASHAGALUKE et al., 2015). Mohamed et al. (2017) aponta que NIR é uma estratégia sistemática não destrutiva para caracterizar e identificar as propriedades do solo, enquanto Abdi et al., (2016) corrobora que a espectroscopia visível e de reflectância do infravermelho próximo (Vis-NIR) é uma técnica rápida, barata e precisa para analisar uma grande variedade de materiais orgânicos e é cada vez mais utilizada na ciência do solo (NOURI et al., 2017).

Adeline et al. (2017) apontam que a espectroscopia Vis-NIR de laboratório é uma boa alternativa para análises físicas e químicas caras do solo para estimar uma ampla gama de propriedades do solo. Segundo Viscarra Rossel et al. (2006) as técnicas espectroscópicas de infravermelho são altamente sensíveis às fases orgânica e inorgânica do solo, tornando seu uso nas ciências agrícola e ambiental particularmente relevante.

Dessa forma, o uso da espectroscopia de reflectância para determinação de atributos químicos e físicos do solo vem sendo cada vez mais difundido (ADELINE et al., 2017), este vem como uma ferramenta acessória a determinação via métodos convencionais (VENDRAME et al., 2012), diminuindo assim variáveis como tempo de análise (SOUSA JUNIOR et al. 2011) e minimizando possíveis impactos na natureza.

A análise quantitativa baseada nos espectros NIR ou MIR requer o desenvolvimento de calibrações que relacionam a informação espectral com as concentrações de analito conhecidas (REEVES, 2009). Devido às diferenças nas origens do solo e nos procedimentos de medição laboratoriais, as funções de calibração de uma região não funcionam bem em outra. A calibração local é preferida para evitar o viés de medição entre laboratórios, e as funções calibradas são mais específicas para os tipos de solo em uma área (MINASNY et al., 2009).

Uma maneira de lidar com dados com um grande número de covariáveis, como os espectros infravermelhos, é a redução de dados (MINASNY E McBRATNEY, 2008), já que a análise espectral quantitativa do solo usando espectroscopia de reflectância visível e infravermelha requer técnicas estatísticas sofisticadas para discernir a resposta dos atributos do solo a partir de características espectrais (VISCARRA ROSSEL et al., 2006).

As estratégias de pré-processamento são incorporadas como um primeiro passo antes da calibração de predição, para melhorar a extração de informações úteis tanto de efeitos aditivos como multiplicativos sobrepostos nos espectros de reflexão (PENG et al., 2014). Diferentes métodos matemáticos são utilizados para vincular os espectros de laboratório Vis-NIR do solo (variáveis preditoras) às propriedades do solo (variáveis de resposta) (NOURI et al., 2017). Dentre essas inclui-se *support vector machine* (Stevens et al., 2010) e regressão parcial de mínimos quadrados (PLSR) (McCARTY et al., 2002).

Segundo Nouri et al (2017) o método de regressão parcial de mínimos quadrados (PLSR) é provavelmente a técnica estatística multivariada mais utilizada para a calibração espectral e a predição das propriedades do solo. É uma técnica que faz a redução dos preditores para um conjunto com menor número de componentes que não sejam correlacionados entre si, e posteriormente realiza a regressão de mínimos quadrados. Os modelos usados mais comuns são regressão linear múltipla (SMLR), regressão parcial de mínimos quadrados (PLSR), splines de regressão adaptativa multivariada (MARS), regressão de componentes principais (PCR) e redes neurais artificiais (ANN) (MOHAMED et al, 2017).

Random Forest é um método de classificação e regressão, desenvolvido por Breiman (2001), nessa técnica são formadas árvores de decisão utilizando um conjunto aleatório inicial, ou seja, cada árvore é gerada a partir dos valores de um vetor aleatório. É bastante robusto e insensível a ruídos.

As redes neurais (ANN) normalmente são organizadas em camadas, essas são constituídas por uma série de "nós" interligados que contêm uma "função de ativação". Os padrões são apresentados à rede através da "camada de entrada", que se comunica com uma ou mais "camadas ocultas" onde o processamento real é feito através de um sistema de "conexões" ponderadas; A maioria das ANNs contém alguma forma de "regra de aprendizagem" que modifica os pesos das conexões de acordo com os padrões de entrada (MOHAMED et al, 2017).

O *Support Vector Machine* (SVM) é uma técnica computacional baseada no reconhecimento de padrão, assim esse determina limites de decisão onde ocorre separação ótima entre as classes com minimização dos erros (NASCIMENTO et al., 2009). Já *Guassian Process Regression* (GPR) é equivalente, em geoestatística, à interpolação por krigagem amplamente conhecida e utilizada na pesquisa pedométrica. No entanto ao invés de usar coordenadas geográficas como dados de entrada, utiliza dados espectrais (RAMIREZ-LOPEZ et al., 2013).

Assim este trabalho objetivou prever os teores de carbono orgânico (CO), argila e P extraível, a partir dos espectros MIR, Vis-NIR e Vis-NIR-MIR, utilizando diferentes métodos de pré-processamento (*Continuum removal-CR*; *Absorbance-ABS*; *Savitzky Golay Derivative-SGD*; *Satandard Normal Variate-SNV* e *Multiplicative Scatter Correction-MS-C*) combinados a cinco modelos de predição (*Partial Least Squares Regression-PLSR*; *Artificial Neural Network-ANN*; *Random Forest-RF*; *Guassian Process Regression-GPR* e *Support Vector Machine-SVM*).

2 REVISÃO BIBLIOGRÁFICA

2.1 Comportamento Espectral Do Solo

O comportamento espectral dos solos, também chamado de assinatura espectral, varia de acordo com a capacidade de troca catiônica-CTC, composição mineralógica, umidade, granulometria, além do teor de matéria orgânica. De acordo com Moraes (2002), esse comportamento é determinado pelas combinações e arranjos dos materiais que constituem os solos.

Há um aumento linear e positivo quanto a absorção de radiação eletromagnética e a capacidade de troca catiônica, ou seja, a medida que há incremento na CTC há aumento na absorção da radiação, o mesmo ocorre com o aumento do teor de matéria orgânica no solo. Quanto a granulometria, o aumento da concentração de minerais de granulação muito fina, como os félsicos, incrementa a reflectância, diminuindo conseqüentemente as bandas de absorção, sendo que o contrário ocorre quando há aumento na concentração de minerais ricos em ferro e magnésio, os chamados máficos. Em relação a umidade, solo úmidos refletem menos, quando comparados a solos secos (FIGUEIREDO, 2005).

Conforme Dalmolin et al. (2005), os componentes que mais afetam o comportamento espectral dos solos são a matéria orgânica e os óxidos de ferro, e, segundo Demattê et al. (2004), através da avaliação das curvas espectrais do solo é possível

realizar a quantificação de determinados atributos do solo como a areia, argila, silte, e até mesmo teores de nutrientes.

A Figura 37, mostra o comportamento espectral para um solo siltoso e arenoso, ambos secos, cujas reflectâncias foram obtidas in situ (SEBEM, 2017b). Enquanto a Figura 38, retrata duas curvas de espectrais, dos horizontes A e B de dois Latossolos, sendo um Latossolo Vermelho distroférico típico e um Latossolo Amarelo ácrico típico.

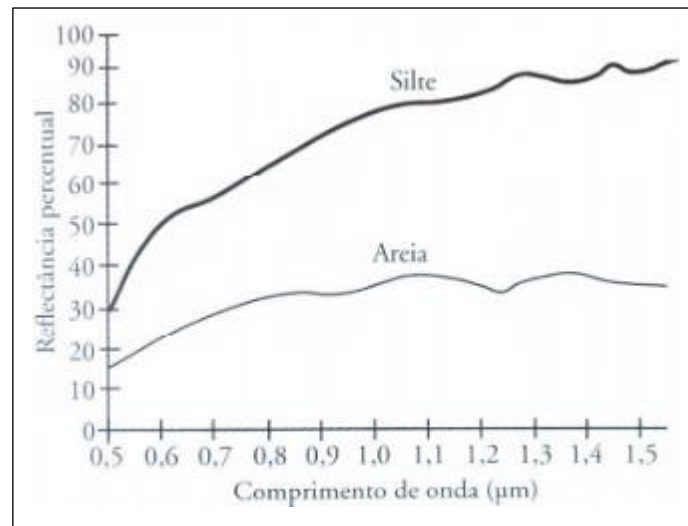


Figura 37- Comportamento espectral para um solo siltoso e arenoso, ambos secos.

Fonte: Sebem, 2017

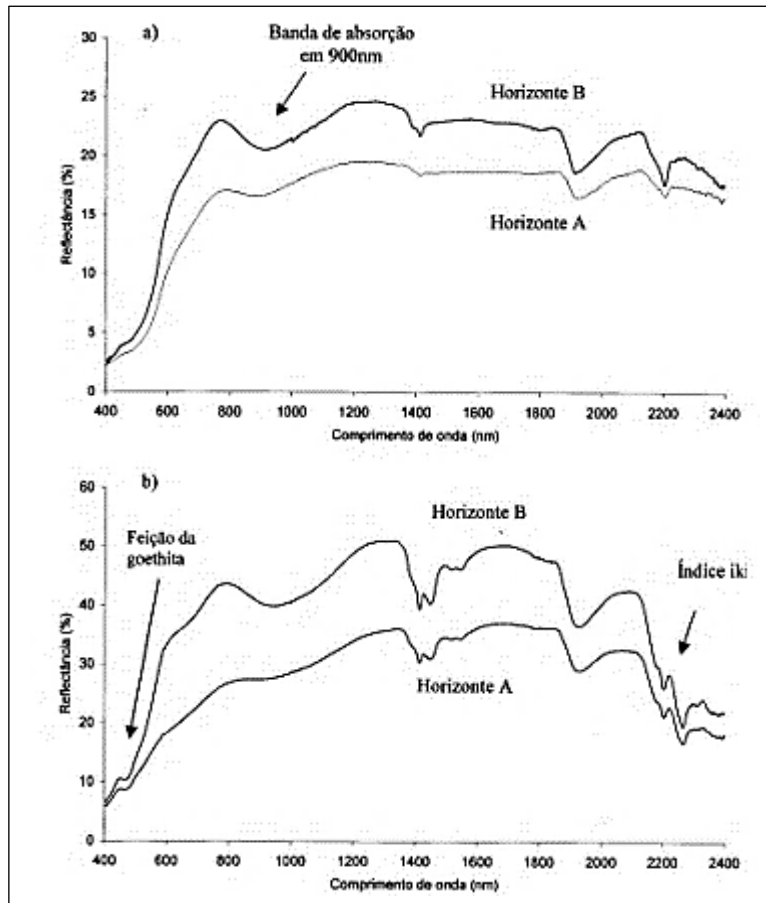


Figura 38-Curvas espectrais dos horizontes A e B de um Latossolo Vermelho distroférico típico (a) e de um Latossolo Amarelo ácrico típico (b).

Fonte:(Dalmolin, 2002)

2.2 Espectroscopia De Reflectância

As curvas espectrais têm uma ligação direta com a estrutura do átomo de um elemento químico, por isso diferentes elementos químicos geram curvas espectrais distintas, já que cada átomo tem capacidade de irradiar em determinado comprimento de onda (UNICAMP, 2017). Dessa forma a espectrometria, tem como base definir qual é uma determinada substância a partir do seu espectro.

A Figura 39, mostra curvas características da gibbsita, um óxido de alumínio presente em grande parte dos solos brasileiros.

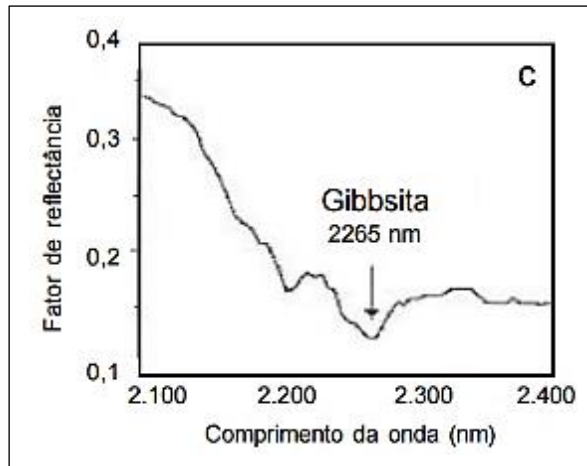


Figura 39- Curva espectral da gibbsita.

Fonte: Madeira Neto (1996)

De acordo com Stoner e Baumgardner (1979), a energia refletida por determinado solo é dada em função de três fatores, sendo eles a energia eletromagnética que incide pelo sol ou uma lâmpada; a quantidade de energia absorvida e a quantidade de energia transmitida. Dalmolin et al. (2005), citam que a relação entre a radiação incidente e a energia absorvida e transmitida, pode ser dada pela equação:

$$R\lambda = I\lambda - (A\lambda + T\lambda) \quad (1)$$

Onde :

R é a reflectância em um determinado comprimento de onda (λ),

I é a energia incidente,

A é a energia absorvida e

T é a energia transmitida.

Os registros espectrais podem ser obtidos em nível terrestre, em campo ou laboratórios, em nível suborbital pelo uso de aeronaves e orbital utilizando-se satélites. A espectrorradiometria laboratorial é especialmente importante para se compreender e definir muitos dos conceitos utilizados na prática do sensoriamento remoto, já que permite o conhecimento de como cada componente se comporta (ALBA et al., 2006).

A Figura 40, mostra um exemplo de espectrorradiômetro utilizado em laboratório, modelo LI-COR 1800.

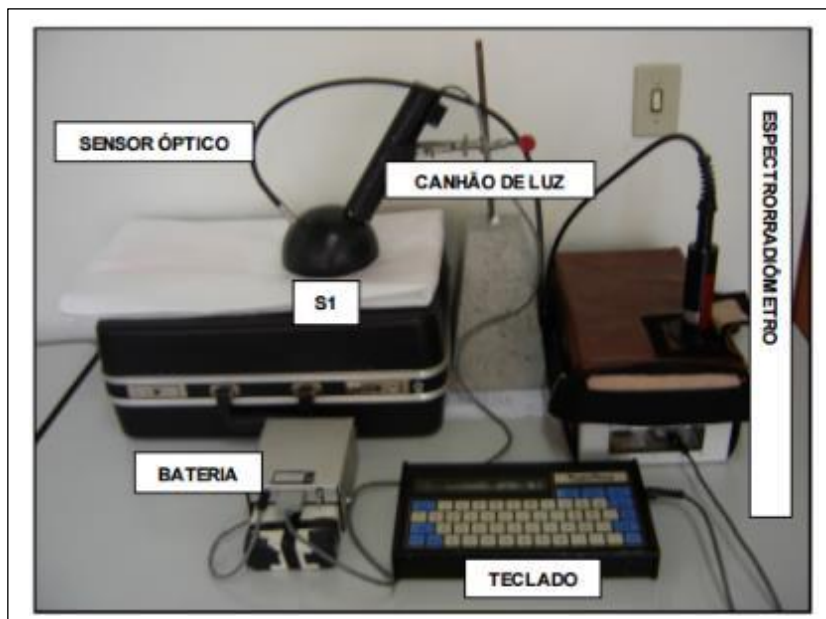


Figura 40- Espectrorradiômetro LI-COR 1800.

Fonte: Alba et al. (2006)

Existem várias formas nas quais a espectroscopia de reflectância pode ser aplicada dentro das ciências agrárias, com especial foco na modelagem digital de atributos e classes de solos.

No Brasil existe a Biblioteca espectral de solos, desenvolvido pela Escola Superior de Agricultura Luiz de Queiroz-ESALQ/USP, a qual já conta com dados de sete estados, além de 500 amostras cedidas por pesquisadores de outras instituições, o que totaliza amostras de 12 estados no total.

2.3 Infravermelho Próximo e Médio (NIR E MIR)

Historicamente, a compreensão do solo e avaliação de sua qualidade foi adquirida através da análise laboratorial física e química. Existe uma tendência global para o desenvolvimento de metodologias mais eficientes em termos de tempo e custo para a análise do solo, uma vez que há uma grande demanda por quantidades maiores de boa qualidade, dados de baixo custo do solo para serem usados em monitoramento ambiental, modelagem e agricultura de precisão (VISCARRA ROSSEL et al, 2006).

Um dos desafios dos pesquisadores atualmente é relacionar a informação do solo existente para fornecer uma técnica rápida e barata para medir com precisão as propriedades funcionais do solo para modelagem e previsão (McBRATNEY et al, 2006). Assim, a espectroscopia de reflectância do infravermelho próximo (NIR) se apresenta como uma técnica rápida, barata, precisa e ambientalmente correta que pode prever a concentração de atributos do solo a partir de um modelo empírico calibrado usando espectros de solo (MALLEY et al ., 2004).

O uso mais moderno de espectrômetros de laboratório para determinar a matéria orgânica do solo (Krishnan et al., 1980), começou nos anos 80 e ampliou-se consideravelmente na década de 90 (REEVES, 2010).

O MIR é capaz de prever propriedades fundamentais do solo relacionadas à matriz do solo ou constituintes sólidos e às concentrações da solução do solo que estão em equilíbrio com as fases sólidas. Os constituintes orgânicos e minerais básicos do solo são bem preditos: capacidade de troca catiônica, teor de argila, C orgânico e total e N. Total. As propriedades químicas relacionadas aos componentes minerais e orgânicos também podem ser previstas, incluindo: K total, P total, Fe elemental, Si e Al (MINASNY et al., 2009).

Ambos NIR e MIR fornecem grandes conjuntos de variáveis preditoras que normalmente são fortemente colineares e ruidosas, características que podem afetar o sucesso da calibração. Para compensar parcialmente esses efeitos, a redução da dimensionalidade é uma abordagem padrão (VOHLAND et al., 2014).

O princípio fundamental do NIR baseia-se nas diferenças nas características moleculares, onde as assinaturas espectrais de diferentes materiais são categorizadas com base em seus espectros de reflectância e absorvência (MOHAMED et al, 2017). A mudança nas assinaturas é referida à extensão e flexão vibratória de átomos que organizam moléculas e cristais (MOHAMED et al, 2017). A maioria dos componentes do solo geralmente são observados nas vibrações da região do infravermelho médio (2.500-25.000 nm), com armadilhas e combinações encontradas na região do infravermelho próximo (700-2.500 nm) (SHEPHERD E WALSH, 2002).

A espectroscopia de infravermelhos baseia-se no princípio das transições nos estados vibratórios e rotacionais de uma molécula e detecta a absorvência por ligações orgânicas e componentes minerais (MINASNY et al., 2009). Assim, a previsão quantitativa das propriedades do solo por espectroscopia infravermelha baseia-se em relações empíricas entre os espectros e as propriedades do solo caracterizadas por laboratório (MINASNY et al., 2009).

As assinaturas espectrais de materiais são definidas por sua reflectância, ou absorvência, em função do comprimento de onda. Em condições controladas, as assinaturas são devidas a transições eletrônicas de átomos e estiramento e flexão vibratória de grupos estruturais de átomos que formam moléculas e cristais. As vibrações fundamentais da maioria dos materiais do solo podem ser encontradas na região do infravermelho médio e infravermelho próximo (BROWN et al., 2006)

Os espectros Vis-NIR resultam da radiação absorvida por várias ligações químicas (por exemplo, C-H, N-H, O-H, C-O, C-N e N-O) encontrado em constituintes orgânicos do solo, bem como Ca-O, Fe-O e Al-O em minerais (SORIANO-DISLA et al., 2014 ; NOCITA et al ., 2015). As características de absorção na faixa visível são dominadas pelas excitações moleculares eletrônicas, e as da faixa NIR contêm uma combinação de vibrações moleculares de alta qualidade (ADELINE et al., 2017).

A espectroscopia de reflectância do infravermelho próximo é um procedimento explicativo rápido e não prejudicial que inclui estimativa de reflectância difusa na região do infravermelho próximo (1000-2500 nm). A espectroscopia de NIR com modelagem de árvore de decisão pode de forma justa e precisa, com incerteza pequena a moderada, prever carbono orgânico do solo (VISCARRA-ROSSEL E HICKS, 2015 e HU et al., 2015).

A aplicação da espectroscopia de infravermelho foi sugerido para ser útil até mesmo para análises biológicas do solo (KUANG et al., 2012 e SORIANO-DISLA et al., 2014).

Assim, diversas propriedades do solo, como a matéria orgânica do solo (MOS), carbono (C), textura, umidade, cálcio (Ca), magnésio (Mg), oxi-hidróxidos de alumínio (Al) / ferro (Fe) e óxidos de ferro, podem ser estimadas usando Vis-NIR (CHANG et al., 2001 ; COZZOLINO e MORÓN, 2006 ; BRUNET et al ., 2007).

Além de medir os constituintes minerais e orgânicos do solo, o Vis-NIR tem sido usado para estimar medidas de fertilidade como saturação de base, pH, bases trocáveis e P extraível (Ben-Dor e Banin, 1995; Dunn et al., 2002; Shepherd e Walsh, 2002; Islam et al., 2003), textura (SHEPHERD e WALSH, 2002; COZZOLINO e MORON, 2003, ISLAM et al., 2003, MORON e COZZOLINO, 2003) e elementos totais Ca, Mg, Fe, Mn, K e Cu (COZZOLINO e MORON, 2003; UDELHOVEN et al., 2003).

Estudos observaram que o aumento do carbono orgânico do solo diminuiu o albedo em todo o espectro de reflectância visível, de ondas curtas e infravermelho próximo (Vis-NIR-SWIR) (STONER e BAUMGARDNER, 1981; HENDERSON et al., 1992). Carbono orgânico foi previsto a partir de várias partes do Vis-NIR-SWIR, isso em grande parte porque ele contém tantos componentes (SUMMERS et al., 2011). Estes componentes incluem compostos tais como a lignina (por exemplo, 2050 e 2351 nm), celulose (1370, 1725 e 2347 nm), pectina (por exemplo, 1320, 1582, 1761 e 2111 nm) e húmus (1929 e 1932 nm) que são opticamente ativos em toda essa região espectral (Summers et al., 2011) e pensa-se que se sobrepõem em lugares (BEN-DOR et al., 1997).

A espectroscopia NIR apresenta uma chance única de analisar os solos como um todo, pois seu padrão espectral contém informações da maioria dos componentes orgânicos da amostra (ROBERTS e COZZOLINO, 2016). Examinando as várias técnicas de coleta de dados da espectroscopia e as análises computacionais disponíveis, destaca que uma combinação adequada de ambas as abordagens seria a maneira mais eficaz de interpretar os espectros, sendo desconsiderada, confiável, rápida e robusta (ROBERTS e COZZOLINO, 2016).

Para análise dos dados provenientes da espectroscopia muitas vezes são necessários alguns pré-processamentos, podendo ser utilizadas técnicas como *Continuum removal-CR*; *Absorbance-ABS*; *Savitzky Golay Derivative-SGD*; *Satandard Normal Variate-SNV* e *Multiplicative Scatter Correction-MS*.

SNV é variação normal padrão que concentra cada espectro pela média e, em seguida, escala-o pelo seu desvio padrão. Esta abordagem foi considerada eficaz na remoção das interferências multiplicativas de artefatos de dispersão nos espectros NIR (BARNES et al., 1989). O MSC é motivado pelas observações empíricas de que os valores espectrais de uma amostra são aproximadamente lineares quando plotados em relação aos valores espectrais médios (HELLAND et al., 1995; GELADI et al., 1985).

Após os pré-processamentos são utilizados diferentes modelos para predição das diferentes propriedades do solo. Alguns exemplos são: *Partial Least Squares Regression-PLSR*; *Artificial Neural Network-ANN*; *Random Forest-RF*; *Guassian Process Regression-GPR* e *Support Vector Machine-SVM*. Os modelos usados mais comuns são regressão linear múltipla (SMLR), regressão parcial de mínimos quadrados (PLSR), splines de regressão adaptativa multivariada (MARS), regressão de componentes principais (PCR) e redes neurais artificiais (ANN) (MOHAMED et al, 2017).

PLSR é uma técnica de modelagem popular usada em quimiometria e é comumente usada para análises espectrais quantitativas (VISCARRA ROSSEL et al., 2006). Ele é usado para construir modelos preditivos quando existem muitas variáveis preditoras altamente colineais. A técnica está intimamente relacionada à regressão dos componentes principais (PCR). No entanto, ao contrário do PCR, o algoritmo selecciona PLSR sucessivas factores ortogonais que maximizam a covariância entre preditor (X espectros) e as variáveis de resposta (y dados de laboratório). Ao montar um modelo PLSR, espera-se encontrar alguns factores de PLSR que explicam a maior parte da variação em preditores e respostas (VISCARRA ROSSEL et al., 2006).

A regressão parcial de mínimos quadrados (PLSR), que projeta os dados espectrais em um espaço de baixa dimensão formado por um conjunto de variáveis

latentes ortogonais, é atualmente o método mais comum para calibrar espectros de propriedades do solo (VISCARRA ROSSEL e BEHRENS, 2010). A regressão PLSR é um método de processamento de dados desenvolvido por Svante Wold e Herman Wold em 1983. É frequentemente usado para entender melhor as relações entre dois conjuntos de dados. O PLSR visa prever um conjunto de dados e do outro conjunto X usando um modelo linear multivariável. Pode ser usado para analisar dados contendo: variáveis numerosas, barulhentas, colineares e até incompletas tanto em X como em Y (WOLD et al., 2001). Ele procura maximizar a covariância entre X e Y através da criação de componentes chamados variáveis latentes relacionados com X, e Y e é usado para expressar a regressão de Y em X (NOURI et al., 2017).

Ou seja, PLSR decompõe as variáveis X e Y e encontra novos componentes, chamados de variáveis latentes, que são combinações lineares ortogonais e ponderadas de variáveis X. Essas novas variáveis X são então usadas para predição de variáveis Y onde X é reflectância do solo e Y é uma propriedade do solo medida (BILGILI et al., 2010). Ao contrário da regressão linear múltipla, ele pode manipular dados com forte colinearidade em variáveis independentes (X) que também podem ser mais numerosos do que observações. Em PLSR, a seleção do número de variáveis latentes é fundamental para prevenir a sobre ou a debilidade dos dados, o que criaria modelos com pouca capacidade de predição (BILGILI et al., 2010).

O ANN é um sistema de computação composto por uma série de elementos de processamento simples e altamente interconectados, que processam a informação por sua resposta de estado dinâmico a entradas externas (CAUDILL, 1987).

3 MATERIAL E MÉTODOS

O trabalho foi realizado na bacia do Ribeirão Inhaúma, localizado em Iconha/ES, sob as coordenadas 21°10'58.82"S e 41°00'08.87"O. Desejou-se quantificar os atributos do solo a partir de dados espectrais medidos nos comprimentos de onda Vis-NIR (Visível e infravermelho próximo), MIR (infravermelho médio) e Vis-NIR-MIR (Visível e infravermelho próximo e infravermelho médio).

Foram utilizadas 184 amostras, coletadas em 92 pontos em duas profundidades (0-10 e 10-30 cm) (Figura 41). Os locais de amostragem foram determinados utilizando o método do hipercubo latino condicionado, sendo considerado as possibilidades de acesso devido às dificuldades de deslocamento na área em função do relevo acentuado. Dessa forma a malha de pontos, ficou sempre próxima às estradas. Em cada um dos pontos, foram registradas as coordenadas utilizando-se receptor GNSS da Leica, modelo GS08Plus, cujos dados foram processados no programa Leca Geoffice 8.0. A base foi ajustada a partir da estação fixa do IBGE existente em Vitória/ES.

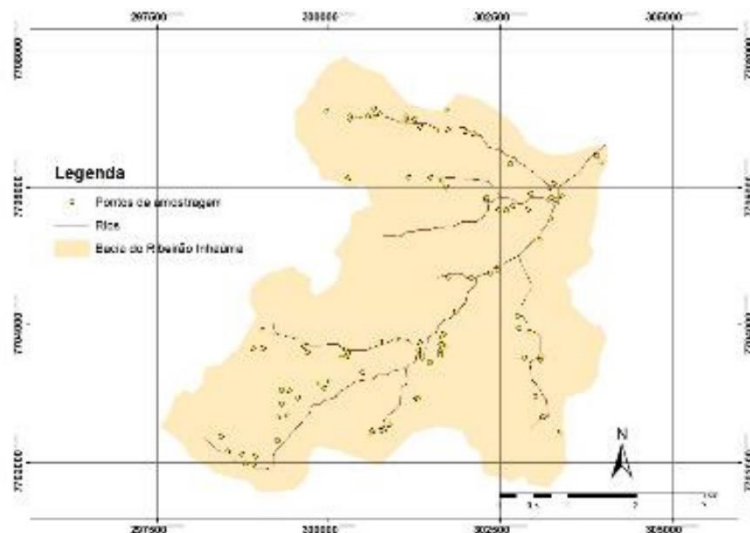


Figura 41- Localização dos pontos de amostragem.

As amostras foram secas ao ar, destorroadas e passadas em peneira com malha de 2 mm, sendo separada entre maior e menor que 2 mm, e quantificada os teores de argila, P extraível e Carbono orgânico. As análises físicas e químicas foram feitas segundo método utilizado no laboratório de rotina de solos, sendo argila pelo método da pipeta, P extraível por Mehlich-1 e carbono orgânico por Walkley & Black.

Na análise espectral, foi utilizado para reflectância bidirecional (350 a 2 500 nm) cerca de 5 g de solo peneirado menor que 2 mm, e para reflectância difusa (infravermelho médio – 2 500 a 25 000 nm) outras 5 g foram processadas, sendo moídas, homogeneizadas e tamisadas a 0,149 mm (100 mesh).

Para obtenção dos dados bidirecionais acondicionou-se as amostras em placas de petri e procedeu-se o nivelamento para diminuir a rugosidade da superfície. Para cada amostra, foram realizadas 300 leituras de modo automático pelo sensor, sendo 100 leituras a cada 90°. O valor final considerado, foi a média das três medidas. A calibração do sensor foi feita utilizando-se placa de Spectralon (sulfato de bário), com reflectância igual a 100%. A calibração foi repetida a cada 20 minutos.

Foi utilizado o sensor *FieldSpec Pro* (*Analytical Spectral Devices, Boulder, Colo*) (HATCHELL, 1999), cuja resolução é de 2 nm para as faixas de 1 100 a 2500 nm, e de 1 nm para os demais comprimentos de onda. O mesmo foi posicionado verticalmente, a 8 cm da amostra, com campo de visada igual a 18°. Duas lâmpadas alógenas de 50 W, foram usadas como fonte de iluminação, sendo essas posicionadas a 35 cm da plataforma, com ângulo zenital igual a 30°.

Ao final foi calculado o fator de reflectância bidirecional, dado pela razão entre a radiância espectral refletida pela amostra de solos e a radiância refletida pela placa de Spelactron.

Já para obtenção dos dados na faixa do infravermelho médio foi usado o *sensor Alpha Sample Compartment RT-DLaTGS ZnSe* (*Bruker Optik GmbH*) equipado com um acessório para aquisição de reflectância difusa (Drift). O equipamento utiliza laser de *HeNe* como fonte de luz posicionada internamente (fonte de iluminação) e com padrão de calibração para cada comprimento de onda. O sensor tem divisor de feixes KBr que permite grande amplitude de radiação incidente sobre a amostra (de 2500 a 25.000 nm).

Cerca de 1 cm³ de amostra de solo, foi acondicionada no recipiente do equipamento afim de proceder-se a leitura. Foram feitas 64 leituras a cada segundo por

espectro e estes foram adquiridos com resolução de 2 cm⁻¹. Antes de cada leitura foi realizada a calibração do sensor, com placa difusa de ouro, que atua na remoção da interferência (*background*) da radiação do espectro da amostra.

Das amostras, foram selecionadas aleatoriamente 70 % para treinamento e 30% para validação (conjunto externo de amostras) dos modelos, utilizou-se também a validação cruzada. Foram testados no pré-processamento os métodos: *Continuum removal* (CR), *Absorbance* (ABS), *Savitzky Golay Derivative* (SGD), *Standard Normal Variate* (SNV) e *Multiplicative Scatter Correction* (MSC). Estes foram combinados com os algoritmos de regressão multivariada: *Partial Least Squares Regression* (PLSR), *Artificial Neural Network* (ANN), *Random Forest* (RF), *Gaussian Process Regression* (GPR linear e radial) e *Support Vector Machine* (SVM). Totalizando, dessa forma, 30 testes.

Para análise de desempenho foi calculado o coeficiente de determinação R², de acordo com a equação:

$$R^2 = \frac{\sum_{i=1}^n (P_i - \bar{o})^2}{\sum_{i=1}^n (O_i - \bar{o})^2} \quad (2)$$

Onde P_i e O_i, são os valores preditos e observados no local i respectivamente; n é o número de amostras.

Foram calculados também a raiz quadrada do erro médio (RMSE), e os índices de padronização do erro de predição, RPD (*Ratio of performance/prediction to deviation*) e RPIQ (*Ratio of performance/prediction to interquatile*). Esses são obtidos pelas equações:

$$RMSE = \sqrt{\frac{1}{l} \sum_{j=1}^l [\hat{z}(S_j) - z^*(S_j)]^2} \quad (3)$$

em que, RMSE é a raiz quadrada do erro médio, e l é o número de pontos destinados a validação.

O erro médio é dado por:

$$MPE = \frac{1}{l} \sum_{j=1}^l [\hat{z}(S_j) - z^*(S_j)]^2 \quad (4)$$

em que, MPE é o erro médio de predição, ($\hat{z}(S_j)$) são os valores estimados e z*(S_j) os dados de validação.

O RPD (Williams, 1987) e o RPIQ (Bellon-Maurel et. al., 2010) são calculados, respectivamente por:

$$RPD = DP/RMSE$$

$$RPIQ = IQ / RMSE$$

Onde DP é o desvio padrão, IQ é a diferença obtida entre o valor referente ao 3º e 1º quartil da distribuição dos dados e RMSE a raiz quadrada do erro médio.

A interpretação dos valores de RPD quanto à qualidade/confiabilidade da predição, foi feita segundo os critérios propostos por Chang e Laird (2002) e Chang et al. (2001), onde valores maiores que 2,00 indicam modelos excelentes para predição de maneira acurada dos atributos; valores entre 1,40 e 2,00 indicam modelos razoáveis e valores abaixo de 1,4 modelos não confiáveis. A mesma analogia foi aplicada para análise de RPIQ (TERRA, 2011).

4 RESULTADOS E DISCUSSÃO

Os espectros sem pré-processamento são mostrados nas Figura 42,43 e 44, enquanto as Figuras 45, 46 e 47 retratam os espectros MIR, Vis-NIR e Vis-NIR-MIR já processados. É possível observar que os pré-processamentos demonstraram-se efetivos, sendo estes mais uniformes em Vis-NIR e Vis-NIR-MIR, embora Kooistra et al. (2001) relataram melhores previsões para argila e matéria orgânica sem pré-processamento de espectros.

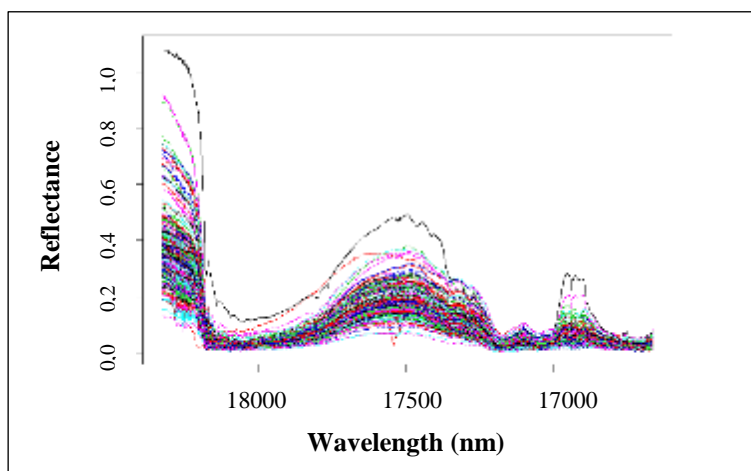


Figura 42- Espectro MIR sem pré-processamento

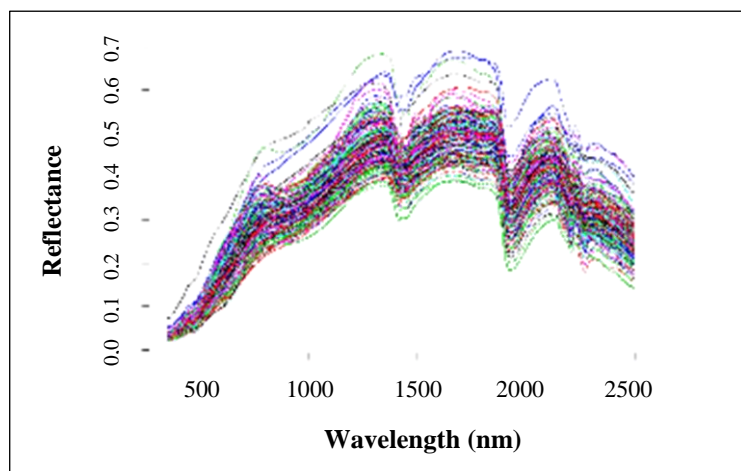


Figura 43- Espectro VIS-NIR sem pré-processamento

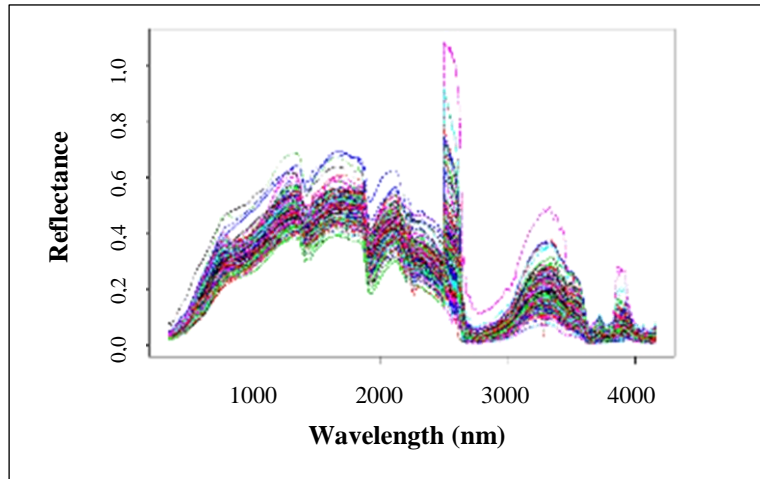


Figura 44- Espectro VIS-NIR-MIR sem pré-processamento

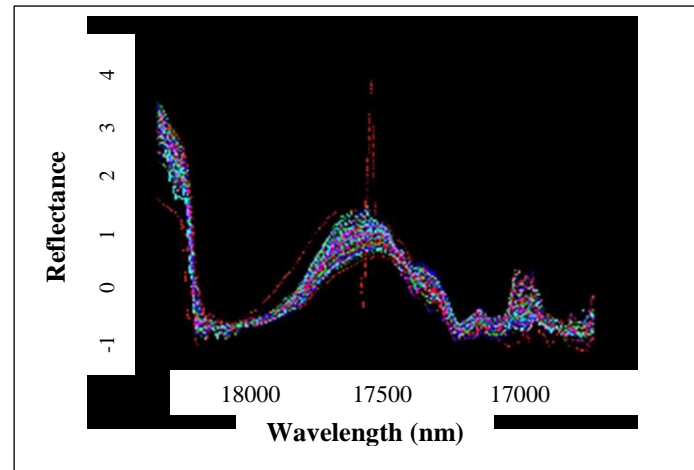
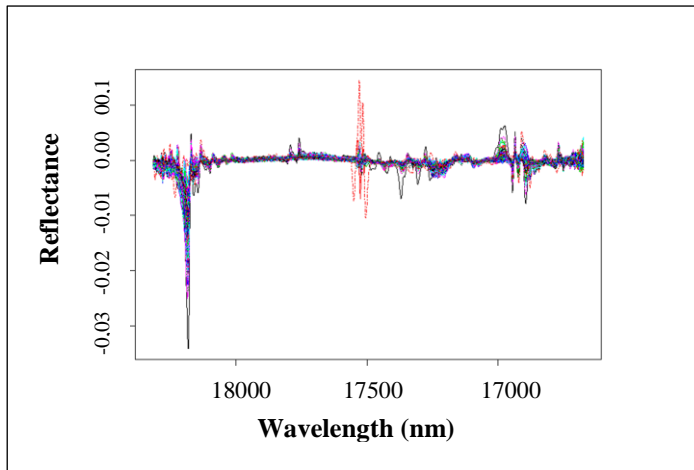
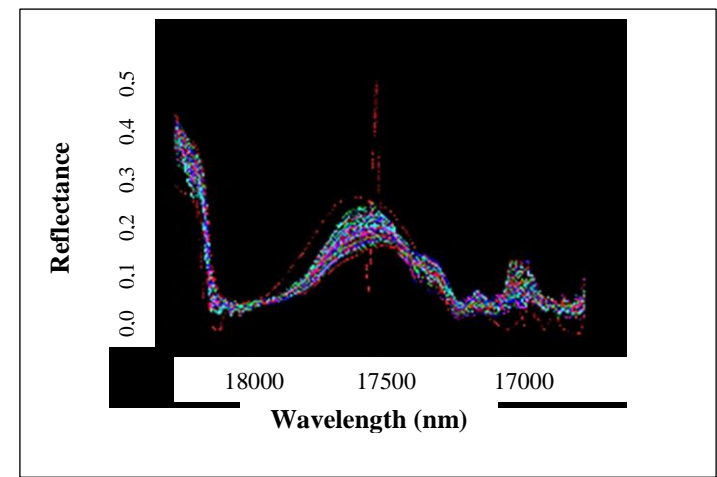
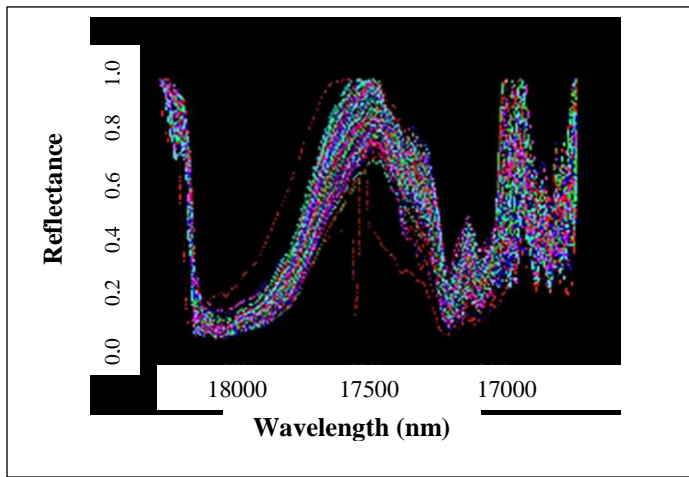
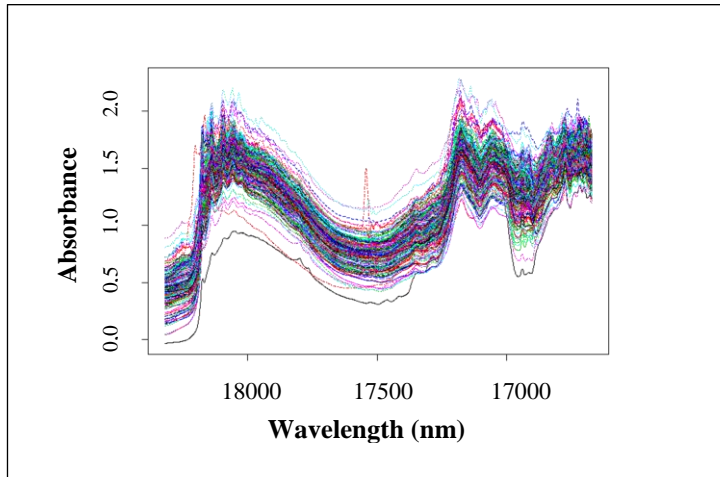


Figura 45- Pré-processamentos aplicados no espectro MIR 1- Absorbance (ABS) 2- Continuum removal (CR) 3- Multiplicative Scatter Correction (MSC) 4- Savitzky Golay Derivative (SGD) e 5- Standard Normal Variate (SNV).

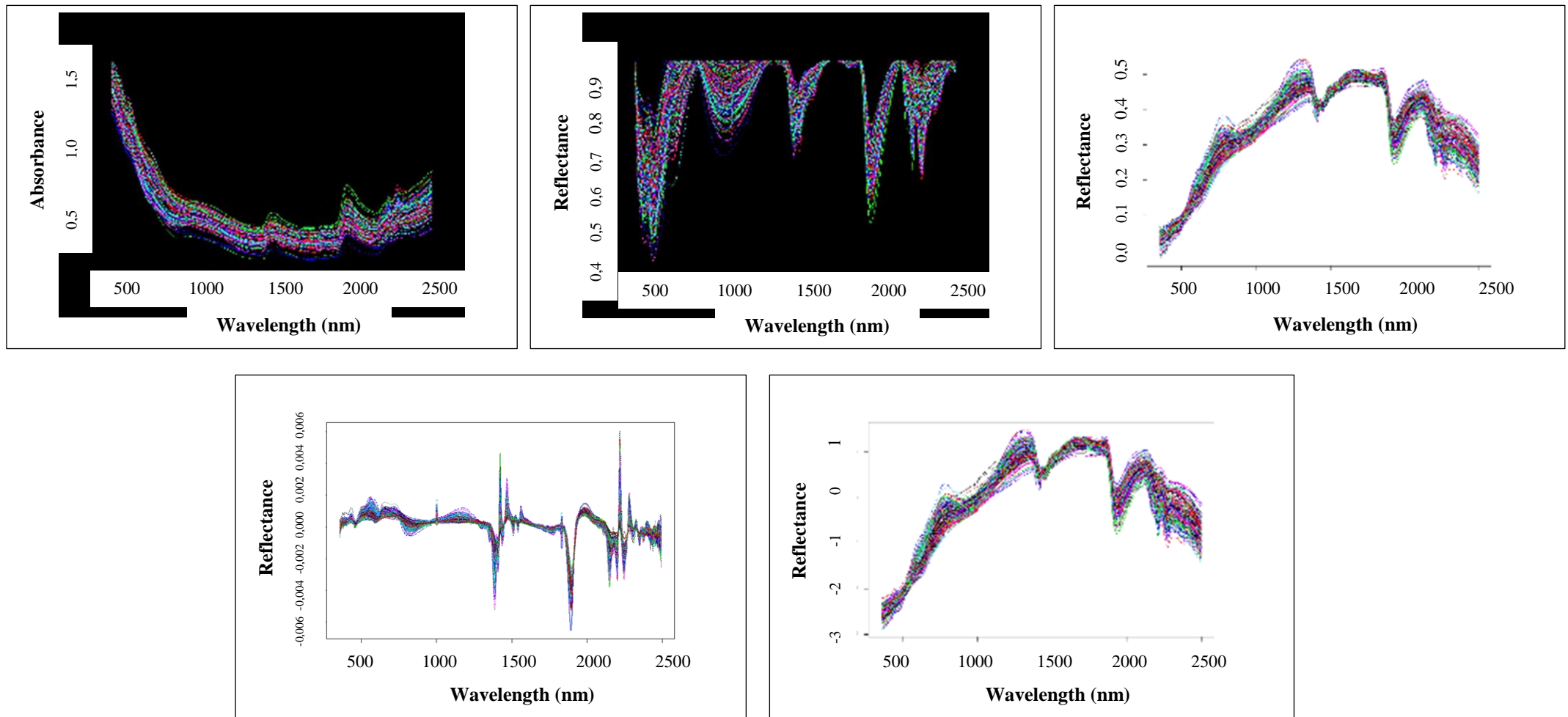


Figura 46- Pré-processamentos aplicados no espectro VIS-NIR 1- Absorbance (ABS) 2- Continuum removal (CR) 3- Multiplicative Scatter Correction (MSC) 4- Savitzky Golay Derivative (SGD) e 5- Standard Normal Variate (SNV).

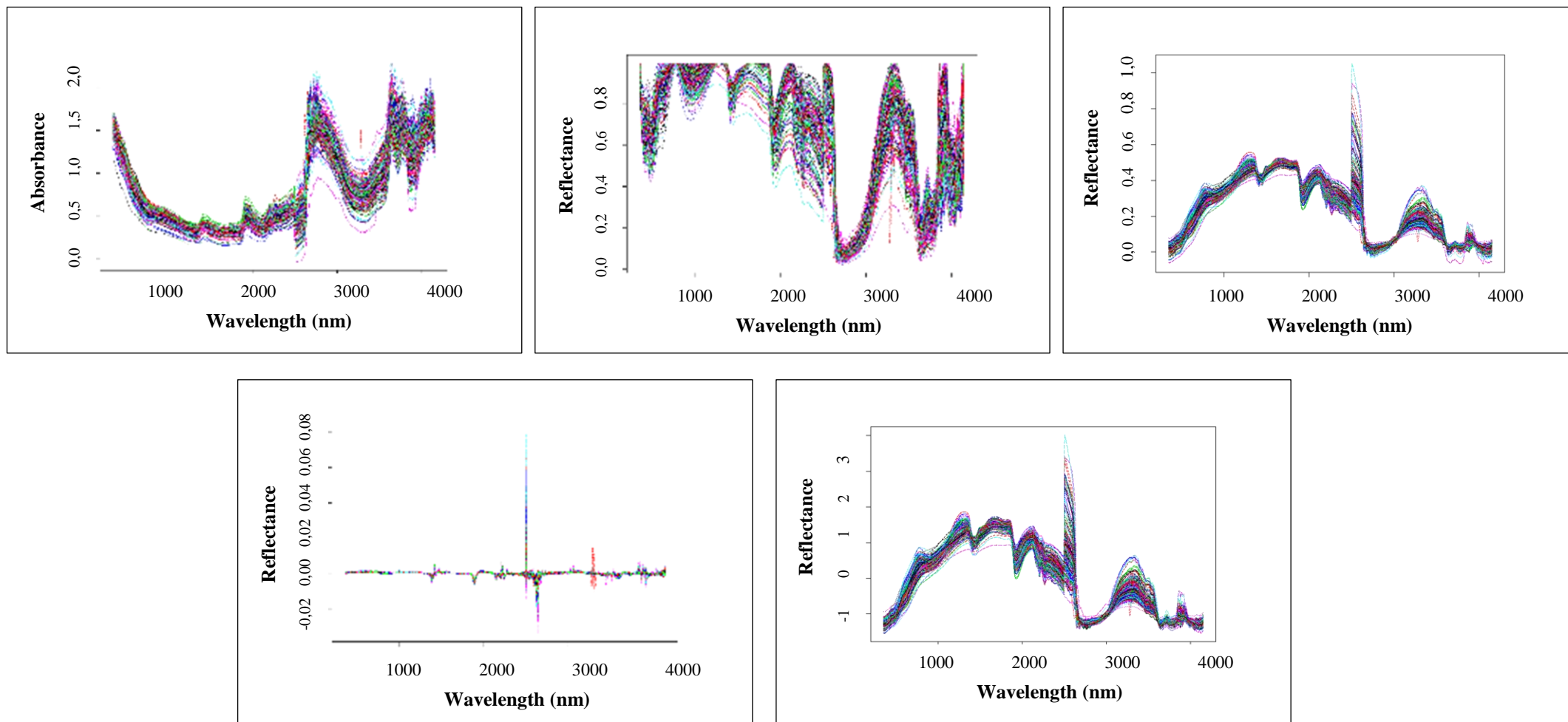


Figura 47- Pré-processamentos aplicados no espectro VIS-NIR-MIR 1- Absorbance (ABS) 2- Continuum removal (CR) 3- Multiplicative Scatter Correction (MSC) 4-Savitzky Golay Derivative (SGD) e 5- Standard Normal Variate (SNV).

As distribuições de frequências são mostradas nos histogramas (Figura 48), onde se verifica que há pequena amplitude na variação dos teores dos três atributos (argila, carbono orgânico e fósforo), tendo os extremos poucos representantes. A estatística descritiva é apresentada na Tabela 38, onde os dados de assimetria e curtose devem estar próximos a 0 e 3, respectivamente (GROENEVELD E MEEDEN, 1984), para satisfazer a condição de normalidade. Sendo importante ressaltar, que nem sempre, devido à natureza de distribuição dos atributos nos solos, estes apresentam distribuição normal (BELLON-MAUREL et al., 2010).

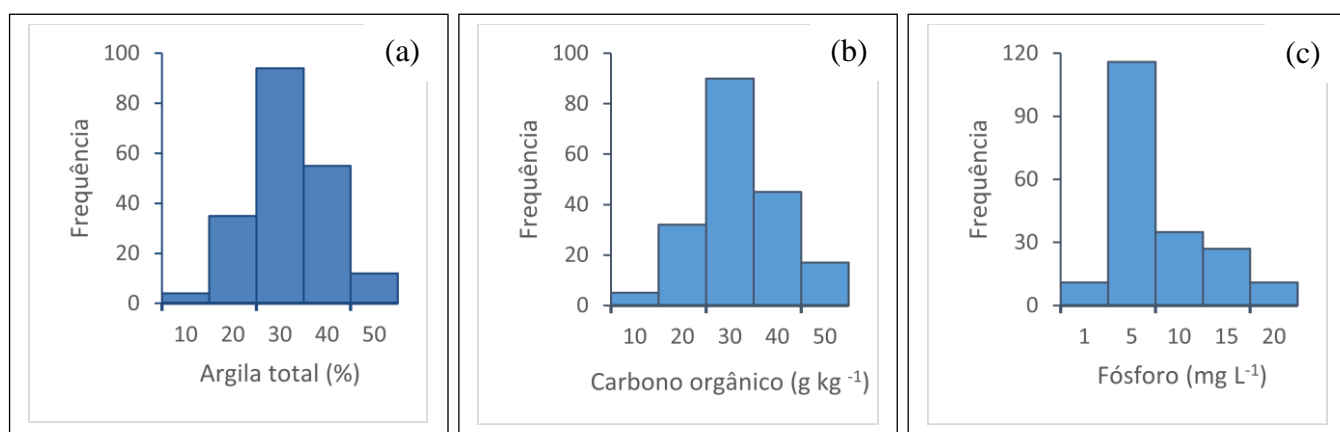


Figura 48- Histogramas de distribuição das três propriedades do solo: (a) argila total; (b) carbono orgânico; (c) fósforo

Bilgili et al. (2010) relatam que as características do solo geralmente apresentam alta variabilidade espacial, mesmo em campos agrícolas únicos, e por isso, normalmente, um grande número de amostras deve ser coletada e analisada para capturar essa variabilidade espacial e estimar adequadamente as propriedades do solo. Resultados positivos, costumam ser alcançados para amostras mais heterogêneas, como as calibrações efetivas alcançadas por Fidencio et al. (2002), Kemper et al. (2005) e Brown et al. (2006).

Shepherd e Walsh (2002) e Brown et al. (2006) usaram a espectroscopia Vis-NIR com base em amostras de vários tipos de solo em grandes áreas geográficas, de forma que este tipo de amostragem geralmente fornece uma ampla gama de valores de indicador de solo, que promove bons resultados regressivos.

Por outro lado, a ampla distribuição dos solos em um conjunto de amostras desafia a metodologia, exigindo uma maior universalidade nas relações de previsão estatística, especialmente quando se incluem materiais parentais diferentes (SHEPHERD e WALSH, 2002). Se as amostras a serem medidas contêm efeitos interferentes que não estão incluídos no conjunto de calibração, os modelos existentes podem ser inválidos para prever os mesmos (JIA et al., 2016), somado ao fato que uma vez calibrado, a escala de aplicação pode limitar a utilidade dos dados (BILGILI et al 2010).

Uma maior variabilidade na fase de treinamento de um modelo estatístico leva a uma robustez melhorada do modelo e a uma maior capacidade de caracterizar uma variedade diversa de amostras, contudo pode também resultar em uma diminuição da precisão da predição (McCARTY et al., 2002). Minasny et al. (2009), em seu estudo

concluíram que é importante ter amostras suficientes e deixar uma parte dos dados como validação, para testar a precisão da predição do MIR (MINASNY et al., 2009).

Hartmann e Appel (2006) apontam que um modelo de previsão de carbono orgânico largamente aplicável, deve basear-se em uma ampla gama de valores dessa variável e solos com diferentes contextos mineralógicos, uma vez que a mineralogia do solo é um dos principais fatores que causam diferenças na reflectância do mesmo.

A quantificação precisa de diferentes propriedades do solo foi obtida a partir de grandes bibliotecas, com amplo número de amostras (BROWN et al., 2005; FERNANDEZ-PIERNA e DARDENNE, 2006; VASQUES et al., 2008; GENOT et al., 2011). Viscarra Rossel e Webster (2012) usaram 21 493 espectros de amostras de solo de aproximadamente 4 000 perfis para prever 24 propriedades do solo. Brown et al. (2006) aplicou o Vis-NIR para caracterizar várias propriedades do solo de mais de 4 500 amostras dos Estados Unidos, África, Ásia, América do Sul e Central e Europa. Shepherd e Walsh (2002) inferiram carbono orgânico, combinando Vis-Nir e *Multivariate Adaptive Regression Splines* (MARS) utilizando mais de 1 000 amostras do solo do leste e sul da África.

Uma revisão de Viscarra Rossel et al. (2006) mostraram que muitos estudos não possuem o procedimento de validação apropriado, por exemplo, para determinar a capacidade de predição do MIR, e isso se deve provavelmente ao pequeno de número de amostras usadas na calibração. Também há uma questão de decidir entre desenvolver funções de calibração usando uma combinação de conjuntos de dados para que a predição funcione em uma grande variedade de tipos de solo ou desenvolver funções de calibração específicas para uma região (MINASNY et al., 2009), como é o caso deste estudo.

Tabela 38- Estatística descritiva para os atributos argila, fósforo (P) e carbono orgânico (CO)

	Argila	P	CO
Observações	184	184	184
Mínimo	8,2	0,05	6,39
Máximo	46,9	19,41	59,22
Média	26,8	5,4	28,44
Mediana	27,28	3,52	27,93
Desvio padrão	7,97	4,72	10,19
Assimetria	-0,04	1,09	0,64
Curtose	2,73	3,21	3,51

Analisando-se o coeficiente de determinação (R^2) para a variável argila no espectro MIR, o melhor resultado ($R^2 = 0,50$) se deu para o pré-processamento SNV e o preditor RF no conjunto de validação (Figura 49), e SGD e SVM linear ($R^2=1,00$) no conjunto de treinamento (Tabela 39). Embora os conjuntos de dados sejam do mesmo domínio espacial, Dardanne et al. (2000) relatam que é incomum obter níveis de previsão similares quando amostras independentes são usadas para validação. Para $R^2 = 0,50$ o RMSE (Tabela 40) foi de 5,61 e para $R^2 = 1,00$ foi de 0,23.

Para essa mesma variável os melhores R^2 (Figura 49) foram encontrados no espectro Vis-NIR ($R^2 = 0,69$, RMSE=4,38) e Vis-NIR-MIR ($R^2 = 0,54$, RMSE = 5,88) no conjunto de validação, tendo o conjunto de treinamento alcançado $R^2 = 1,00$. Viscarra Rossel et al. (2006) acharam em seu trabalho pequenas melhorias nas previsões de argila

usando a gama combinada Vis-NIR-MIR. Mohamed et al. (2017) apontam que os espectros Vis-Nir contêm informações valiosas para prever a textura do solo, concordando os resultados obtidos. Segundo Hunt (1980) muitos minerais de argila têm reflectância espectral única em comprimentos de onda visíveis e NIR-SWIR (Infravermelho próximo –Infravermelho de ondas curtas).

Para o conjunto Vis-NIR, os melhores valores de R^2 foram obtidos na combinação MSC-SVM linear na validação e SGD- GPR linear (RMSE = 0,49) no treinamento. Já para Vis-NIR-MIR pela combinação SGD-PLSR (validação) e CR e SGD, ambos com GPR linear (RMSE = 0,6 e 0,01; respectivamente) no treinamento.

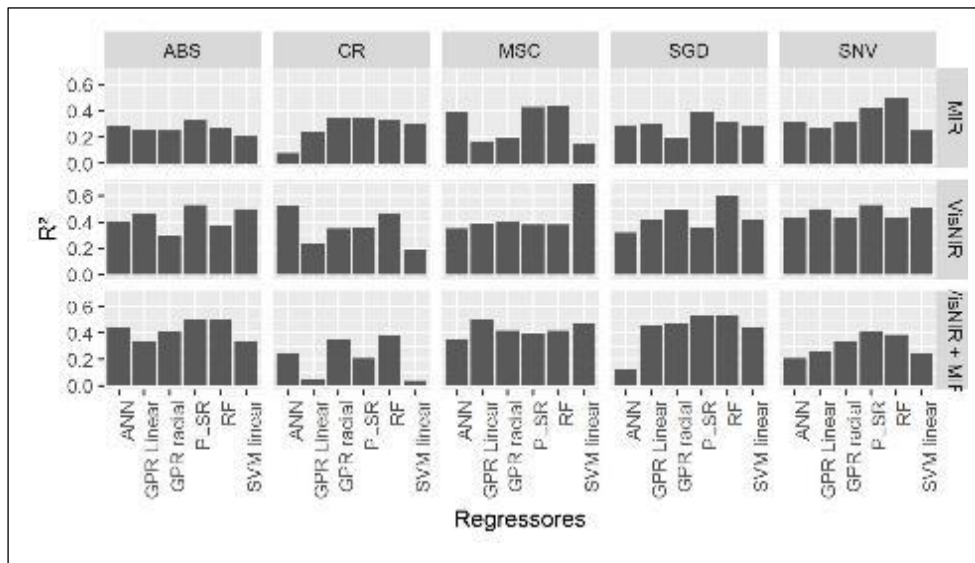


Figura 49- Valores do coeficiente de determinação (R^2), para variável argila., encontrados para os diferentes métodos e pré-processamentos no grupo na validação.

Absorbance (ABS); Continuum removal (CR); Multiplicative Scatter Correction (MSC); Savitzky Golay Derivative (SGD); Standard Normal Variate (SNV) Partial Least Squares Regression (PLSR), Artificial Neural Network (ANN), Random Forest (RF), Gaussian Process Regression (GPR linear e radial) e Support Vector Machine (SVM). Mir (Infravermelho médio); Vis-NIR (Visível e infravermelho próximo) e Vis-NIR-MIR (Visível, infravermelho próximo e médio).

Tabela 39- Valores do coeficiente de determinação (R^2), para variável argila, encontrados para os diferentes métodos e pré-processamentos no grupo no treinamento, nos espectros Mir, Vis-NIR e Vis-NIR-MIR

R^2					R^2				
		Vis-NIR-MIR					Vis-NIR-MIR		
Modelo	Pré-processamento	Mir	Vis-NIR	Treinamento	Modelo	Pré-processamento	Mir	Vis-NIR	Treinamento
PLSR	CR	0,44	0,61	0,82	GPR linear	SGD	1,00	1,00	1,00
ANN	CR	0,10	0,44	0,48	GPR radial	SGD	0,38	0,39	0,49
RF	CR	0,43	0,37	0,47	SVM linear	SGD	0,99	0,99	0,99
GPR linear	CR	0,98	0,95	1,00	PLSR	SNV	0,43	0,69	0,68
GPR radial	CR	0,37	0,38	0,48	ANN	SNV	0,31	0,58	0,42
SVM linear	CR	0,99	0,97	0,99	RF	SNV	0,35	0,53	0,41
PLSR	ABS	0,49	0,70	0,48	GPR linear	SNV	0,99	0,78	0,99
ANN	ABS	0,44	0,39	0,41	GPR radial	SNV	0,37	0,45	0,44
RF	ABS	0,31	0,35	0,39	SVM linear	SNV	0,99	0,77	0,99
GPR linear	ABS	0,94	0,79	0,97	PLSR	MSC	0,40	0,73	0,40
GPR radial	ABS	0,91	0,35	0,42	ANN	MSC	0,36	0,60	0,44
SVM linear	ABS	0,96	0,78	0,99	RF	MSC	0,40	0,49	0,38
PLSR	SGD	0,54	0,74	0,63	GPR linear	MSC	0,98	0,79	0,95
ANN	SGD	0,13	0,28	0,07	GPR radial	MSC	0,97	0,49	0,44
RF	SGD	0,32	0,41	0,44	SVM linear	MSC	0,99	0,70	0,99

Absorbance (ABS); Continuum removal (CR); Multiplicative Scatter Correction (MSC); Savitzky Golay Derivative (SGD); Standard Normal Variate (SNV) Partial Least Squares Regression (PLSR), Artificial Neural Network (ANN), Random Forest (RF), Gaussian Process Regression (GPR linear e radial) e Support Vector Machine (SVM). Mir (Infravermelho médio); Vis-NIR (Visível e infravermelho próximo) e Vis-NIR-MIR (Visível, infravermelho próximo e médio).

Tabela 40- Valores de RMSE, para variável argila., encontrados para os diferentes métodos e pré-processamentos no grupo no treinamento (T) e validação (V), nos espectros Mir, Vis-NIR e Vis-NIR-MIR

		RMSE					
		Mir		Vis-NIR		Vis-NIR-MIR	
Modelo	Pré-processamento	T	V	T	V	T	V
PLSR	CR	5,94	6,57	4,99	6,53	3,58	7,64
ANN	CR	7,59	7,81	6,01	5,37	6,02	6,47
RF	CR	6,05	6,67	6,39	5,66	6,18	5,41
GPR linear	CR	1,12	8,14	1,80	8,76	0,60	17,23
GPR radial	CR	6,90	7,15	6,65	6,41	6,46	5,68
SVM linear	CR	0,81	7,73	1,48	8,93	0,81	18,42
PLSR	ABS	5,66	6,47	4,40	5,72	5,69	5,71
ANN	ABS	5,96	6,77	6,24	6,15	6,07	6,02
RF	ABS	6,63	6,82	6,43	6,24	6,25	6,08
GPR linear	ABS	2,04	7,50	3,68	6,05	1,40	7,79
GPR radial	ABS	3,93	7,04	6,82	6,81	6,55	6,73
SVM linear	ABS	1,52	8,47	3,72	5,71	0,93	8,31
PLSR	SGD	5,55	5,84	4,23	6,14	4,66	5,88
ANN	SGD	7,65	6,37	6,73	6,70	7,40	8,06
RF	SGD	7,03	6,33	6,16	5,40	5,97	6,17
GPR linear	SGD	0,23	7,04	0,49	7,85	0,01	6,52
GPR radial	SGD	7,62	7,02	6,64	6,42	6,15	6,78
SVM linear	SGD	0,83	7,06	0,78	8,03	0,76	6,67
PLSR	SNV	6,07	5,90	4,52	5,28	4,52	6,44
ANN	SNV	6,71	6,43	5,24	5,93	6,06	7,31
RF	SNV	6,59	5,61	5,61	5,72	6,05	6,52
GPR linear	SNV	0,94	8,30	3,76	5,53	0,90	8,88
GPR radial	SNV	7,29	6,82	6,34	6,01	6,22	6,93
SVM linear	SNV	0,82	8,97	3,92	5,59	0,77	9,13
PLSR	MSC	6,43	5,16	4,15	6,82	5,82	6,81
ANN	MSC	6,65	5,33	5,03	6,70	5,67	7,10
RF	MSC	6,58	5,16	5,73	6,24	6,01	6,78
GPR linear	MSC	1,16	8,98	3,62	6,67	1,06	6,65
GPR radial	MSC	3,88	6,37	6,13	6,23	6,09	7,18
SVM linear	MSC	0,88	9,77	4,49	4,38	0,74	7,02

Absorbance (ABS); Continuum removal (CR); Multiplicative Scatter Correction (MSC); Savitzky Golay Derivative (SGD); Standard Normal Variate (SNV) Partial Least Squares Regression (PLSR), Artificial Neural Network (ANN), Random Forest (RF), Gaussian Process Regression (GPR linear e radial) e Support Vector Machine (SVM). Mir (Infravermelho médio); Vis-NIR (Visível e infravermelho próximo) e Vis-NIR-MIR (Visível, infravermelho próximo e médio).

Em MIR os maiores valores de RPIQ (Figura 50) e RPD (Figura 51) no conjunto validação, foram encontrados para a combinação pré-processamento SGD e preditor PLSR (1,82), e MSC com os preditores PLSR e RF (1,38), respectivamente. Pela análise do índice RPD tem-se um modelo “não confiável” (RPD menor que 1,4), enquanto o RPIQ eleva a predição para modelos razoáveis.

Nos outros espectros, foram encontrados valores máximos iguais a 2,15 e 1,62 em Vis-NIR e 1,87 e 1,31 em Vis-NIR-MIR, para RPIQ e RPD, respectivamente. Esses foram dados pela combinação MSC-SVM linear (2,15 e 1,62), MSC-GPR linear (1,87) e CR-RF (1,31). Pela análise de RPIQ os modelos foram classificados como excelentes (2,15) e razoáveis (1,87), enquanto RPD os classificou como razoáveis (1,62) e não confiáveis (menores que 1,4).

Em relação a capacidade de predição analisada dentro de cada pré-processamento, segundo os valores de RPIQ e RPD, percebeu-se que para a variável argila, PLSR foi

superior em todas as faixas de espectro, exceto para RPD na faixa Vis-NIR, cujo SVM foi superior. Sendo importante ressaltar, que apesar de ter se apresentado como método que atingiu mais vezes maiores valores dos índices, o valor máximo foi encontrado pelos preditores relacionados acima.

Em relação aos pré-processamentos para argila, SGD se destacou com maior parte dos valores de RPIQ superior a 1,4 na faixa do MIR, assim como MSC e SNV no espectro Vis-NIR e MSC em Vis-NIR-MIR. Na faixa do Vis-NIR foram encontrados valores de RPIQ superiores a 2,0. Quanto ao RPD, MSC foi superior em Mir; SNV em Vis-NIR e ABS, CR e SGD mostraram-se eficientes em Vis-NIR-MIR. Valores de RPD superiores a 1,4 foram encontrados apenas para MSC na faixa de Vis-NIR. Assim como na faixa do MIR, os modelos ao serem classificados pelo índice RPIQ apresentaram qualidade superior ao determinado por RPD, sendo aqueles com índice igual ou maior que 2,0 considerado acurado para predição do atributo.

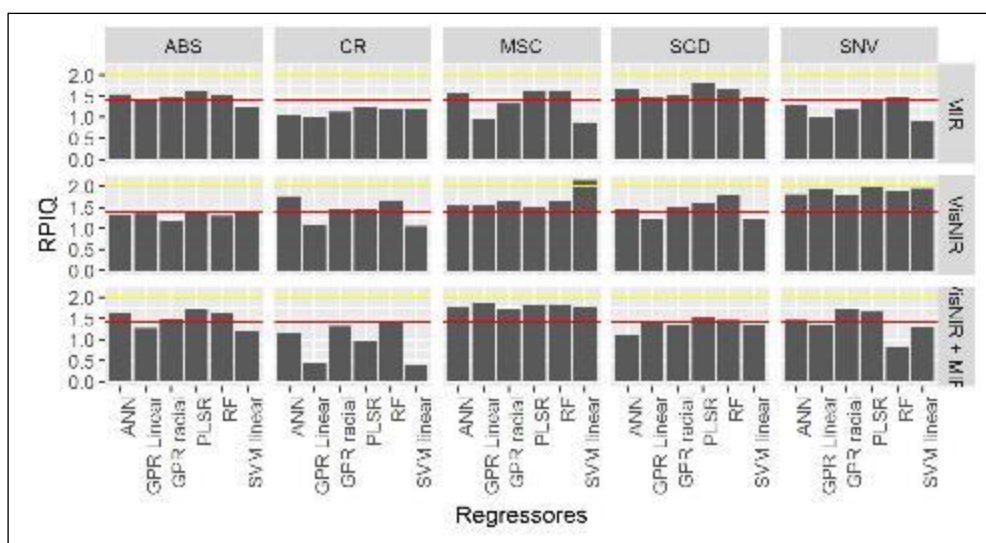


Figura 50- Valores de RPIQ encontrados para os diferentes métodos e pré-processamentos no grupo na validação, para variável argila.

Absorbance (ABS); Continuum removal (CR); Multiplicative Scatter Correction (MSC); Savitzky Golay Derivative (SGD); Satandard Normal Variate (SNV) Partial Least Squares Regression (PLSR), Artificial Neural Network (ANN), Random Forest (RF), Guassian Process Regression (GPR linear e radial) e Support Vector Machine (SVM). Mir (Infravermelho médio); Vis-NIR (Visível e infravermelho próximo) e Vis-NIR-MIR (Visível, infravermelho próximo e médio).

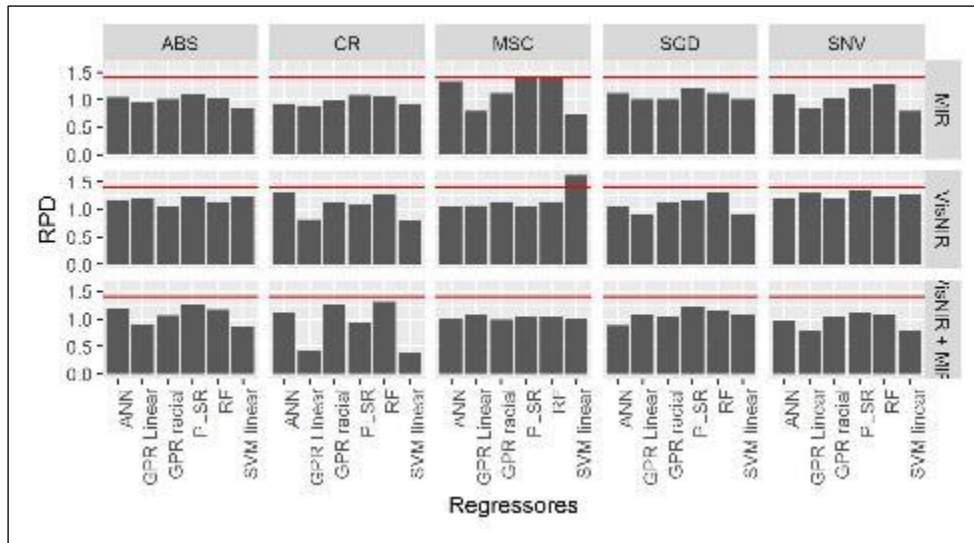


Figura 51- Valores de RPD encontrados para os diferentes métodos e pré-processamentos no grupo na validação, para variável argila.

Absorbance (ABS); Continuum removal (CR); Multiplicative Scatter Correction (MSC); Savitzky Golay Derivative (SGD); Satandard Normal Variate (SNV) Partial Least Squares Regression (PLSR), Artificial Neural Network (ANN), Random Forest (RF), Guassian Process Regression (GPR linear e radial) e Support Vector Machine (SVM). Mir (Infravermelho médio); Vis-NIR (Visível e infravermelho próximo) e Vis-NIR-MIR (Visível, infravermelho próximo e médio).

Para predição de argila, alguns resultados satisfatórios foram encontrados por Ben-Dor & Banin (1995); Chang et al. (2001) e Morón & Cozzolino (2003). Ben-Dor e Banin (1995) e Islan et al. (2003) modelaram argila a partir do espectro NIR, achando R^2 de 0,56 e 0,75 respectivamente. Chang et al (2001) e Shepherd e Walsh (2002) trabalharam no espectro Vis-NIR, obtendo valores ligeiramente melhores de 0,67 e 0,78. Silva et al. (2016) mostraram a possibilidade de se predizer argila na região do infravermelho próximo utilizando PLSR.

Viscarra Rossel et al. (2006), modelando argila encontrou valores de R^2 iguais a 0,43 (Vis); 0,60 (Nir); 0,67 (Mir) e 0,67 (Vis-NIR-MIR), KANIA E GRUBA (2016), testando a predição de argila pelo espectro Nir, encontrou valores de R^2 para calibração de 0,67 e 0,95; e para validação de 0,57 e 0,21; Valores esses inferiores aos encontrados nesse trabalho, enquanto Ben-Dor e Banin (1995) obtiveram valores superiores, com R^2 de 0,86.

As Figuras 52 e 53 mostram a distribuição dos dados dentro de cada preditor nas diferentes faixas do espectro para RPIQ E RPD, respectivamente, para a variável argila. Em ambos é possível observar a boa distribuição dos dados, tendo RPIQ maior variabilidade em relação a RPD. SVM linear foi o método que apresentou maior inconstância, em ambos índices.

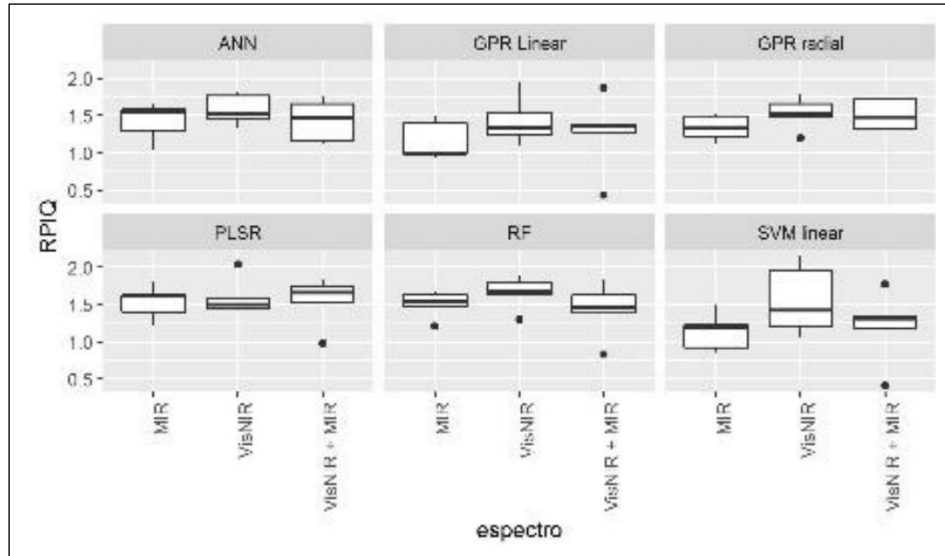


Figura 52- Box plot da distribuição dos dados de RPIQ nos diferentes pré-processamentos e faixas do espectro, para variável argila.

Partial Least Squares Regression (PLSR), Artificial Neural Network (ANN), Random Forest (RF), Gaussian Process Regression (GPR linear e radial) e Support Vector Machine (SVM). Mir (Infravermelho médio); Vis-NIR (Visível e infravermelho próximo) e Vis-NIR-MIR (Visível, infravermelho próximo e médio).

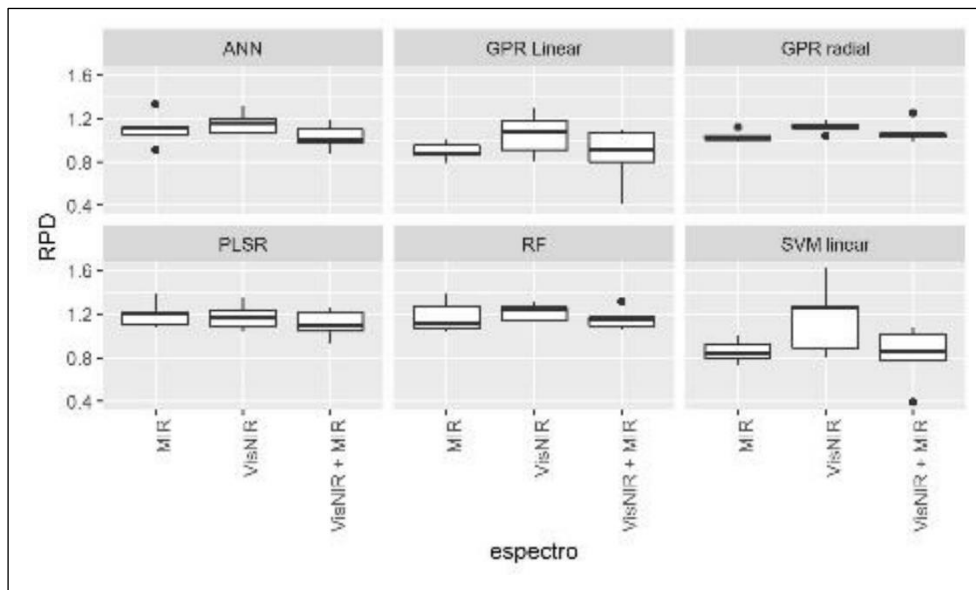


Figura 53- Box plot da distribuição dos dados de RPD nos diferentes pré-processamentos e faixas do espectro, para variável argila.

Partial Least Squares Regression (PLSR), Artificial Neural Network (ANN), Random Forest (RF), Gaussian Process Regression (GPR linear e radial) e Support Vector Machine (SVM). Mir (Infravermelho médio); Vis-NIR (Visível e infravermelho próximo) e Vis-NIR-MIR (Visível, infravermelho próximo e médio).

Segundo Viscarra Rossel e McBratney (1998) os comprimentos de onda significativos para estimar o teor de argila na faixa NIR são 1 600 nm, 1 800 nm, 2 000

nm e 2 100 nm . Já Nawara et al. (2016) relataram que o comprimento de onda de 2206 nm seria a banda ideal para quantificação deste atributo. Nesse trabalho encontrou-se em Vis-NIR, como variáveis mais importantes na predição (Figura 54) os espectros em torno de 1 500 nm, 1 800 nm e 2 100 nm , concordando com o encontrado por Viscarra Rossel e McBratney (1998).

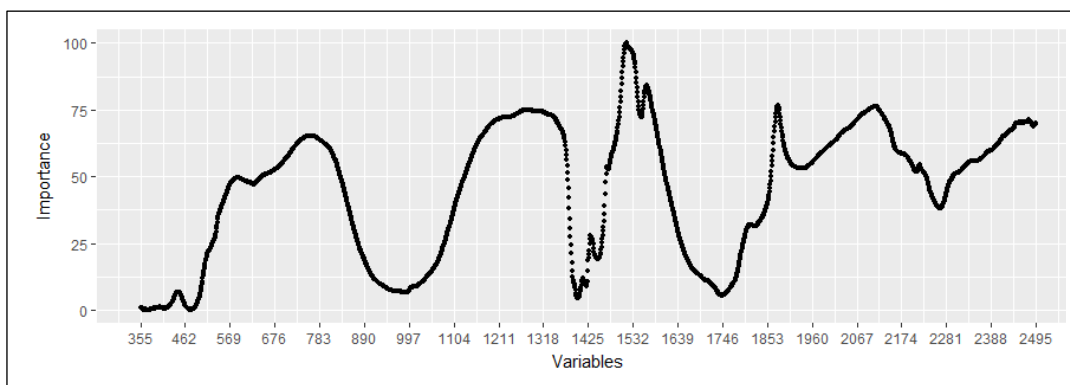


Figura 54- Variáveis importantes para predição de argila, no espectro Vis-NIR (Visível e infravermelho próximo) usando pré-processamento *Multiplicative Scatter Correction* (MSC) e modelo *Support Vector Machine* (SVM).

Para a variável carbono orgânico (CO), no espectro MIR, os melhores valores de R^2 (Figura 55 e Tabela 41) e RMSE (Tabela 42), foram iguais a 0,65 (RMSE =8,31) e 1,00 (RMSE= 0,03) para validação e treinamento, respectivamente. Sendo estes obtidos pela combinação CR-PLSR (0,65) e SGD-GPR linear (1,00). Os maiores índices RPIQ e RPD (Figuras 56 e 57) foram 1,58 e 1,32, sendo o primeiro obtido pelo pré-processamento SGD e preditor PLSR; e o segundo pela combinação SNV-PLSR. Ben-Dor e Banin (1995) obtiveram valores de R^2 de 0,55 para matéria orgânica usando um conjunto de amostras separado para testar a precisão da predição, valor este inferior aos encontrados neste trabalho.

Dentro de Vis-NIR e Vis-NIR-MIR os valores de R^2 (Figura 55 e Tabela 41) foram ligeiramente inferiores a MIR para o conjunto validação e se mantiveram para o conjunto de treinamento, já os valores de RPIQ (Figura 56) e RPD (Figura 57) foram levemente superiores. De acordo com Vohland et al. (2014) os mecanismos físicos diferem basicamente entre os domínios Vis-NIR e MIR, enquanto as vibrações moleculares fundamentais dos componentes do solo podem ser medidas apenas no MIR, em NIR são detectadas as repercussões e as combinações dessas vibrações fundamentais.

Em Vis-NIR encontrou-se R^2 igual a 0,63 (RMSE = 8,18) e em Vis-NIR-MIR igual a 0,62 (RMSE=8,28) dados pela combinação SGD-RF e CR-ANN, respectivamente; enquanto no conjunto de treinamento foi igual a 1,00 dado por SGD-GPR linear para ambos (RMSE = 0,66 e 0,02; para Vis-NIR e Vis-NIR-MIR, respectivamente), tendo GR-GPR linear também obtido R^2 igual a 1,00 (RMSE =0,8) no espectro Vis-NIR-MIR.

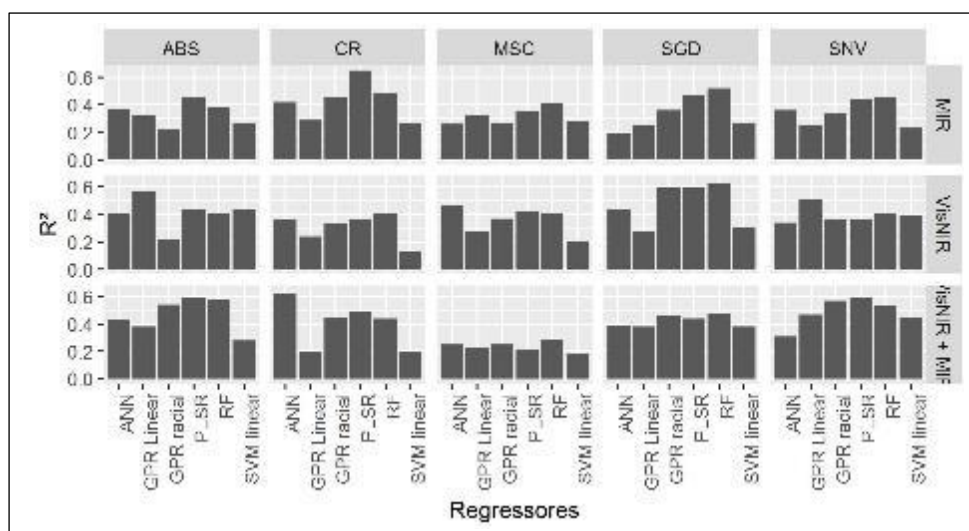


Figura 55- Valores do coeficiente de determinação (R^2), encontrados para variável carbono orgânico, nos diferentes métodos e pré-processamentos no grupo na validação.

Absorbance (ABS); Continuum removal (CR); Multiplicative Scatter Correction (MSC); Savitzky Golay Derivative (SGD); Satandard Normal Variate (SNV) Partial Least Squares Regression (PLSR), Artificial Neural Network (ANN), Random Forest (RF), Guassian Process Regression (GPR linear e radial) e Support Vector Machine (SVM). Mir (Infravermelho médio); Vis-NIR (Visível e infravermelho próximo) e Vis-NIR-MIR (Visível, infravermelho próximo e médio).

Tabela 41- Valores do coeficiente de determinação (R^2), para variável carbono orgânico, encontrados para os diferentes métodos e pré-processamentos no grupo no treinamento, nos espectros Mir, Vis-NIR e Vis-NIR-MIR

Modelo	Pré-processamento	R^2			Modelo	Pré-processamento	R^2		
		Mir	Vis-NIR	Vis-NIR-MIR			Mir	Vis-NIR	Vis-NIR-MIR
PLSR	CR	0,47	0,46	0,61	GPR Linear	SGD	1,00	1,00	1,00
ANN	CR	0,41	0,48	0,48	GPR radial	SGD	0,50	0,41	0,52
RF	CR	0,35	0,44	0,48	SVM linear	SGD	0,99	0,99	0,99
GPR Linear	CR	0,98	0,93	1,00	PLSR	SNV	0,55	0,49	0,45
GPR radial	CR	0,39	0,46	0,54	ANN	SNV	0,47	0,48	0,53
SVM linear	CR	0,98	0,93	0,99	RF	SNV	0,33	0,33	0,30
PLSR	ABS	0,50	0,46	0,45	GPR Linear	SNV	0,98	0,70	0,98
ANN	ABS	0,45	0,45	0,50	GPR radial	SNV	0,37	0,44	0,33
RF	ABS	0,32	0,26	0,24	SVM linear	SNV	0,99	0,74	0,99
GPR Linear	ABS	0,92	0,63	0,97	PLSR	MSC	0,52	0,56	0,63
GPR radial	ABS	0,90	0,92	0,35	ANN	MSC	0,34	0,43	0,57
SVM linear	ABS	0,94	0,65	0,99	RF	MSC	0,36	0,34	0,43
PLSR	SGD	0,39	0,48	0,68	GPR Linear	MSC	0,97	0,73	0,98
ANN	SGD	0,31	0,35	0,36	GPR radial	MSC	0,41	0,39	0,46
RF	SGD	0,41	0,34	0,41	SVM linear	MSC	0,98	0,70	0,99

Absorbance (ABS); Continuum removal (CR); Multiplicative Scatter Correction (MSC); Savitzky Golay Derivative (SGD); Satandard Normal Variate (SNV) Partial Least Squares Regression (PLSR), Artificial Neural Network (ANN), Random Forest (RF), Guassian Process Regression (GPR linear e radial) e Support Vector Machine (SVM). Mir (Infravermelho médio); Vis-NIR (Visível e infravermelho próximo) e Vis-NIR-MIR (Visível, infravermelho próximo e médio).

Tabela 42- Valores de RMSE, para variável carbono orgânico, encontrados para os diferentes métodos e pré-processamentos no grupo no treinamento (T) e validação (V), nos espectros Mir, Vis-NIR e Vis-NIR-MIR

Modelo	Pré-processamento	RMSE					
		Mir		Vis-NIR		Vis-NIR-MIR	
		T	V	T	V	T	V
PLSR	CR	7,14	8,31	7,79	8,93	6,75	8,00
ANN	CR	7,52	9,85	7,62	9,01	7,12	8,28
RF	CR	7,92	9,75	7,93	8,62	7,79	8,26
GPR Linear	CR	1,36	12,25	2,87	11,52	0,80	14,15
GPR radial	CR	8,39	10,85	8,25	9,45	8,21	8,48
SVM linear	CR	1,47	12,88	2,89	13,37	1,04	14,13
PLSR	ABS	7,28	8,69	7,65	8,69	7,72	7,53
ANN	ABS	7,59	9,39	7,72	8,87	7,49	8,54
RF	ABS	8,49	9,66	8,91	9,13	9,06	8,10
GPR Linear	ABS	2,92	10,24	6,33	7,84	2,04	10,15
GPR radial	ABS	4,82	10,63	5,12	8,71	9,02	9,60
SVM linear	ABS	2,47	11,41	6,77	8,06	1,17	11,73
PLSR	SGD	8,17	8,43	7,43	7,76	6,16	8,17
ANN	SGD	8,70	10,84	8,27	9,09	8,26	9,18
RF	SGD	8,48	9,00	8,39	8,18	8,37	7,61
GPR Linear	SGD	0,03	11,48	0,66	12,48	0,02	9,71
GPR radial	SGD	9,42	10,28	8,42	9,45	8,30	8,21
SVM linear	SGD	1,03	11,12	1,01	12,34	1,05	9,33
PLSR	SNV	7,45	8,11	7,61	8,82	7,31	8,03
ANN	SNV	7,65	8,71	7,64	9,01	7,40	9,19
RF	SNV	8,67	8,40	8,69	8,52	8,26	8,88
GPR Linear	SNV	1,53	10,69	5,87	7,90	1,61	9,55
GPR radial	SNV	8,93	9,51	8,66	9,33	8,58	10,43
SVM linear	SNV	1,05	11,63	5,58	8,90	0,98	10,02
PLSR	MSC	7,16	9,50	7,28	8,30	0,63	10,23
ANN	MSC	8,38	10,06	8,25	7,71	7,44	8,09
RF	MSC	8,27	9,03	8,87	8,01	8,37	8,75
GPR Linear	MSC	1,72	11,70	5,69	10,23	1,71	11,15
GPR radial	MSC	8,58	10,30	9,06	8,58	8,74	8,94
SVM linear	MSC	1,47	12,98	5,95	11,40	1,08	12,57

Absorbance (ABS); Continuum removal (CR); Multiplicative Scatter Correction (MSC); Savitzky Golay Derivative (SGD); Standard Normal Variate (SNV) Partial Least Squares Regression (PLSR), Artificial Neural Network (ANN), Random Forest (RF), Gaussian Process Regression (GPR linear e radial) e Support Vector Machine (SVM). Mir (Infravermelho médio); Vis-NIR (Visível e infravermelho próximo) e Vis-NIR-MIR (Visível, infravermelho próximo e médio).

No que diz respeito aos valores de RPIQ (Figura 56) e RPD (Figura 57), para a variável carbono orgânico, PLSR foi superior para Mir e Vis-NIR-MIR, e GPR linear foi superior na faixa Vis-NIR. Em relação aos pré-processamentos para carbono orgânico, SGD se destacou com maior parte dos valores de RPIQ superior a 1,4 na faixa do MIR e Vis-NIR-MIR, e SNV no espectro Vis-NIR-MIR. Quanto ao RPD apenas a faixa Vis-NIR-MIR apresentou valores acima de 1,4 com ABS e SGD se destacando como os melhores pré-processamentos para a variável.

RPIQ apresentou valores máximos iguais a 1,74 (Vis-NIR) e 1,84 (Vis-NIR-MIR), e RPD valores de 1,39 (Vis-NIR) e 1,42 (Vis-NIR-MIR), estes foram obtidos pelo preditor PLSR, combinado ao pré-processamento SGD (1,74), SNV (1,84) e ABS (1,42), o RPD de 1,39 foi alcançado combinando-se MSC-ANN. Ao contrário do encontrado nesse estudo, cujo PLSR se destacou, Xiaoting et al. (2014), trabalharam com o preditor SVM, nos espectros Vis-NIR como forma de melhorar a precisão de carbono orgânico do solo, combinando-o com um modelo de algoritmo de projeção (SPA).

Brown et al. (2006) também apontam que outras técnicas de mineração de dados como as Árvores de Regressão Impulsionadas (BRT) podem superar o PLSR, devido à sua capacidade de incluir interações e relacionamentos não-lineares. Outros trabalhos são citados por aplicarem diferentes algoritmos sucesso (MOUAZEN et al., 2010; VISCARRA ROSSEL e BEHRENS, 2010; VOHLAND et al., 2011).

Entretanto é enfatizado por Adeline et al. (2017) que o PLSR é o método estatístico multivariado mais comumente utilizado na ciência do solo, sendo destaque para calibração da reflectância do solo para estimação de propriedades dos mesmos (VISCARRA ROSSEL et al., 2006), principalmente devido à sua superioridade em relação aos métodos tradicionais para lidar com a multicolinearidade em dados de alta dimensão (BILGILI et al., 2010).

Bilgili et al (2010), trabalhou com 512 amostras e diferentes solos, conseguindo prever matéria orgânica com R^2 igual a 0,73 a partir de Vis-NIR com pré-processamento dos dados usando SGD. Tendo os valores de R^2 no conjunto de validação sido menores do que os valores correspondentes no conjunto de calibração.

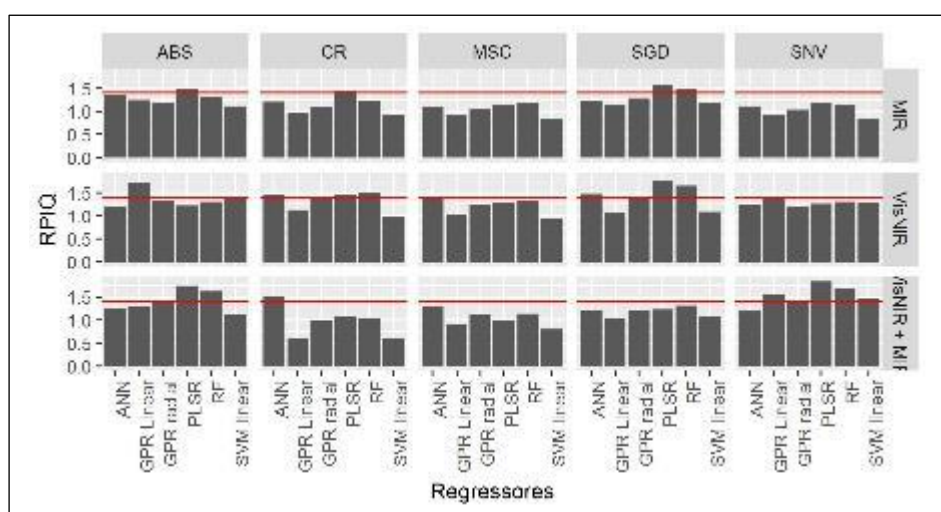


Figura 56- Valores de RPIQ encontrados para os diferentes métodos e pré-processamentos no grupo na validação, para variável carbono orgânico.

Absorbance (ABS); Continuum removal (CR); Multiplicative Scatter Correction (MSC); Savitzky Golay Derivative (SGD); Standard Normal Variate (SNV) Partial Least Squares Regression (PLSR), Artificial Neural Network (ANN), Random Forest (RF), Gaussian Process Regression (GPR linear e radial) e Support Vector Machine (SVM). Mir (Infravermelho médio); Vis-NIR (Visível e infravermelho próximo) e Vis-NIR-MIR (Visível, infravermelho próximo e médio).

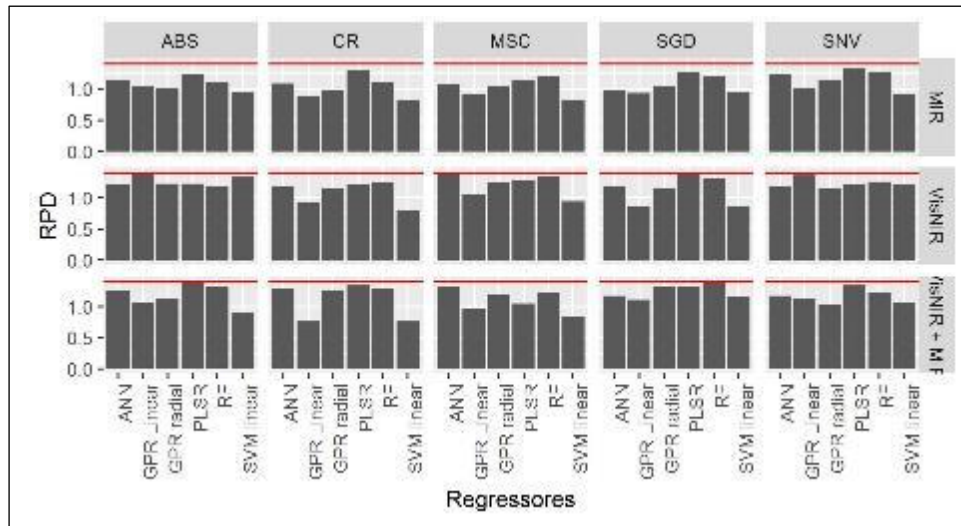


Figura 57- Valores de RPD encontrados para os diferentes métodos e pré-processamentos no grupo na validação, para variável carbono orgânico.

Absorbance (ABS); Continuum removal (CR); Multiplicative Scatter Correction (MSC); Savitzky Golay Derivative (SGD); Standard Normal Variate (SNV) Partial Least Squares Regression (PLSR), Artificial Neural Network (ANN), Random Forest (RF), Gaussian Process Regression (GPR linear e radial) e Support Vector Machine (SVM). Mir (Infravermelho médio); Vis-NIR (Visível e infravermelho próximo) e Vis-NIR-MIR (Visível, infravermelho próximo e médio).

Há uma série de trabalhos que demonstram a possibilidade de predição do carbono orgânico com base no MIR (ZIMMERMANN et al., 2007; BORNEMANN et al., 2008; YANG et al., 2012). A pesquisa que compara o MIR e o NIR nas mesmas amostras demonstrou que constantemente o MIR supera o NIR na análise dos solos, principalmente para as frações C e N (MADARI et al., 2005; McCARTY et al., 2002; REEVES et al., 2002).

Bellon-Maurel e McBratney (2011) analisaram as técnicas espectroscópicas NIR e MIR para estudos de carbono do solo, e destacaram que a espectroscopia MIR produziu melhores modelos (entre 10 a 40% de melhoria) em relação à aqueles desenvolvidos a partir de espectros NIR. Porém, dentre os estudos analisados, nenhum deles comparou a espectroscopia NIR e MIR nas mesmas amostras de solo (KNOX et al., 2015).

Em geral, MIR é considerado superior ao Vis-NIR (VOHLAND et al., 2014). McCarty et al. (2002) e McCarty e Reeves (2006) também encontraram MIR melhor do que NIR e Vis-NIR para a estimativa dos teores de carbono do solo. Porém, essa superioridade não é reconhecida em todos os trabalhos (MADARI et al.,2006; MICHEL et al., 2009 e LUDWIG et al.,2008). Nesse trabalho, a melhor predição de carbono orgânico foi obtida no espectro MIR, em relação a R^2 , porém os melhores valores de RPIQ e RPD foram para Vis-NIR. Segundo Ludwig et al. (2016) a espectroscopia MIR é um método estabelecido para estimar o conteúdo de carbono orgânico do solo.

Isso pode estar relacionado ao fato da separação de cada contribuição dos componentes do solo nos espectros Vis-NIR, ser uma tarefa desafiadora devido à natureza complexa da matriz do solo com múltiplas sobreposições de características espectrais, além das fortes colinearidades existentes entre as propriedades do solo (GOBRECHT et al., 2013).

Knox et al., (2015) inferiram que a espectroscopia MIR demonstrou prever carbono orgânico e total na maior parte das vezes com maior precisão do que os modelos

derivados da Vis-NIR. Eles encontraram valores de R^2 variando de 0,58 a 0,87 para Vis-NIR; 0,87 a 0,96 para Mir e 0,88 a 0,95 para Vis-NIR-MIR, porém utilizaram uma gama muito maior de amostras que este trabalho, tendo sido amostrados 1014 locais, usando 696 para calibração e 296 para validação.

Vohland et al. (2014), encontrou valores de R^2 para carbono orgânico em Vis-NIR de 0,60 e em Mir de 0,78; com RPD de 1,58 e 2,12; respectivamente. Todas as análises foram realizadas em um conjunto de 60 amostras de solo extraídas de terras aráveis com diferentes tipos de solo, com diferentes texturas, desenvolvidos a partir de diferentes rochas-mãe e em diferentes posições de paisagem. Apesar de poucas amostras, percebe-se a heterogeneidade do material analisado, ao contrário do utilizado neste estudo.

Shepherd & Walsh (2002) acharam valores de R^2 igual a 0,8 para C orgânico, utilizando uma biblioteca espectral com mais de 1 000 amostras, enquanto bashagaluke et al. (2015), predizendo carbono encontraram R^2 igual a 0,72 utilizando 530 amostras de solo compostas. Viscarra Rossel et al. (2006) relataram exatidão similar ($R^2 = 0,73$) para um conjunto de validação de 118 amostras em uma área de 18 ha na Austrália.

Daniel et al. (2003) desenvolvendo estudo na Tailândia usando redes neurais artificiais no espectro Vis-NIR, conseguiu valores de $R^2 = 0,86$ para predição de matéria orgânica, assim como Ben-Dor et al. (2002) que utilizou imagens hiperespectrais para o mapeamento de matéria orgânica com base nos espectros de reflectância da argila, conseguindo R^2 de 0,82.

Arachchi et al., (2016), encontraram, dentre os espectros testados, maior correlação entre o carbono medido e previsto, na faixa do MIR, com R^2 entre 0,63-0,85. Kania e Gruba (2016), testando a predição de carbono total utilizando o espectro NIR, encontrou valores de R^2 para calibração de 0,80 e 0,48 ; e para validação de 0,03 e 0,22, valores este inferiores ao encontrados neste trabalho.

Summers et al. (2011), predizendo carbono orgânico na faixa de 400-2500 nm, com 228 amostras, conseguiram R^2 de 0,57 com RPD de 1,8. Reeves e Smith (2009) predisse carbono orgânico, nos espectros MIR e NIR, encontrando valores de R^2 para carbono orgânico igual a 0,58 e 0,53; respectivamente.

Reeves (2010) observaram que a determinação de orgânico ou matéria orgânica é geralmente viável, mas fatores de confusão como tamanho de partícula, cor do solo e o tipo de solo, entre outros, poderia causar problemas com o desenvolvimento da calibração.

Também foi especificamente observado que o uso do procedimento Walkley-Black (WALKLEY e BLACK, 1934) para carbono orgânico, pode ser problemático devido à não linearidade nos valores medidos, entretanto, esse procedimento é o método mais comumente utilizado (MALLEY et al. 2004). Esses fatores poderiam explicar os valores encontrados, ligeiramente abaixo de alguns estudos citados na literatura com R^2 acima de 0,8 e RPD-RPIQ superiores a 2,0.

As Figuras 58 e 59 mostram a distribuição dos dados dentro de cada preditor nas diferentes faixas do espectro para RPIQ E RPD respectivamente, para a variável carbono orgânico. Em ambos é possível observar que RPIQ, de maneira geral, apresentou maior variabilidade em relação a RPD, no entanto apresentou valores discrepantes apenas em ANN. GPR linear e SVM apresentaram maior oscilação para RPD, de forma mais acentuada em Vis-NIR, enquanto ANN e GPR radial têm menor amplitude dentro de RPIQ.

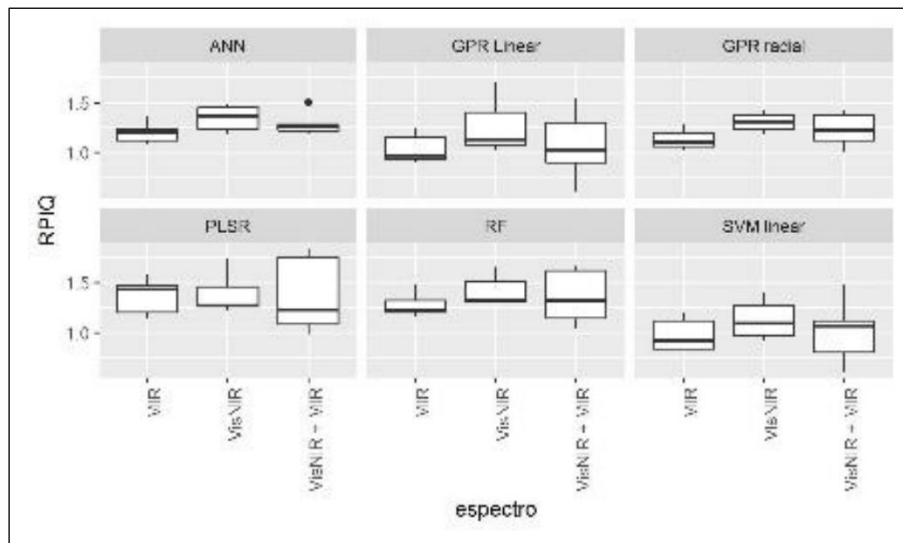


Figura 58- Box plot da distribuição dos dados de RPIQ nos diferentes modelos e faixas do espectro, para variável carbono orgânico.

Partial Least Squares Regression (PLSR), Artificial Neural Network (ANN), Random Forest (RF), Gaussian Process Regression (GPR linear e radial) e Support Vector Machine (SVM). Mir (Infravermelho médio); Vis-NIR (Visível e infravermelho próximo) e Vis-NIR-MIR (Visível, infravermelho próximo e

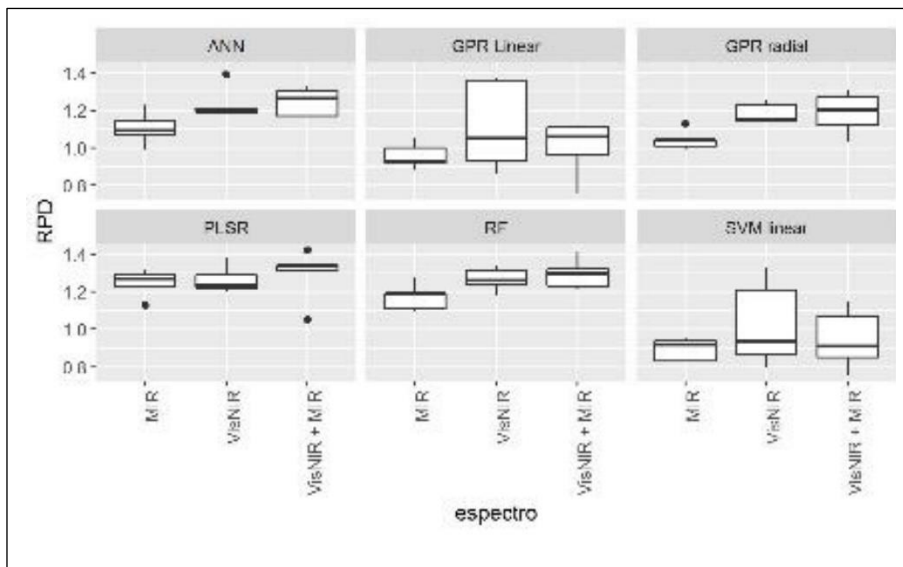


Figura 59- Box plot da distribuição dos dados de RPD nos diferentes modelos e faixas do espectro, para variável carbono orgânico.

Partial Least Squares Regression (PLSR), Artificial Neural Network (ANN), Random Forest (RF), Gaussian Process Regression (GPR linear e radial) e Support Vector Machine (SVM). Mir (Infravermelho médio); Vis-NIR (Visível e infravermelho próximo) e Vis-NIR-MIR (Visível, infravermelho próximo e

A espectroscopia de infravermelhos é bem adaptada e utilizada na predição de propriedades como carbono orgânico do solo (REEVES, 2010; BELLON-MAUREL et al., 2010), que absorve diretamente na região infravermelha, devido à grande sensibilidade dessa região para os grupos como C-H, C-O e C-N que predominam na matéria orgânica (SORIANO-DISLA et al., 2014). BEN-DOR et al. (1999) apontam que

os grupos OH e CH dominam o NIR e as transições eletrônicas das porções visíveis do espectro eletromagnético.

A matéria orgânica inclui a biomassa viva de plantas e remanescentes de vegetação (Bartholomeus et al.; 2008), sendo o carbono orgânico uma indicação do teor de matéria orgânica, já que este é um dos seus principais componentes (STEINER et al. 2011).

De acordo com Beyer et al. (2001, estes contém componentes bioquímicos como clorofila, celulose, pectina, amido, lignina e ácidos húmicos, que influenciam a refletância no visível (400-700 nm) e no infravermelho próximo (700-1400 nm) do espectro eletromagnético. Hartmann e Appel (2006) explicaram a variação nas concentrações de celulose por meio do espectro NIR. Viscarra Rossel E Hicks (2015) relataram o uso de Vis-NIR para prever frações de carbono.

Sendo assim, teoricamente, a matéria orgânica, devido toda sua complexidade, é espectralmente ativa em praticamente toda a região NIR (BEN-DOR e BANIN, 1990), porém muitas vezes é relatado que os sinais da matéria orgânica, nessa região, podem ser fracos (VISCARRA ROSSEL e McBRATNEY,1998), já que pode haver sobreposição de características espectrais de alguns minerais e a matéria orgânica (BEN-DOR e BANIN, 1990; VISCARRA ROSSEL e McBRATNEY,1998). Ben-Dor e Banin (1990) expuseram que as bandas de 1 744 nm, 1 870 nm e 2 052 nm eram importantes para as previsões de carbono orgânico.

Nesse trabalho, a melhor predição foi encontrada no espectro MIR, cuja variáveis mais importantes na predição são apresentadas na Figura 60. As bandas mais relevantes foram 18.193 e 16.879. Ainda se destacaram bandas entre 17.247 e 17.326.

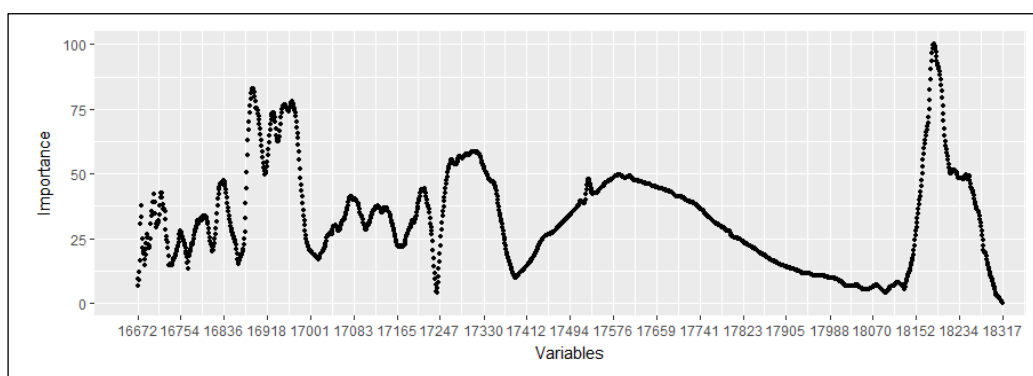


Figura 60- Variáveis importantes para predição de carbono orgânico, no espectro MIR (Infravermelho médio) usando pré-processamento *Continuum removal* (CR) e modelo *Partial Least Squares Regression* (PLSR).

Para a variável fósforo (P), os valores de R^2 (Figura 61) foram os menores dentre os atributos modelados neste trabalho. Abdi et al. (2016) relataram que há influência entre a predição do P disponível no espectro Vis-NIR, e o tipo de solução usada na extração durante a análise do solo em laboratório. Minasny et al. (2009) relatam que o P disponível extraído por qualquer bicarbonato ou o método Bray não são bem previstos. Bons resultados de Vis-NIR foram encontrados quando o P disponível foi extraído utilizando o método de Olsen (VAN GROENIGEN et al., 2003).

Nduwamungu et al. (2009) mostraram que P extraído com Mehlich-3 (M-3) foram mal preditos por Vis-NIR; assim como Abdi et al. (2012), que acreditam que há uma má correlação entre P extraído do M-3 e o carbono orgânico; além do fato que as extrações químicas do solo que alteram o equilíbrio entre as fases podem complicar ainda mais a

interpretação dos resultados (VISCARRA ROSSEL et al., 2006). Porém, historicamente, a compreensão do sistema do solo e avaliação de sua qualidade foi obtida através desse tipo de análise laboratorial (VISCARRA ROSSEL et al., 2006).

Dentro do espectro MIR não foi encontrado nenhum valor satisfatório de R^2 para o conjunto validação, e para o conjunto treinamento obteve-se valor igual a 1,00 (RMSE= 0,13 , Tabela 43) para SGD-GPR linear (Tabela 43), essa combinação proporcionou valores de R^2 semelhantes nos espectros Vis-NIR (RMSE =0,01) e Vis-NIR-MIR (RMSE = 0,01). Maleki et al. (2006) observaram a diferença na reflectância espectral Vis-NIR, de acordo com a variação nos teores de P e a hipótese de que P se correlaciona indiretamente com o infravermelho próximo através de diferentes componentes do solo.

Nos espectros Vis-NIR e Vis-NIR-MIR os valores de R^2 (Figura 61), foram de 0,57 (SNV-PLSR com RMSE = 3,09) e de 0,35 (MSC-PLSR com RMSE= 4,26), respectivamente.

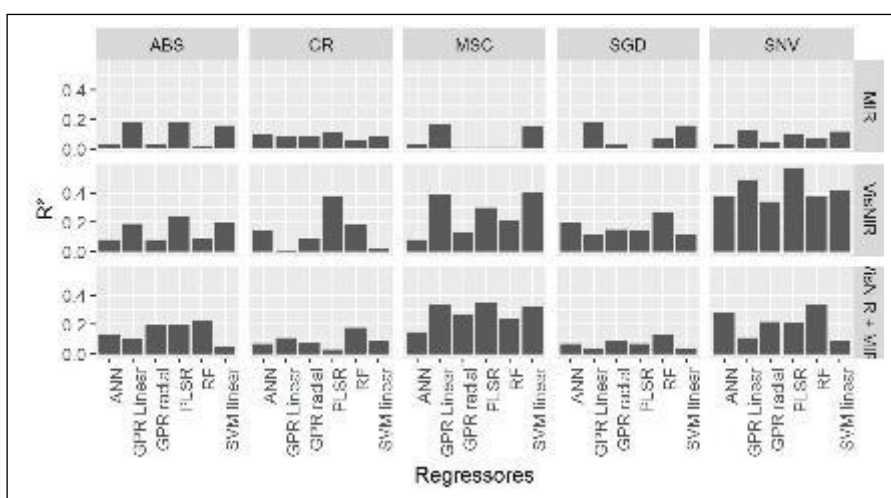


Figura 61- Valores do coeficiente de determinação (R^2), encontrados para variável fósforo, nos diferentes métodos e pré-processamentos no grupo na validação.

Absorbance (ABS); Continuum removal (CR); Multiplicative Scatter Correction (MSC); Savitzky Golay Derivative (SGD); Standard Normal Variate (SNV) Partial Least Squares Regression (PLSR), Artificial Neural Network (ANN), Random Forest (RF), Gaussian Process Regression (GPR linear e radial) e Support Vector Machine (SVM). Mir (Infravermelho médio); Vis-NIR (Visível e infravermelho próximo) e Vis-NIR-MIR (Visível, infravermelho próximo e médio).

Tabela 43- Valores do coeficiente de determinação (R^2), para variável fósforo, encontrados para os diferentes métodos e pré-processamentos no grupo no treinamento, nos espectros Mir, Vis-NIR e Vis-NIR-MIR

R^2					R^2				
		Vis-Mir	Vis-NIR	Vis-NIR-MIR			Vis-Mir	Vis-NIR	Vis-NIR-MIR
Modelo	Pré-processamento	Treinamento			Modelo	Pré-processamento	Treinamento		
PLSR	CR	0,14	0,18	0,18	GPR Linear	SGD	1,00	1,00	1,00
ANN	CR	0,12	0,16	0,16	GPR radial	SGD	0,16	0,36	0,36
RF	CR	0,10	0,16	0,16	SVM linear	SGD	0,99	0,99	0,99
GPR Linear	CR	0,98	0,99	0,99	PLSR	SNV	0,53	0,87	0,87
GPR radial	CR	0,14	0,22	0,22	ANN	SNV	0,22	0,25	0,25
SVM linear	CR	0,98	0,99	0,99	RF	SNV	0,08	0,15	0,15
PLSR	ABS	0,42	0,30	0,30	GPR Linear	SNV	0,98	0,99	0,99
ANN	ABS	0,10	0,11	0,11	GPR radial	SNV	0,18	0,25	0,25
RF	ABS	0,00	0,08	0,08	SVM linear	SNV	0,99	0,99	0,99
GPR Linear	ABS	0,93	0,96	0,96	PLSR	MSC	0,23	0,68	0,68
GPR radial	ABS	0,98	0,20	0,20	ANN	MSC	0,17	0,28	0,28
SVM linear	ABS	0,95	0,98	0,98	RF	MSC	0,13	0,16	0,16
PLSR	SGD	0,04	0,49	0,49	GPR Linear	MSC	0,97	0,97	0,97
ANN	SGD	0,00	0,22	0,22	GPR radial	MSC	0,96	0,24	0,24
RF	SGD	0,00	0,27	0,27	SVM linear	MSC	0,99	0,99	0,99

Absorbance (ABS); Continuum removal (CR); Multiplicative Scatter Correction (MSC); Savitzky Golay Derivative (SGD); Standard Normal Variate (SNV) Partial Least Squares Regression (PLSR), Artificial Neural Network (ANN), Random Forest (RF), Gaussian Process Regression (GPR linear e radial) e Support Vector Machine (SVM). Mir (Infravermelho médio); Vis-NIR (Visível e infravermelho próximo) e Vis-NIR-MIR (Visível, infravermelho próximo e médio).

Tabela 44- Valores de RMSE, para variável fósforo, encontrados para os diferentes métodos e pré-processamentos no grupo no treinamento (T) e validação (V), nos espectros Mir, Vis-NIR e Vis-NIR-MIR

		RMSE					
Modelo	Pré-processamento	Mir		Vis-NIR		Vis-NIR-MIR	
		T	V	T	V	T	V
PLSR	CR	4,46	4,27	4,30	4,16	4,30	4,61
ANN	CR	4,49	4,28	4,48	4,12	4,48	4,10
RF	CR	4,54	4,41	4,35	3,90	4,35	4,19
GPR Linear	CR	0,80	5,52	0,56	14,93	0,56	5,88
GPR radial	CR	4,57	4,34	4,34	4,05	4,34	4,46
SVM linear	CR	0,54	6,09	0,46	14,03	0,46	6,51
PLSR	ABS	3,57	4,30	4,04	3,70	4,04	4,04
ANN	ABS	4,46	4,67	4,39	4,57	4,39	4,57
RF	ABS	4,90	4,73	4,64	3,82	4,64	3,92
GPR Linear	ABS	1,37	4,88	1,10	3,89	1,10	6,27
GPR radial	ABS	2,43	4,67	4,49	3,79	4,49	4,09
SVM linear	ABS	1,11	5,42	0,72	3,94	0,72	8,13
PLSR	SGD	4,50	5,06	3,28	4,07	3,28	5,22
ANN	SGD	4,60	4,97	4,19	3,84	4,19	4,72
RF	SGD	4,63	4,84	3,97	3,73	3,97	4,63
GPR Linear	SGD	0,13	4,85	0,01	5,21	0,01	6,03
GPR radial	SGD	4,45	4,92	3,96	3,94	3,96	4,73
SVM linear	SGD	0,46	4,82	0,45	5,10	0,45	5,90
PLSR	SNV	3,32	4,69	1,58	3,09	1,58	4,90
ANN	SNV	4,31	4,40	3,88	3,80	3,88	4,50
RF	SNV	4,69	4,27	4,05	3,83	4,05	4,68
GPR Linear	SNV	0,71	5,37	0,54	3,34	0,54	5,52
GPR radial	SNV	4,68	4,20	3,92	4,08	3,92	4,99
SVM linear	SNV	0,49	5,70	0,43	3,62	0,43	5,89
PLSR	MSC	4,15	4,76	2,74	3,94	2,74	4,26
ANN	MSC	4,33	4,57	4,07	4,23	4,07	4,20
RF	MSC	4,44	4,74	4,47	3,83	4,47	3,77
GPR Linear	MSC	0,84	5,42	0,88	3,60	0,88	5,24
GPR radial	MSC	2,36	4,60	4,36	4,01	4,36	3,77
SVM linear	MSC	0,47	5,97	0,48	3,45	0,48	5,75

Absorbance (ABS); Continuum removal (CR); Multiplicative Scatter Correction (MSC); Savitzky Golay Derivative (SGD); Standard Normal Variate (SNV) Partial Least Squares Regression (PLSR), Artificial Neural Network (ANN), Random Forest (RF), Gaussian Process Regression (GPR linear e radial) e Support Vector Machine (SVM). Mir (Infravermelho médio); Vis-NIR (Visível e infravermelho próximo) e Vis-NIR-MIR (Visível, infravermelho próximo e médio).

Para RPIQ (Figura 62), o maior valor foi encontrado em Vis-NIR-MIR, sendo igual a 2,04 (SNV-ANN), seguido por Vis-NIR igual a 1,92 (SNV-PLSR) e Mir igual a 1,64 (SGD-SVM linear). Já para RPD (Figura 63), a faixa do Vis-NIR apresentou-se com valor igual a 1,53 (SNV-PLSR), seguido por Vis-NIR-MIR com índice de 1,25 (MSC-RF e MSC-GPR radial), e Mir com valor de 1,11 (SNV-RF). Assim, o espectro Vis-NIR-MIR se destacou pelo índice RPIQ sendo classificado como excelente, enquanto os outros foram tidos como razoáveis, e pelo índice RPD, Vis-NIR foi qualificado com um modelo razoável.

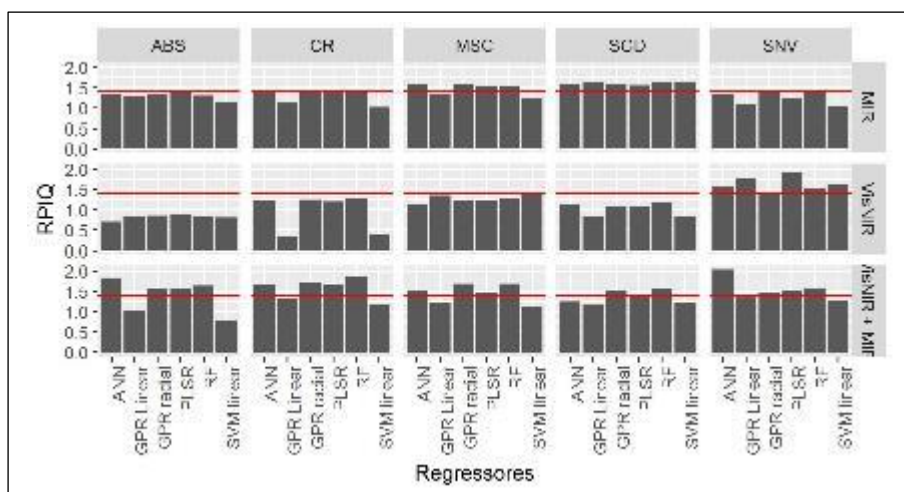


Figura 62- Valores de RPIQ encontrados para os diferentes métodos e pré-processamentos no grupo na validação, para variável fósforo.

Absorbance (ABS); Continuum removal (CR); Multiplicative Scatter Correction (MSC); Savitzky Golay Derivative (SGD); Standard Normal Variate (SNV) Partial Least Squares Regression (PLSR), Artificial Neural Network (ANN), Random Forest (RF), Gaussian Process Regression (GPR linear e radial) e Support Vector Machine (SVM). Mir (Infravermelho médio); Vis-NIR (Visível e infravermelho próximo) e Vis-NIR-MIR (Visível, infravermelho próximo e médio).

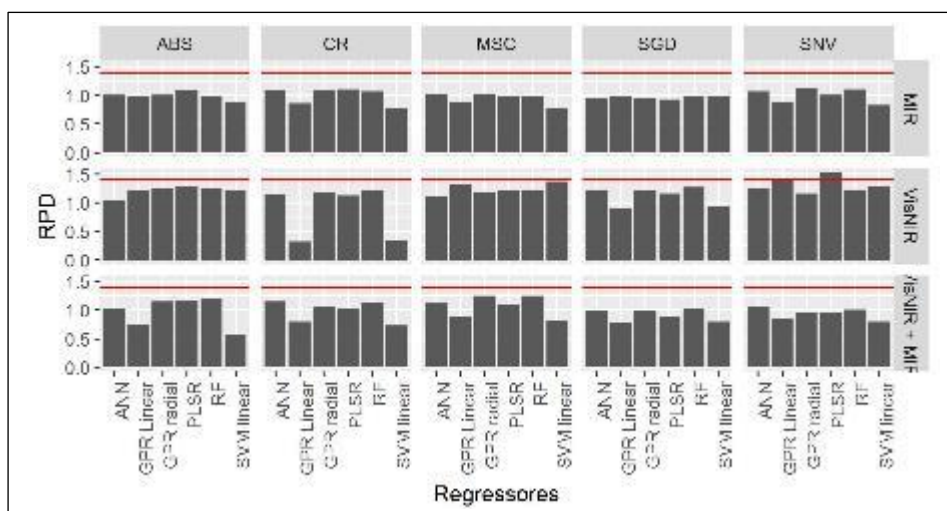


Figura 63- Valores de RPD encontrados para os diferentes métodos e pré-processamentos no grupo na validação, para variável fósforo.

Absorbance (ABS); Continuum removal (CR); Multiplicative Scatter Correction (MSC); Savitzky Golay Derivative (SGD); Standard Normal Variate (SNV) Partial Least Squares Regression (PLSR), Artificial Neural Network (ANN), Random Forest (RF), Gaussian Process Regression (GPR linear e radial) e Support Vector Machine (SVM). Mir (Infravermelho médio); Vis-NIR (Visível e infravermelho próximo) e Vis-NIR-MIR (Visível, infravermelho próximo e médio).

Quanto aos valores de RPIQ e RPD para a variável fósforo, PLSR foi superior em Vis-NIR e em Mir juntamente com GPR radial (para o índice RPIQ); em Vis-NIR-MIR se destacaram RF e ANN. Em referência aos pré-processamentos para fósforo, SGD se destacou com maior parte dos valores de RPIQ superior a 1,4 na faixa do MIR e SNV nos demais, tendo SNV apresentado ainda, valor de RPIQ superior a 2,0. Quanto ao RPD apenas a faixa Vis-NIR apresentou valores acima de 1,4 com SNV se destacando como os melhores pré-processamentos para a variável.

Segundo Malley et al. (2004) boas previsões no espectro NIR são menos frequentes para predição de P, sendo encontrados valores de R^2 entre 0,4 e 0,5 (Malley et al, 2002, para P e N mineral.; Chang et al., 2001, e Shepherd & Walsh, 2002). As possíveis causas são a natureza do elemento estudado, a interação com outros elementos, além do método de extração utilizado (Chang et al., 2001). Podem ainda estar relacionado a heterogeneidade do conjunto de amostras e a calibração.

Por outro lado, Niederberger et al., 2015, conseguiram bons resultados trabalhando no espectro NIR, onde perceberam que modelos preditos com base nas frações orgânicas do P tinham resultados superiores ao P inorgânico, isso porque os compostos orgânicos eram mais facilmente excitados no espectro NIR.

Abdi et al. (2012) encontraram R^2 entre 0,7 e 0,8 predizendo P total, utilizando o espectro Vis-NIR para mapeamento dos teores do elemento em um solo extremamente arenoso no Canadá, atribuindo esses valores a correlação existente entre o mesmo e a matéria orgânica. Viscarra Rossel et al. (2006), encontrou modelando P disponível, valores de R^2 iguais a 0,06 (Vis); 0,01 (Nir); 0,20 (Mir) e 0,02 (Vis-NIR-MIR). Janik et al. (1998) modelando teores de P disponível com espectro MIR encontrou valores de R^2 igual a 0,07 enquanto Daniel et al. (2003), com Vis-NIR encontraram valores de R^2 de 0,81.

Lee et al. (2003) fizeram a predição de P em 540 amostras de solo de quatro grandes tipos de solo na Flórida na faixa de comprimento de onda de 400 a 2500 nm, trabalharam com PLSR e os valores de R^2 variaram de 0,52 a 0,66. Knox et al. (2015) encontraram valores de R^2 variando entre 0,51 e 0,95 para diferentes faixas do espectro. Minasny et al. (2009), não encontram valores satisfatórios na predição de P disponível no espectro MIR, McCarty e Reeves (2006) também não considerou satisfatório a predição de P utilizando esse comprimento de onda.

Viscarra Rossel et al. (2006) usando PLSR, observaram que a precisão das previsões em cada uma das regiões espectrais VIS, NIR, MIR e Vis-NIR-MIR variou consideravelmente entre as diferentes propriedades. Obtiveram-se previsões mais precisas utilizando o MIR para carbono orgânico, argila e P.

Reeves, et al. (2009) concluíram de NIR poderia ser satisfatório para predizer algumas propriedades, porém não obteve resultados adequados, ou com grande variabilidade de estudo para estudo, para formas minerais de Ag, Al, Cd, Cu, Co, Fe, K, N, P, Pb, Na, Ni, Se, Si, Zn e pH. Esses autores encontraram valores de R^2 para P igual a 0,85 e 0,09, nos espectros MIR e NIR, respectivamente.

Bogrekci e Lee (2007) encontraram coeficientes de determinação entre 0,76 e 0,93 na região do infravermelho próximo, usando PLSR para P total, P extraído por Mehlich-1 e P solúvel em água. E na região do visível, valores entre 0,61 e 0,83.

As Figuras 64 e 65 mostram a distribuição dos dados dentro de cada preditor nas diferentes faixas do espectro para RPIQ E RPD respectivamente, para a variável fósforo. Em ambos é possível visualizar a distribuição satisfatória dos dados, com baixa amplitude e variabilidade. RPIQ teve distribuição ligeiramente menos uniforme em relação a RPD. GPR linear e SVM apresentaram maior oscilação, em ambos índices e de maneira mais acentuada no espectro Vis-NIR.

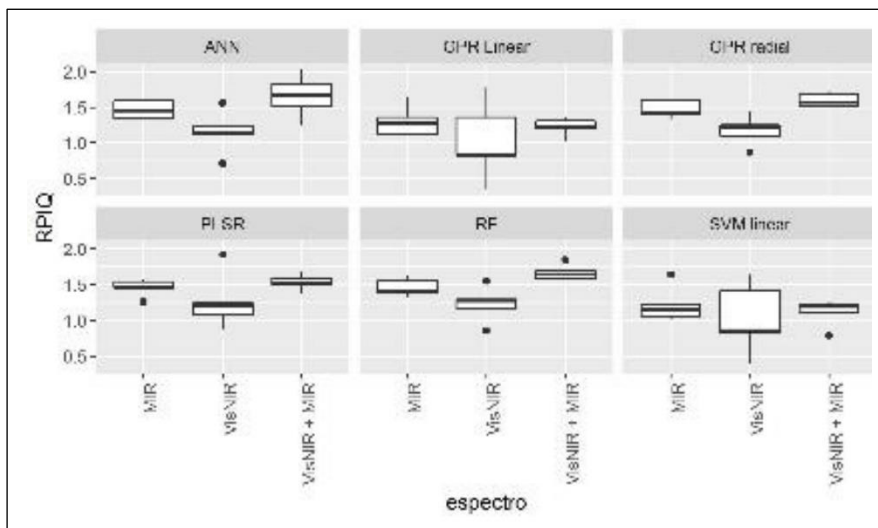


Figura 64- Box plot da distribuição dos dados de RPIQ nos diferentes modelos e faixas do espectro, para variável fósforo.

Partial Least Squares Regression (PLSR), Artificial Neural Network (ANN), Random Forest (RF), Gaussian Process Regression (GPR linear e radial) e Support Vector Machine (SVM). Mir (Infravermelho médio); Vis-NIR (Visível e infravermelho próximo) e Vis-NIR-MIR (Visível, infravermelho próximo e

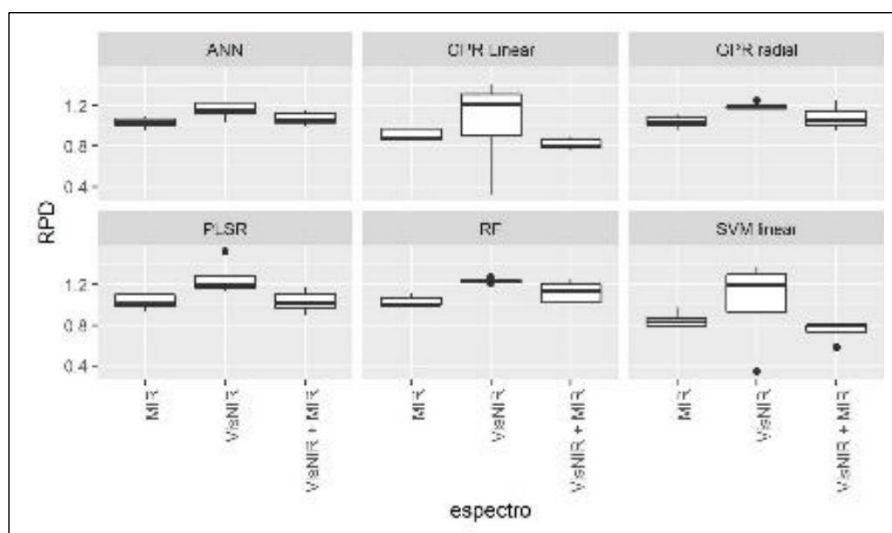


Figura 65- Box plot da distribuição dos dados de RPD nos diferentes modelos e faixas do espectro, para variável fósforo.

Partial Least Squares Regression (PLSR), Artificial Neural Network (ANN), Random Forest (RF), Gaussian Process Regression (GPR linear e radial) e Support Vector Machine (SVM). Mir (Infravermelho médio); Vis-NIR (Visível e infravermelho próximo) e Vis-NIR-MIR (Visível, infravermelho próximo e

Quanto as variáveis mais importantes para predição de P (Figura 66), constatou-se que as mais importantes foram as faixas em torno de 620 nm, 2220 nm e 2495 nm. Tendo ainda destaque as faixas 1479 a 1580 nm, 1853 nm e 1907nm.

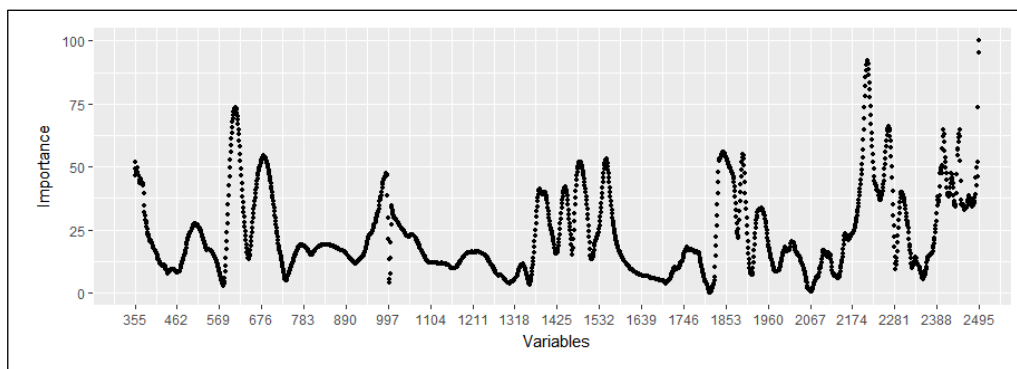


Figura 66- Variáveis importantes para predição de fósforo, no espectro Vis-NIR (Visível e Infravermelho próximo) usando pré-processamento *Satandard Normal Variate (SNV)* e modelo *Partial Least Squares Regression (PLSR)*.

Considerando o afirmado por Bellon-Maurel e McBratney (2011), onde seria preferível a avaliação do índice RPIQ sobre o RPD, este estudo encontrou modelos satisfatórios (RPIQ e/ou RPD superior a 1,4) e excelentes (RPIQ superior a 2,0) capazes de prever P, CO e argila, podendo estes serem utilizados para classificar solos de acordo com suas propriedades, conforme sugerido por Shepherd & Walsh (2002).

Segundo Ludwig et al. (2016) muitos estudos visam principalmente, alcançar os maiores valores de RPD possíveis, desconsiderando outros aspectos importantes como a representatividade da amostra para uma população específica, um tratamento apropriado de casos suspeitos e uma discussão sobre a generalização dos resultados do MIR dependendo da escala.

Shepherd & Walsh (2002) foram os primeiros a propor o uso da análise NIR laboratorial para a discriminação de solos a partir de valores especificados de corte (acima ou abaixo de um critério estabelecido) para a maioria das propriedades relacionadas à fertilidade do solo. Eles mostraram que as amostras poderiam ser discriminadas de forma aproximada usando árvores de classificação mesmo para propriedades como K trocável e P que são mal previstos pelos modelos de regressão.

De acordo com Cecillon et al (2009) a qualidade do solo nem sempre precisa ser quantificado com precisão, já que muitas aplicações industriais ou agrícolas requerem apenas uma classificação da condição do solo em relação a um valor de teste crítico para uma propriedade chave.

O solo é um sistema heterogêneo cujos processos e mecanismos são complexos e difíceis de compreender plenamente (Viscarra Rossel et al., 2006). Muitas técnicas analíticas de solo convencionais são usadas na tentativa de estabelecer a relação entre propriedades físicas e químicas do solo e componentes individuais do solo, muitas vezes desconsiderando as complexas interações de múltiplos componentes (Viscarra Rossel et al., 2006).

Isso leva mais uma vez a enxergar a importância do uso de novas técnicas analíticas, dessa forma, a espectroscopia visível e infravermelha pode, em instâncias, ser mais direta do que a análise convencional do solo e, às vezes, mais precisa (Viscarra Rossel et al., 2006). McCauley et al., já em 1993, sugeriram que a espectroscopia VIS pode ser mais precisa do que as digestões de dicromato para análise de carbono orgânico do solo.

5 CONCLUSÕES

1. Este estudo encontrou modelos satisfatórios (RPIQ e/ou RPD superior a 1,4) e excelentes (RPIQ superior a 2,0) capazes de prever argila, CO e P.

2. Os modelos encontrados podem ser testados para discriminar solos segundo um valor de teste crítico para argila, CO e P. Como, por exemplo, separar solos arenosos e argilosos.

3. Acredita-se que a adição de mais amostras no treinamento e validação, pode melhorar os modelos com RPIQ e RPD entre 1,4 e 2,0 tornando-os modelos acurados, com excelente predição.

4. Melhor faixa do espectro para predição de argila, CO e P, foram Vis-NIR (argila e P) e Mir (carbono orgânico);

5. Os pré-processamentos MSC, CR e SNV foram os mais eficientes para predição de argila, CO e P, respectivamente, enquanto os métodos PLSR (CO e P) e SVM (argila) se destacaram com as melhores predições, sendo por isso, recomendados para modelagem desses atributos na área estudada.

6 REFERÊNCIAS BIBLIOGRÁFICAS

ABDI, D., TREMBLAY, G.F., ZIADI, N., BÉLANGER, G., PARENT, L.-É. Predicting soil phosphorus-related properties using near-infrared reflectance spectroscopy. *Soil Sci. Soc. Am. J.* 76, 2318–2326.2012.

ABDI, D.; CADE-MENUN, B.J.; ZIADI, N.; TREMBLAY, G.F.; PARENT, L.-É. Visible near infrared reflectance spectroscopy to predict soil phosphorus pools in chernozems of Saskatchewan, Canada. *Geoderma Regional*, [s.l.], v. 7, n. 2, p.93-101, jun. 2016.

ADELIN, K.R.M.; GOMEZ, C.; GORRETTA, N.; ROGER, J.M. Predictive ability of soil properties to spectral degradation from laboratory Vis-NIR spectroscopy data. *Geoderma*, [s.l.], v. 288, p.143-153, fev. 2017.

ALBA, J.M.F.; ALCOVER NETO, A.; SANTOS, H.P DOS; SANDRINI, W.C. Caracterização espectrorradiométrica de minerais e rochas sedimentares. *Embrapa Clima Temperado: Série Documentos: 172*. 13 p. 2006.

ANDRONIKOV, V.L.; DOBROLVSHIY, G.V. Theory and methods for the use of remote sensing in the study of soils. *Mapping Science and Remote Sensing*, v.28, p.92-101, 1991.

ARACHCHI, M.P.N.K. H; FIELD, D.J.; McBRATNEY, A.B.. Quantification of soil carbon from bulk soil samples to predict the aggregate-carbon fractions within using near- and mid-infrared spectroscopic techniques. *Geoderma*, [s.l.], v. 267, p.207-214, abr. 2016.

BARTHOLOMEUS, H.M.; SCHAEPMAN, M.E.; KOOISTRA, L.; STEVENS, A.; HOOGMOED, W.B.; SPAARGAREN, O.S.P. Spectral reflectance based indices for soil organic carbon quantification. *Geoderma*, [s.l.], v. 145, n. 1-2, p.28-36, maio 2008.

BASHAGALUKE JANVIER; NSHOBOLE NICOLE; FATAKI DIDIER, MOCHOGE BENSON, MUGWE JAYNE, WALANGULULU JEAN. Application of infrared technique in soil properties characterization in South Kivu province of DR Congo. *African Journal of Food Science and Technology*. Vol. 6(2) pp. 58-67, Feb 2015.

BELLON-MAUREL, V., FERNANDEZ-AHUMADA, E., PALAGOS, B., ROGER, J.-M.; McBRATNEY, A. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *TrAC Trends Anal. Chem.* 29, 1073–1081. 2010.

BELLON-MAUREL, V., FERNANDEZ-AHUMADA, E., PALAGOS, P., ROGER, J.M., McBRATNEY, A.B. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *Trends Anal. Chem.* 29, 1073–1081.2010.

BELLON-MAUREL, V., McBRATNEY, A. Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils — critical review and research perspectives. *Soil Biol. Biochem.* 43, 1398–1410.2011.

BELLON-MAUREL, V.; FERNANDEZ-AHUMADA, E.; PALAGOS, B.; ROGER, J.-M.; McBRATNEY, A. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *Trends in Analytical Chemistry*. Amsterdam, v.29, n 9, p. 1073-1081,2010.

BEN-DOR, E., BANIN, A. Diffuse reflectance spectra of smectite minerals in the near infrared and their relation to chemical composition, *Sci. Geol. Bull.* 43.117–128.1990.

BEN-DOR, E., BANIN, A. Near infrared analysis as a rapid method to simultaneously evaluate several soil properties. *Soil Science Society of America Journal* 59, 364–372. 1995.

BEN-DOR, E., IRONS, J.R., EPEMA, J.F. Soil reflectance: remote sensing for the Earth Science, In: Rencz, Andrew N. (Ed.), 3 ed. *Manual of Remote Sensing*, vol.3.1999.

BEN-DOR, E., PATKIN, K., BANIN, A., KARNIELI, A., 2002. Mapping of several soil properties using DAIS- 7915 hyperspectral scanner data – a case study over clayey soils in Israel. *Int. J. Remote Sensing* 23, 1043–1062.2002.

BEYER, L., KAHLE, P., KRETSCHMER, H., WU, Q. Soil organic matter composition of man-impacted urban sites in North Germany. *Journal of Plant Nutrition and Soil Science* 164, 359–364.2001.

BILGILI; A.V., VAN ES H.M.;AKBAS, F., DURAK; A., HIVELY, W.D. Visible-near infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey. *Journal of Arid Environments*. v.74. p. 229–238. 2010.

- BOGREKCI, I.; LEE, W.s.. Comparison of Ultraviolet, Visible, and Near Infrared Sensing for Soil Phosphorus. *Biosystems Engineering*, [s.l.], v. 96, n. 2, p.293-299, fev. 2007.
- BORNEMANN, L., WELP, G., BRODOWSKI, S., RODIONOV, A., AMELUNG, W. Rapid assessment of black carbon in soil organic matter using mid-infrared spectroscopy. *Org. Geochem.* 39, 1537–1544. 2008.
- BROWN, D.J., BRICKLEMYER, R.S., MILLER, P.R. Validation requirements for diffuse reflectance soil characterization models with a case study of VNIR soil C prediction in Montana. *Geoderma* 129, 251–267.2005.
- BROWN, D.J., SHEPHERD, K.D., WALSH, M.G., DEWAYNE MAYS, M., REINSCH, T.G. Global soil characterization with VNIR diffuse reflectance spectroscopy. *Geoderma* 132, 273–290.2006.
- CÉCILLON, L.; BARTHÉS, B.; GOMEZ, C.; ERTLEN, D.; GÉNOT, V.; HEDDE, M.; STEVENS, A.; BRUN, J-J. Assessment and monitoring of soil quality using near-infrared reflectance spectroscopy (NIRS). *European Journal Of Soil Science*, [s.l.], v. 60, n. 5, p.770-784, out. 2009.
- CEPSRM-CENTRO ESTADUAL DE PESQUISAS EM SENSORIAMENTO REMOTO E METEOROLOGIA. UFRGS-Rio Grande do Sul-RS. 2017. Disponível em:< <http://www.ufrgs.br/engcart/PDASR/rem.html>>. Acesso em: Mar, 2017.
- CHANG, C.W, LAIRD, D.A., MAUSBACH, M.J., HURBURGH, C.R.J. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. *Soil Sci. Soc. Am. J.* 65, 480–490.2001.
- CHANG, C.W.; LAIRD, D.A. Near-infrared reflectance spectroscopic analysis of soil C and N. *Soil Science*, Philadelphia, v. 167, n.2, p.110-116, 2002.
- CHANG, C.W.; LAIRD, D.A; MAUSBACH, M.J.; HUBURGH,C.R. Near-infrared reflectance spectroscopy-principal componentes regression analyses od soil properties. *Soil Science Society of America Journal*, Madison, v.65, p.480-900, 2001.
- COURROL, L.C; PRETO, A. DE O. Apostila teórica: Óptica Técnica I. Faculdade de Tecnologia de São Paulo-FATEC. 2017. Disponível em: < http://www.fatecsp.br/paginas/apostila_teorica.pdf> Acesso em: Mar, 2017.
- DALMOLIN, R.S.D. Matéria orgânica e características físicas, químicas, mineralógicas e espectrais de Latossolos de diferentes ambientes. 2002. 151f. Tese (Doutorado em Ciência do Solo) - Programa de Pós-graduação em Ciência do Solo, Universidade Federal do Rio Grande do Sul.
- DALMOLIN, R.S.D.; GONÇALVES, C.N.; KLAMT, E.; DICK,D.P. Relação entre os constituintes do solo e seu comportamento espectral. *Ciência Rural*, Santa Maria, v.35, n.2, p.481-489, mar-abr, 2005.
- DANIEL, K.W., TRIPATHI, N.K., HONDA, K. Artificial neural network analysis of laboratory and in situ spectra for the estimation of macronutrients in soils of Lop Buri (Thailand). *Australian Journal of Soil Research* 41, 47–59. 2003.

DARDANNE, P., SINNAEVE, G., BAETEN, V. Multivariate calibration and chemometrics for near infrared spectroscopy: which method? *Journal of Near Infrared Spectroscopy* 8, 229–237. 2000.

DEMATTÊ, J.A.M.; GENU, A. M.; FIORIO, P.R.; ORTIZ, J.L.; MAZZA, J.A.; LEONARDO, H.C.L. Comparação entre mapas de solos obtidos por sensoriamento remoto espectral e pelo método convencional. *Pesquisa agropecuária brasileira*, Brasília, v.39, n.12, p.1219-1229, dez. 2004.

EMBRAPA. Sistema Brasileiro de Classificação de Solos. Embrapa Solos. Centro Nacional de Pesquisa de Solos (Rio de Janeiro, RJ). Sistema brasileiro de classificação de solos. 2. ed. – Rio de Janeiro: EMBRAPA-SPI, 2006. 306 p.

ESALQ. Escola Superior de Agricultura Luiz de Queiróz. 2017. Disponível em: <<http://bibliotecaespectral.wixsite.com/esalq/about>>. Acesso em: Maio, 2017.

FERNANDEZ-PIERNA, J.A., DARDENNE, P. Soil Parameter Quantification by NIRS as a Chemometric Challenge at ‘Chimiometrie 2006’, 9th Annual Chemometrics Congress. Elsevier Science Bv, Paris, France 94–98. 2006

FIDENCIO, P.H., POPPI, R.J., DE ANDRADE, J.C. Determination of organic matter in soils using radial basis function networks and near infrared spectroscopy. *Analytica Chimica Acta* 453, 125–134. 2002.

Figueiredo, D. Conceitos básicos de sensoriamento remoto. CONAB. Brasília - DF 2005. Disponível em <http://www.conab.gov.br/conabweb/download/SIGABRASIL/manuais/conceitos_sm.pdf> Acesso em: Mar, 2017.

FLORENZANO, T.G. Imagens de satélites para estudos ambientais. Cap 1: Imagens obtidas por sensoriamento remoto. São Paulo: Oficina de Textos, 97 p., 2002.

GENOT, V., COLINET, G., BOCK, L., VANVYVE, D., REUSEN, Y., DARDENNE, P. Near infrared reflectance spectroscopy for estimating soil characteristics valuable in the diagnosis of soil fertility. *Journal of Near Infrared Spectroscopy* 19 (2), 117–138. 2011.

GERARDI, M. Física. Universidade Federal de Juiz de Fora-UFJF. 90 p. 2014.

GOBRECHT, A., ROGER, J.M., BELLON-MAUREL, V. Major issues of diffuse reflectance NIR spectroscopy in the specific context of soil carbon content estimation: a review. *Adv. Agron.* 123, 145–175. 2013.

GROENEVELD, R.A.; MEEDEN, G. Measuring skewness and kurtosis. *Journal of the Royal Statistical Society*. London, v.33, n.4, p.391-399, 1984.

HARTMANN, H.P., APPEL, T. Calibration of near infrared spectra for measuring decomposing cellulose and green manure in soils. *Soil Biology & Biochemistry* 38, 887–897. 2006.

HARTMANN, H.P., APPEL, T. Calibration of near infrared spectra for measuring decomposing cellulose and green manure in soils. *Soil Biology & Biochemistry* 38, 887–897. 2006.

HATCHELL, D.C. Analytical spectral devices, Inc.technical guide. Boulder: Analytical spectral devices, 1999. 140 p.

HUNT, G.R., 1980. Electromagnetic radiation – the communication link in remote sensing. In: Siegal, B.S., Gillespie, A.R. (Eds.), Remote Sensing in Geology. Wiley, New York.

ISLAM, K., SINGH, B., McBRATNEY, A.B. Simultaneous estimation of various soil properties by ultra-violet, visible and near-infrared reflectance spectroscopy. Australian Journal of Soil Research 41, 1101– 1114.2003.

JANIK, L.J., MERRY, R.H., SKJEMSTAD, J.O. Can mid-infrared diffuse reflectance analysis replace soil extractions? Aust. J. Exp. Agric. 38, 681–696.1998.

JIA, S.; LI, H.; WANG, Y.; TONG, R., LI, Q. Recursive variable selection to update near-infrared spectroscopy model for the determination of soil nitrogen and organic carbon. Geoderma, [s.l.], v. 268, p.92-99, abr. 2016.

KANIA, MATEUSZ; GRUBA, PIOTR. Estimation of selected properties of forest soils using near-infrared spectroscopy (NIR). Soil Science Annual, [s.l.], v. 67, n. 1, p.32-36, 1 jan. 2016.

KEMPER, T., BÖTTCHER, K., MACHWITZ, M., SOMMER, S., MEHL, W. Na approach to chemometric methods for soil organic matter estimation from laboratory and remote sensing data. 4th Workshop on Imaging Spectroscopy, Warsaw, pp. April 27–29.2005.

KNOX, N.M.S., GRUNWALD, S., MCDOWELL, M.L., BRULAND, G.J., MYERS, D.B., HARRIS, W.G. Modelling soil carbon fractions with visible near-infrared (VNIR) and mid-infrared (MIR) spectroscopy. Geoderma 239–240, 229–239.2015.

KOOISTRA, L., WEHRENS, R., LEUVEN, R.S.E.W., BUYDENS, L.M.C. Possibilities of visible-near-infrared spectroscopy for assessment of soil contamination in river floodplains. Analytica Chimica Acta 446, 97–105.2001.

LEE W S; SANCHEZ J F; MYLAVARAPU R S; CHOE J S. Estimating chemical properties of Florida soils using spectral reflectance. Transactions of the ASAE, 46(5), 1443–1453.2003.

LUDWIG, B., NITSCHKE, R., TERHOEVEN-URSELMANS, T., MICHEL, K., FLESSA, H. Use of midinfrared spectroscopy in the diffuse-reflectance mode for the prediction of the composition of organic matter in soil and litter. J. Plant Nutr. Soil Sci. 171, 384–391.2008.

LUDWIG,B.; LINSLER,D.; HÖPER, H.; SCHMIDT, H.; PIEPHO, H-P.; VOHLAND,M. Pitfalls in the use of middle-infrared spectroscopy: representativeness and ranking criteria for the estimation of soil properties. Geoderma, [s.l.], v. 268, p.165-175, abr. 2016.

MADARI, B.E., REEVES III, J.B., COELHO, M.R., MACHADO, P.L.O.A., DE POLLI, H., COELHO, R.J., BENITES, V., SOUZA,M., MCCARTY, L.F., G.W. Mid- and near-

infrared spectroscopic determination of total and organic carbon in a diverse set of soils from the Brazilian national soil collection. *Spectrosc. Lett.* 38, 721–740.2005.

MADARI, B.E., REEVES III, J.B., MACHADO, P.L.O.A., GUIMARÃES, C.M., TORRES, E., McCARTY, G.W. Mid- and near-infrared spectroscopic assessment of soil compositional parameters and structural indices in two Ferralsols. *Geoderma* 136, 245–259.2006.

MALEKI, M.R., VAN HOLM, L., RAMON, H., MERCKX, R., DE BAERDEMAEKER, J., MOUAZEN, A.M. Phosphorus sensing for fresh soils using visible and near infrared spectroscopy. *Biosystems Engineering* 95, 425–436.2006.

MALLEY, D.F., MARTIN, P.D., DEN-DOR, E. Application in analysis of soils. In: Roberts, C.A., Workman Jr., J., Reeves III, J.B. (Eds.), *Near-Infrared Spectroscopy in Agriculture*. American Society of Agronomy, Madison, WI, pp. 729–784. Chapter 26.2004.

MALLEY, D.F., YESMIN, L. & EILERS, R.G. Rapid analysis of hog manure and manure-amended soils using near-infrared spectroscopy. *Soil Science Society of America Journal*, 66, 1677- 1686.2002.

McCARTY, G.W., REEVES III, J.B. Comparison of near infrared and mid infrared diffuse reflectance spectroscopy for field-scale measurement of soil fertility parameters. *Soil Sci.* 171, 94–102.2006.

McCARTY, G.W., REEVES III, J.B., REEVES, V.B., FOLLETT, R.F., KIMBLE, J.M. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. *Soil Sci. Soc. Am. J.* 66, 640–646. 2002.

McCARTY, G.W., REEVES III, J.B., REEVES, V.B., FOLLETT, R.F., KIMBLE, J.M. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. *Soil Sci. Soc. Am. J.* 66, 640–646.2002.

McCAULEY, J.D., ENGEL, B.A., SCUDDER, C.E., MORGAN, M.T., ELLIOT, P.W. Assessing the spatial variability of organic matter. ASAE Paper No. 93–1555, American Society of Agricultural Engineers, St. Joseph, MI, USA.1993.

MENESES, P.R; ALMEIDA, T.DE. Introdução ao processamento de imagens de sensoriamento remoto. Universidade de Brasília. 276 p. 2012.

MICHEL, K., TERHOEVEN-URSELMANS, T., NITSCHKE, R., STEFFAN, P., LUDWIG, B. Use of near and mid-infrared spectroscopy to distinguish carbon and nitrogen originating from char and forest-floor material in soils. *J. Plant Nutr. Soil Sci.* 172, 63–70.2009.

MINASNY, B; TRANTER, G.; MCBRATNEY, A.B.; BROUGH, D.M.; MURPHY, B.W. Regional transferability of mid-infrared diffuse reflectance spectroscopic prediction for soil chemical properties. *Geoderma*, [s.l.], v. 153, n. 1-2, p.155-162, out. 2009.

- MOHAMED, E.S.; SALEH, A.M.; BELAL, ABD_ALLAH GAD, A.B. Application of near-infrared reflectance for quantitative assessment of soil properties. *The Egyptian Journal Of Remote Sensing And Space Science*, [s.l.], p.1-14, fev. 2017.
- MORAES, E.C. De. Fundamentos do sensoriamento remoto. Ministério da Ciência e Tecnologia. Instituto Nacional de Pesquisas Espaciais-INPE . INPE 8984-PUD/62. São José dos Campos. 22 p. 2002.
- MORÓN, A. & COZZOLINO, D. Exploring the use of near infrared reflectance spectroscopy to study physical properties and microelements in soils. *Journal of Near Infrared Spectroscopy*, 11, 145-154.2003.
- MOUAZEN, A.M., KUANG, B., DE BAERDEMAEKER, J., RAMON, H. Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. *Geoderma* 158, 23–31.2010.
- NAWARA, S.; BUDDENBAUM, H.; HILL, J.; KOZAK, J.; MOUAZEN, A.M. Estimating the soil clay content and organic matter by means of different calibration methods of vis-NIR diffuse reflectance spectroscopy, *Soil Tillage Res.* 155. 510–522. 2016.
- NDUWAMUNGU, C., ZIADI, N., PARENT, L.-É., TREMBLAY, G.F. Mehlich 3 extractable nutrients as determined by near-infrared reflectance spectroscopy. *Can. J. Soil Sci.* 89, 579–587.2009.
- NIEDERBERGER, J., TODT, B., BOČA, A., NITSCHKE, R., KOHLER, M., KÜHN, P., BAUHUS, J. Use of near-infrared spectroscopy to assess phosphorus fractions of different plant availability in forest soils. *Biogeosciences* 12, 3415–3428.2015.
- NOVO, E. M. L. de M. Sensoriamento Remoto: Princípios e Aplicações. Editora Edgar Blücher Ltda. São José dos Campos, 1989. 308p.
- REEVES III, J.B. Near-versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: where are we and what needs to be done? *Geoderma* 158, 3–14.2010.
- REEVES III, J.B., FOLLETT, R.F., McCARTY, G.W., KIMBLE, J.M. Can near or mid-infrared diffuse reflectance spectroscopy be used to determine soil carbon pools? *Commun. Soil Sci. Plant Anal.* 37, 2307–2325. 2006.
- REEVES, J. B.; SMITH, D. B. The potential of mid- and near-infrared diffuse reflectance spectroscopy for determining major- and trace-element concentrations in soils from a geochemical survey of North America. *Applied Geochemistry*, [s.l.], v. 24, n. 8, p.1472-1481, ago. 2009.
- REEVES, J., McCARTY, G., MIMMO, T. The potential of diffuse reflectance spectroscopy for the determination of carbon inventories in soils. *Environmental Pollution* 116, 277–284.2002.

REEVES, JAMES B.. Near- versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: Where are we and what needs to be done?. *Geoderma*, [s.l.], v. 158, n. 1-2, p.3-14, ago. 2010.

SAVITZKY, A., GOLAY, M.J.E. Smoothing and differentiation of data by simplified least square procedure. *Analytical Chemistry* 36, 1627–1639.1964.

SEBEM, E . Colégio Politécnico da Universidade Federal de santa Maria-UFSM. Notas de aula. 2017b. Disponível em:<http://200.132.36.199/elodio/downloads/sr/SR_Aula_08_Solos.pdf> Acesso em: Mar, 2017.

SEBEM, E. Colégio Politécnico da Universidade Federal de santa Maria-UFSM. Notas de aula. 2017. Disponível em:<http://200.132.36.199/elodio/downloads/sr/SR_Aula_05.pdf> Acesso em: Mar, 2017.

SHEPHERD, K.D., WALSH, M.G. Development of reflectance spectral libraries for characterization of soil properties. *Soil Sci. Soc. Am. J.* 66, 988–998.2002.

SILVA, E.B., TEN CATEN, A. Estimating Soil Texture from a Limited Region of the Visible/Near-Infrared Spectrum. *Digital Soil Morphometr.*, 73–87.2016.

SORIANO-DISLA, J. M; JANIK, L.J; VISCARRA ROSSEL,R.A; MacDONALD, L.M; McLAUGHLIN, M.J. The Performance of Visible, Near-, and Mid-Infrared Reflectance Spectroscopy for Prediction of Soil Physical, Chemical, and Biological Properties. *Applied Spectroscopy Reviews*, [s.l.], v. 49, n. 2, p.139-186, 12 ago. 2013.

STEINER, F.; PIVETTA, L.A.; CASTOLDI, G.; COSTA, M.S.S.DE M.; COSTA; L.A. de M. Carbono orgânico e carbono residual do solo em sistema de plantio direto, submetido a diferentes manejos. *Revista Brasileira de Ciências Agrárias - Brazilian Journal Of Agricultural Sciences*, [s.l.], v. 6, n. 3, p.401-408, 19 set. 2011.

STONER, E.R.; BAUMGARDNER, M.F. Data acquisition through remote sensing. In: BEATTY, M.T. et al. *Planning the uses and management of land*. Madison : Soil Science Society of America, 1979. p.159-185.

SUMMERS, D.; LEWIS, M.; OSTENDORF, B.; CHITTLEBOROUGH, D. Visible near-infrared reflectance spectroscopy as a predictive indicator of soil properties. *Ecological Indicators*, [s.l.], v. 11, n. 1, p.123-131, jan. 2011.

TERRA, F. da S. Espectroscopia de reflectância do visível ao infravermelho médio aplicada aos estudos qualitativos e quantitativos de solos. (Tese de doutorado). Piracicaba, 2001, 374 p. Escola Superior de Agricultura Luiz de Queiroz-ESALQ.

UNICAMP. Universidade Estadual de Campinas. Instituto de Geociências. Princípios Básicos da Espectrometria. Disponível em<<http://www.ige.unicamp.br/espectrobauxita/espectroscopia.html>> Acesso em: Maio, 2017.

VAN GROENIGEN, J.W., MUTTERS, C.S., HORWATH, W.R., VAN KESSEL, C. NIR and DRIFT-MIR spectrometry of soils for predicting soil and crop parameters in a flooded field. *Plant Soil* 250, 155–165.2003.

VASQUES, G.M., GRUNWALD, S., SICKMAN, J.O. Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. *Geoderma* 146 (1–2), 14–25.2008.

VISCARRA ROSSEL R.A.,T; WALVOORT, D.J.J., MCBRATNEY, A.B.; JANIK, L.J.; SKJEMSTA, J.O. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma*, [s.l.], v. 131, n. 1-2, p.59-75, mar. 2006.

VISCARRA ROSSEL, R.A., BEHRENS, T. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma* 158 (1–2), 46–54.2010.

VISCARRA ROSSEL, R.A., HICKS, W.S. Soil carbon and its fractions estimated by visible near infrared transfer functions. *Eur. J. Soil Res.* 66, 438–450.2015.

VISCARRA ROSSEL, R.A., WALVOORT, D.J.J., McBRATNEY, A.B., JANIK, L.J., SKJEMSTAD, J.O. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* 131 (1–2), 59–75.2006.

VISCARRA ROSSEL, R.A., WEBSTER, R. Predicting soil properties from the Australian soil visible near infrared spectroscopic database. *Eur. J. Soil Sci.* 63, 848–860.2012.

VISCARRA ROSSEL, R.A.; McBRATNEY A.B. Laboratory evaluation of a proximal sensing technique for simultaneous measurement of clay and water content, *Geoderma*. 85. p.19–39.1998.

VOHLAND, M. ; M. LUDWIG ; S. THIELE-BRUHNB; LUDWIG, B. Determination of soil properties with visible to near- and mid-infrared spectroscopy: Effects of spectral variable selection. *Geoderma*, [s.l.], v. 223-225, p.88-96, jul. 2014.

VOHLAND, M., BESOLD, J., HILL, J., FRÜND, H.-C. Comparing different multivariate calibration methods for the determination of soil organic carbon pools with visible to near infrared spectroscopy. *Geoderma* 166 (1), 198–255.2011.

WALKLEY, A., BLACK, I.A. An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Science* 37, 29–38. 1934.

WILLIAMS, P.C. Variables affecting near-infrared reflectance spectroscopic analysis. In: WILLIAMS, P.C.; NORRIS, K. (Ed). *Near infrared technology in agriculture and food industries*. St Paul: American Association of Cereal Chemists, 1987. Cap 7. p. 143-167.

XIAOTING, P., TIEZHU, S., AIHONG, S., YIYUN, C., WENXIU, G. Estimating soil organic carbon using VIS/NIR spectroscopy with SVMR and SPA methods. *Remote Sens.* 6, 2699–2717.2014.

YANG, X.M., XIE, H.T., DRURY, C.F., REYNOLDS, W.D., YANG, J.Y., ZHANG, X.D. Determination of organic carbon and nitrogen in particulate organic matter and

particle size fractions of Brookston clay loam soil using infrared spectroscopy. *Eur. J. Soil Sci.* 63.177–188.2012.

ZIMMERMANN, M., LEIFELD, J., FURHRER, J. Quantification of soil organic carbon fractions by infrared spectroscopy. *Soil Biol. Biochem.* 39, 224–231.2007.

CAPÍTULO V

LEVANTAMENTO DE ATRIBUTOS DO SOLO UTILIZANDO ESPECTROMETRIA DE RAIOS-X : COMPARAÇÃO COM DADOS AEROGEOFÍSICOS E MAPA DE USO E OCUPAÇÃO DO SOLO

RESUMO

A fluorescência de raios X (XRF) é uma nova técnica que tem sido utilizada para análise de solos, sendo uma ferramenta dinâmica e extremamente proficiente para a determinação quantitativa e qualitativa da presença de elementos químicos em diversos tipos de amostras, de interesse biológico, industrial, geológico e ambiental. A análise por fluorescência de Raios-X é um processo quantitativo fundamentado na medida das intensidades dos Raios-X característicos, emitidos pelos elementos que constituem a amostra, quando excitada por partículas ou ondas eletromagnéticas. Ou seja, a técnica utilizada por esses aparelhos se baseia no conceito de emissão e absorção atômica, que podem ser explicados com base na energia utilizada para a excitação dos átomos. Utilizando uma energia baixa, pode-se levar um elétron a um nível superior na sua órbita, e ao retirar a energia ele retorna a seu nível inicial, tanto quando se eleva o nível e quando se retorna ao nível inicial o elétron libera um espectro luminoso característico que se utilizando das ferramentas adequadas, é possível detectá-lo e dependendo da intensidade do espectro numa amostra é possível quantificar o elemento. O princípio básico da Espectrometria por Fluorescência de Raios X consiste em utilizar uma grande energia para excitação dos átomos, fazendo com que o elétron seja removido do átomo, e outro átomo em um nível energético maior o substitua liberando uma energia na forma de fóton com espectro na faixa dos raios X. O conhecimento das características geofísicas da área de estudo comparadas ao mapa de uso e ocupação do solo podem ajudar a compreender a dinâmica do ambiente; e estes dados correlacionados à aqueles obtidos por meio da termofluorescência podem aprimorar este estudo permitindo a correlação desses dados com muitas outras variáveis. A área de pesquisa fica compreendida no município de Iconha/ES, e corresponde a bacia hidrográfica do Ribeirão Inhaúma. Os mapas de uso e ocupação do solo foram criados através de fotointerpretação. Os dados de gamaespectrometria foram obtidos através da CPRM. Para análises de fluorescência de raios-x foram coletados 100 pontos em duas profundidades, e analisados os teores totais de Al, Ca, Fe, K, Mg, Na, Si, Ti, Cr, Cu, Mn, Ni, P, Pb, V, Zn, Zr, S, Cl. Em relação aos dados geofísicos a área de estudo apresentou baixos teores de K, fato este que pode estar associado ao material de origem. As regiões de floresta foram aquelas que apresentam temperaturas mais amenas na região enquanto afloramento rochoso apresentou as máximas. Quatro componentes principais foram capazes de explicar a variação total nos dados provenientes de análise de raios-x. Dez variáveis apresentaram valores de correlação não linear acima de 0,8. O uso e ocupação do solo mostrou-se correlacionado com os dados geofísicos e de temperatura. Em relação aos dados aerogeofísicos foram encontradas alta correlação com altitude e baixa correlação com dados obtidos pela análise de raios-x.

Palavras-chave: fluorescência de raios-X, gamespectromectria, magnetometria

ABSTRACT

X-ray fluorescence (XRF) is a new technique that has been used for soil analysis, being a dynamic and extremely proficient tool for the quantitative and qualitative determination

of the presence of chemical elements in several types of samples of biological, industrial interest, geological and environmental. X-ray fluorescence analysis is a quantitative process based on the measurement of the characteristic X-ray intensities emitted by the sample elements when excited by particles or electromagnetic waves. That is, the technique used by these devices is based on the concept of atomic emission and absorption, which can be explained based on the energy used for the excitation of atoms. Using a low energy, one can take an electron to a higher level in its orbit, and when withdrawing the energy it returns to its initial level, both when the level rises and when it returns to the initial level the electron releases a luminous spectrum characteristic that using the appropriate tools, it is possible to detect it and depending on the intensity of the spectrum in a sample it is possible to quantify the element. The basic principle of X-Ray Fluorescence Spectrometry is to use a large energy to excite the atoms, causing the electron to be removed from the atom, and another atom at a higher energy level to replace it by releasing an energy in the form of a photon with spectrum in the X-ray range. Knowledge of the geophysical characteristics of the study area compared to the map of land use and occupation can help to understand the dynamics of the environment; and these data correlated to those obtained by means of thermofluorescence can improve this study allowing the correlation of this data with many other variables. The research area is comprised in the municipality of Iconha / ES, and corresponds to the hydrographic basin of Ribeirão Inhaúma. The maps of land use and occupation were created through photointerpretation. Gamma spectrometry data were obtained through CPRM. The total contents of Al, Ca, Fe, K, Mg, Na, Si, Ti, Cr, Cu, Mn, Ni, P and Pb were analyzed for fluorescence analysis of x- V, Zn, Zr, S, Cl. In relation to the geophysical data the study area presented low levels of K, a fact that may be associated with the source material. The forest regions were those that presented milder temperatures in the region while rocky outcropping showed the maximum. Four main components were able to explain the overall variation in data from x-ray analysis. Ten variables presented non-linear correlation values above 0.8. The use and occupation of the soil was correlated with the geophysical and temperature data. In relation to aerogeophysical data, high correlation with altitude and low correlation with data obtained by x-ray analysis were found.

Key words: X-ray fluorescence, gamma spectrometry, magnetometry

1 INTRODUÇÃO

São variadas as formas de se estudar a superfície terrestre. Dados de sensoriamento remoto aliados a estudos de campo se apresentam como uma dessas formas consideradas eficientes. Os métodos aerogeofísicos compreendem também tipo de técnica de sensoriamento remoto, onde se realiza medidas indiretas de investigação das estruturas em subsuperfície (SANTOS et al., 2008), sendo a magnometria e a gamaespectrometria os principais resultados desta técnica, onde a segunda é baseada nos princípios da geofísica nuclear, onde existem três tipos de partículas provenientes de núcleos instáveis de substâncias radioativas: α ; β e γ (FERREIRA et al., 2010).

A radioatividade dos solos está diretamente relacionada ao seu material originário e a processos que atuaram na sua formação. O levantamento gamaespectrométrico reflete a variação geoquímica do potássio (K), do urânio (U) e do tório (Th) nos 30 cm superiores da superfície da terra (SANTOS et al., 2008), ou seja a gamaespectrometria mede as contagens/concentrações destes elementos nas rochas,

sedimentos e solos através da detecção dos raios gama emitidos durante o decaimento radioativo natural destes elementos (ADAMS e GASPARINI, 1970). A abundância do urânio na crosta terrestre oscila entre 2 e 3 ppm, a do tório entre 8 e 12 ppm e a do potássio entre 2 e 2,5% (MINTY, 1997; BONOTTO, 2004).

O tório é um elemento altamente eletropositivo e com afinidade litófila, ocorrendo principalmente em óxidos, silicatos e fosfatos (SILVA et al., 2010). Ulbricht et al. (2009) relatam que mesmo em regiões de clima tropical o tório pode comportar-se como o mais estável dos três radioelementos. O urânio ocorre em baixa concentração na crosta terrestre, sendo quimicamente dominado por seus estados de valência U^{+4} e U^{+6} (Becegato e Ferreira, 2005). O primeiro estado geralmente está contido em minerais não solúveis, enquanto o segundo associa-se com ânions nos carbonatos, sulfatos e fosfatos para formarem espécies solúveis (DICKSON e SCOTT, 1997). A principal ocorrência do K nos continentes é nos feldspatos potássicos (Ulbricht et al. 2009) e é extremamente móvel e susceptível ao intemperismo, principalmente em ambientes com clima tropicais e subtropicais, com forte regimes de chuvas (WILFORD et al., 1997).

A gamaespectrometria depende de muitas variáveis, podendo ser citados os contrastes das propriedades físicas e geométricas das fontes radioativas, o tamanho e a resolução dos detectores, os efeitos do meio ambiente, como umidade do solo, e, no caso de levantamentos aéreos, o movimento das fontes aéreas de radiação na atmosfera baixa (ULBRICH et al., 2009).

Se por um lado a gamaespectrometria já é uma técnica amplamente utilizada em estudo de solos, não há grandes relatos da comparação destes dados com valores obtidos por análises de Raios-X. A análise por Fluorescência de Raios-X é um método qualitativo baseado na medida das intensidades dos Raios-X característicos emitidos pelos elementos que constituem a amostra (POTTS, 2008), ou seja, realiza análises da composição química de uma amostra através da detecção dos raios-X característicos dos seus elementos.

Em resumo, a análise por fluorescência de Raios-X consiste de três fases: I. Excitação dos elementos que constituem a amostra; II. Dispersão dos espectros de linha emitidos pela amostra; III. Detecção dos espectros de linha emitidos (QUEIRÓS, 2011). Os limites de detecção dos elementos analisados dependem: da energia dos Raios-X fluorescentes recebida pelo analisador (consequência do número atômico do elemento), do tipo e energia da fonte de excitação, do tipo de matriz, de interferências interespectrais, da capacidade de resolução do equipamento e da duração dos ensaios (QUEIRÓS, 2011).

Utilizando-se uma fonte de energia, pode-se levar um elétron a um nível superior na sua órbita, e ao retirar a energia ele retorna a seu nível inicial, tanto quando se eleva o nível e quando se retorna ao nível inicial o elétron libera um espectro luminoso característico, que se utilizando das ferramentas adequadas é possível detectá-lo e dependendo da intensidade do espectro numa amostra é possível quantificar o elemento (BÖGEL NETO, 2012).

A comparação desses dados com o mapa de uso e ocupação do solo permitem realizar inferências sobre a área. Conforme Santos & Santos (2010), o mapeamento do uso e cobertura do solo é considerado por vários autores uma importante ferramenta de conhecimento das rápidas transformações da paisagem, pois permite a obtenção de informações para construção de cenários ambientais e indicadores que poderão servir de subsídios práticos à avaliação da capacidade de suporte ambiental, direcionando assim o melhor uso do solo naquele ambiente.

Dessa forma este trabalho objetivou analisar o grau de correlação entre informações dos dados aerogeofísicos e aqueles obtidos pela análise de

termofluorescência feita com analisador de raios-X, comparando os valores encontrados com o mapa de uso e ocupação do solo.

2 REVISÃO BIBLIOGRÁFICA

2.1 Espectrometria Por Fluorescência De Raios-X

A análise por Fluorescência de Raios X é um método quali-quantitativo baseado na medida das intensidades dos Raios X característicos emitidos pelos elementos que constituem a amostra (POTTS, 2008). Segundo Jenkins (1988), a espectrometria de fluorescência de raios-X permite identificar elementos através da medição do comprimento de onda ou da energia de uma emissão de raios-X.

Os raios X emitidos por tubos de raios X, ou raios X ou gama por uma fonte radioativa, excitam os elementos que constituem, os quais, por sua vez, emitem linhas espectrais com energias características do elemento e cujas intensidades estão relacionadas com a concentração do elemento na amostra (NASCIMENTO FILHO, 1999). Ou seja, a espectrometria de fluorescência de raios-X realiza análises da composição química de uma amostra através da detecção dos raios-X característicos dos seus elementos (URBANO E BRANDÃO, 2013).

Num átomo estável, o número de elétrons é igual ao número de prótons. As órbitas ou camadas têm um número específico de elétrons e são caracterizadas por apresentarem diferentes níveis de energia. As órbitas mais internas têm estados de energias mais baixos do que as mais externas, o que significa que têm menor quantidade de energia armazenada.

A nomenclatura destas órbitas ou camadas é feita alfabeticamente, começando na mais interna que é designada pela letra K, seguindo-se as letras L, M, N, etc. O preenchimento das camadas orbitais por elétrons inicia-se nas órbitas mais próximas do núcleo e vai-se processando sucessivamente para as órbitas mais externas. Quando um átomo é atingido por fótons de alta energia, como os Raios-X, provenientes de uma fonte radioativa, os elétrons absorvem essa energia ou usam-na para se ejetarem do interior da estrutura atômica.

Se a fonte irradiar o átomo com energia superior à energia de absorção dos elétrons da camada mais interna, estes são ejetados para fora da estrutura, criando-se valências nesta camada. Estas valências serão preenchidas por elétrons que se deslocarão das camadas mais externas para as mais internas, de forma a garantir o rearranjo estrutural do átomo. Como os elétrons das órbitas mais externas se encontram em estados de energia mais elevados, ao passarem para camadas menos enérgicas, dissipam o excesso de energia sob a forma de emissão de radiação secundária. Este tipo de emissão de radiação secundária produzida como consequência do rearranjo atômico é específica de cada elemento, ou seja, cada elemento da tabela periódica produz, quando excitado, uma radiação secundária única e característica. A esta radiação emitida dá-se o nome de radiação característica ou fluorescência (JORGE & ROQUE, 2005).

Em resumo, a análise por fluorescência de Raios-X consiste de três fases: I. Excitação dos elementos que constituem a amostra; II. Dispersão dos espectros de linha emitidos pela amostra; III. Detecção dos espectros de linha emitidos (QUEIRÓS, 2011).

Os limites de detecção dos elementos analisados dependem: da energia dos Raios-X fluorescentes recebida pelo analisador (consequência do número atômico do elemento), do tipo e energia da fonte de excitação, do tipo de matriz, de interferências interespectrais, da capacidade de resolução do equipamento e da duração dos ensaios (QUEIRÓS, 2011).

Existem ainda outros fatores que podem condicionar a qualidade dos resultados das análises realizadas, como o teor de água da amostra, a temperatura ambiente, o incorreto posicionamento do analisador, uma imensa variedade nas propriedades físicas (dimensão, uniformidade, etc.) das partículas ou a presença simultânea de elementos químicos com concentrações muito distintas (QUEIRÓS, 2011).

2.2 Estudo Da Correlação Entre Dados Aerogeofísicos E Dados Obtidos Por Termofluorescência

A radioatividade dos solos está diretamente relacionada ao seu material originário e a processos que atuaram na sua formação. O levantamento gamaespectrométrico reflete a variação geoquímica do K, do U e do Th nos 30 cm superiores da superfície da terra (SANTOS et al., 2008), ou seja a gamaespectrometria mede as contagens/concentrações de K, U e Th em rochas, sedimentos e solos através da detecção dos raios gama emitidos durante o decaimento radioativo natural destes elementos (ADAMS E GASPARINI, 1970). A abundância do urânio na crosta terrestre oscila entre 2 e 3 ppm, a do tório entre 8 e 12 ppm e a do potássio entre 2 e 2,5% (BONOTTO, 2004).

O tório é um elemento altamente eletropositivo e com afinidade litófila, ocorrendo principalmente em óxidos, silicatos e fosfatos (SILVA et al., 2010). Ulbricht et al. (2009) relatam que mesmo em regiões de clima tropical o tório pode comportar-se como o mais estável dos três radioelementos. O urânio ocorre em baixa concentração na crosta terrestre, sendo quimicamente dominado por seus estados de valência U^{+4} e U^{+6} . (BECEGATO E FERREIRA, 2005). O primeiro estado geralmente está contido em minerais não solúveis, enquanto o segundo associa-se com ânions nos carbonatos, sulfatos e fosfatos para formarem espécies solúveis (DICKSON E SCOTT, 1997). A principal ocorrência do K nos continentes é nos feldspatos potássicos (ULBRICH et al. 2009), e é extremamente móvel e susceptível ao intemperismo, principalmente em ambientes com clima tropicais e subtropicais, com forte regimes de chuvas (WILFORD et al., 1997).

A gamaespectrometria depende de muitas variáveis, podendo ser citados os contrastes das propriedades físicas e geométricas das fontes radioativas, o tamanho e a resolução dos detectores, os efeitos do meio ambiente e, no caso de levantamentos aéreos, o movimento das fontes aéreas de radiação na atmosfera baixa (ULBRICH et al., 2009).

Sendo assim, ambas formas de análises mostram resultados relativos aos elementos totais presentes no solo.

A comparação desses dados com o mapa de uso e ocupação do solo permitiram realizar inferências sobre a área. Conforme Santos & Santos (2010), o mapeamento do uso e cobertura do solo é considerado por vários autores uma importante ferramenta de conhecimento das rápidas transformações da paisagem, pois permite a obtenção de informações para construção de cenários ambientais e indicadores que poderão servir de subsídios práticos à avaliação da capacidade de suporte ambiental, direcionando assim o melhor uso do solo naquele ambiente.

3 MATERIAL E MÉTODOS

O trabalho foi conduzido na bacia hidrográfica do Ribeirão Inhaúma, com área de 2.403,9 ha, coordenada central de 21°10'58.82"S e 41°00'08.87"O, localizada quase que totalmente no município de Iconha, no sul do estado do Espírito Santo. A área apresenta-se com relevo predominantemente montanhoso e forte ondulado. Segundo o levantamento de solos realizado pela Embrapa, e publicado pelo IBGE (2001), os solos presentes na região são: Cambissolo Háplico, Latossolo vermelho-amarelo e afloramentos de rocha. A geologia predominante na área de estudo é o metatonalito.

Os dados geofísicos foram obtidos através do levantamento realizado pelo Serviço Geológico do Brasil - CPRM no ano de 2010. Foram utilizados gamaespectrômetro *Radiations Solutions*, modelo RS-500, de 1024 canais espectrais, onde o espectro de cada um dos cristais detectores é analisado individualmente para determinação precisa dos fotopicos de K, U e Th. Uma correção linear foi aplicada individualmente a cada cristal, mantendo o espectro permanentemente alinhado. As radiações gama detectadas foram somadas e as leituras reduzidas (normalizadas) a uma única saída de 256 canais espectrais. As leituras foram realizadas a cada segundo, representando medições a intervalos de amostragem de aproximadamente 75 m no terreno (CPRM, 2010).

Os dados obtidos foram submetidos previamente pela CPRM (2010) as correções do tempo morto, filtragem (não linear e do tipo passa-baixa), correlação do erro de paralaxe, cálculo da altura efetiva de vôo, remoção do *background* da aeronave e da radiação cósmica, remoção do *background* do radônio, estimativa dos coeficientes *Skyshine*, correção do efeito *Compton*, correção altimétrica e conversões para determinar as concentrações dos elementos. A resolução espacial é de 125 m.

Previamente a comparação com os dados provenientes da análise de termofluorescência os dados de gamaespectrometria foram comparados com o mapa de uso e ocupação do solo, este foi produzido utilizando o software ArcGis 10.2 a partir da observação de elementos básicos de interpretação, como: tonalidade e cor, forma e tamanho, padrão, textura, sombra. As imagens ortorretificadas da área, com resolução de 25 cm, utilizadas neste estudo, foram cedidas pelo Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural-INCAPER, por meio do Sistema Integrado de Bases Geoespaciais do Estado do Espírito Santo- GEOBASES.

As classes definidas foram: afloramento rochoso com vegetação, água, área construída, bambú, banana, café, café com banana, capoeira, estradas, eucalipto, floresta, palmeira, pastagem e solo exposto (Figura 67). Para auxiliar nas discussões esses mapas foram comparados também ao modelo digital de elevação-MDE, obtido a partir de dados do satélite ALOS com resolução de 12,5 m. Foi utilizada ferramenta adicional (toolbox) do Arcgis 10.2 para recuperação automática da temperatura da superfície terrestre (TST) a partir de dados do sensor TIRS do Landsat 8, conforme procedimento de estimação da temperatura da superfície terrestre desenvolvido de Sobrino e Jimenez-Munoz (2014).

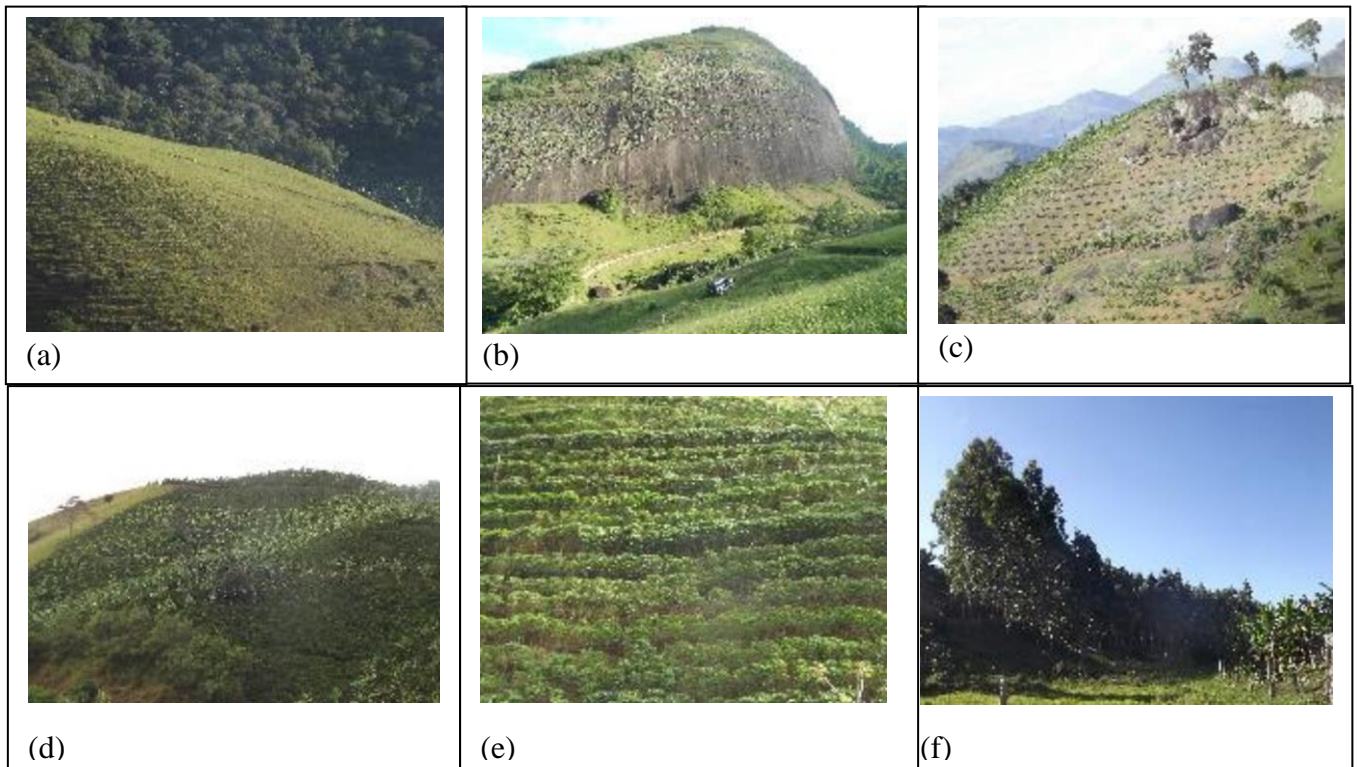


Figura 67- Principais classes de uso e ocupação do solo definidas na Bacia do Ribeirão Inhaúma em Iconha/ES.

a- Pastagem e floresta; b- afloramento rochoso com vegetação; c- banana; d- café com banana; e- café; f- eucalipto.

As amostras para análise de termofluorescência foram coletadas ao longo de toda bacia, em 100 pontos, na profundidade 0-10 cm. Os locais de amostragem foram determinados utilizando o método do hipercubo latino condicionado, sendo considerado as possibilidades de acesso devido às dificuldades de deslocamento na área em função do relevo acentuado. Dessa forma a malha de pontos, ficou sempre próxima às estradas (Figura 68). Em cada um dos pontos, foram registradas as coordenadas utilizando-se receptor GNSS da Leica, modelo GS08Plus, cujos dados foram processados no programa Leca Geoffice 8.0. A base foi ajustada a partir da estação fixa do IBGE existente em Vitória/ES.

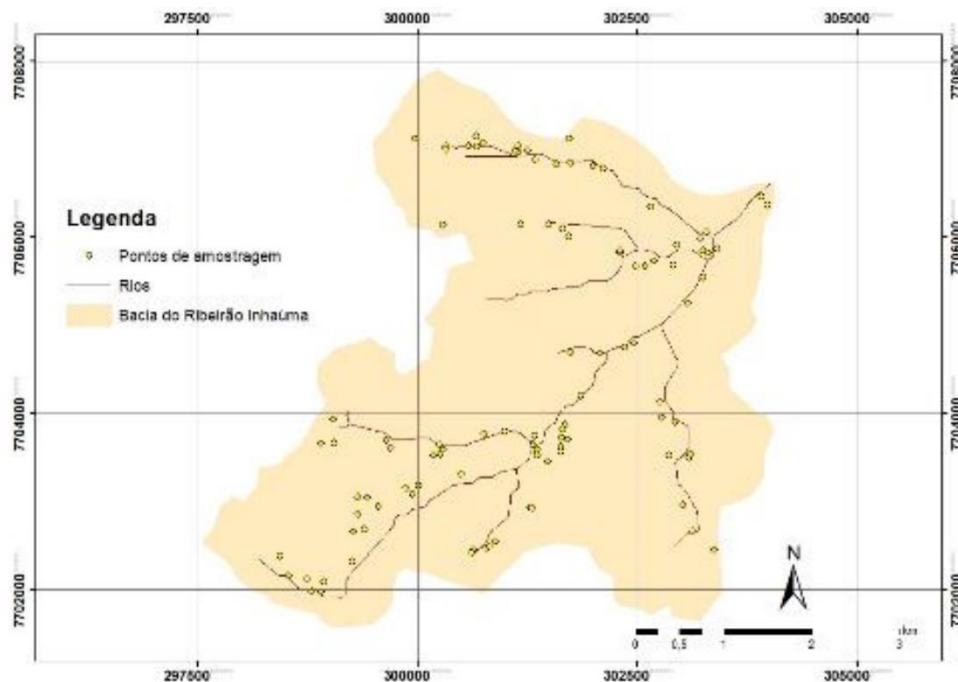


Figura 68- Localização dos pontos de amostragem na bacia do Ribeirão Inhaúma, Iconha/ES.

As amostras foram secas ao ar, destorroadas e passadas em peneira com malha de 2 mm. No laboratório foram maceradas em almofariz de ágata e novamente peneiradas em malha 1 mm, posteriormente colocadas em molde padrão e prensadas manualmente, a fim de se formarem pastilhas.

Essas pastilhas foram usadas para leitura no analisador de termofluorescência de raios-X do Laboratório de Solos, da Universidade Federal de Viçosa-UFV. Foi utilizado para tal o modelo Micro-EDX-1300 da Shimadzu, foram usadas as voltagens padrão do equipamento.

Após a análise de termofluorescência foi feita a verificação dos dados quanto a presença de *outliers* e dados perdidos (*missing data*), esses foram realizados com auxílio do software SPSS, utilizando-se a aba *analyse*, e o comando *descriptive statistic* para *outliers*, e para dados perdidos usou-se o comando *analyse* e posteriormente *Missing value analyse*. Em ambos casos para substituição desses dados utilizou-se a imputação por regressão, no qual o valor faltante/outlier de uma variável é estimado por intermédio da regressão dessa variável às demais.

Foram feitas análise de correlação de Pearson; correlação não linear pelo *Randomized Dependence Coefficient-RDC* (Lopez-Paz et al., 2013), utilizando-se o software R, e análise de componentes principais – PCA para os dados obtidos a partir da termofluorescência, através do software SPSS. Previamente a PCA foi feito o teste de normalidade de Kolmogorov Smirnov (K-S) ($p < 0,05$).

4 RESULTADOS E DISCUSSÃO

A Figura 69 mostra o mapa de uso e ocupação do solo. Os mapas com as concentrações de Urânio (U) em ppm; Tório (Th) em ppm e Potássio (K) em %, são mostrados na figura 70. A geologia do local é formada por metatonalitos, que são classificados como rochas magmáticas. Petrograficamente os metatonalitos tem como principais minerais essenciais o plagioclásio, o quartzo, o ortopiroxênio e o clinopiroxênio e como minerais secundários a hornblenda e a biotita; petroquimicamente os metatonalitos são rochas de filiação cálcio-alcalina de baixo teor de potássio (PINHO, 2005).

Rochas de composição mineralógica caracterizada essencialmente por quartzo e plagioclásio apresentam baixas concentrações de K, U e Th, enquanto que as rochas compostas por micas, feldspatos e minerais acessórios como o zircão, a apatita e a monazita, contêm teores maiores desses elementos (SAPUCAIA et al, 2005). Alguns minerais acessórios portadores de Th e U podem ser citados, como zircão, titanita, epidoto primário, monazita, apatita, xenotima e allanita (ULBRICH, 2009). Essa informação é corroborada por Bea (1996) que ressalta que esses dois elementos estão comumente presentes em minerais acessórios encontrados em variadas rochas ígneas e metamórficas quartzo-feldspáticas.

Já em relação ao K, os locais onde esse elemento está presente em maior quantidade pode estar relacionado com as áreas usadas para agricultura, já que os produtores da região fazem grandes tratamentos culturais incluindo adubação potássica, principalmente na cultura da banana, bastante exigente desse mineral e com área bem representativa na bacia (72,81 ha). Porém, vale ressaltar que o local com maior teor de K encontra-se sob áreas de pastagem, floresta e afloramento rochoso com vegetação, nas regiões de maior altitude.

A concentração de U pode também estar relacionada com a adição de adubos, já que os locais que apresentam maiores teores deste correspondem a ocupação de áreas com banana, café e pastagem. Já em 1976, Pfister et al. observaram elevada concentração do urânio em superfosfato triplo, e super-simples. Righi et al. (2000) obtiveram valores de U e Th em amostras de adubos fosfatados (super triplo).

Em relação ao Th, a concentração de U é cerca de seis a sete vezes menor, isso pode estar relacionado ao clima, onde regiões quente e úmida reúne condições necessárias para que ocorra lixiviação do urânio presente nas rochas. Esse tipo de clima promove e favorece circulação de águas, em geral ácidas, com oxigênio e dióxido de carbono dissolvidos, fator esse determinante para que os minerais contendo U, exceto os refratários, sejam hidrolisados, oxidados e lixiviados na forma do íon hidroxila $(\text{UO}_2)^{2+}$, que em ambiente oxidante tem alta mobilidade, passando para a solução do solo (WEIDJEN e WEIDJEN, 1995).

Além disso, a mobilidade do U^{+6} é modificada pela adsorção a óxidos hidratados de ferro, minerais de argila e coloides, como também pela transformação em minerais de U^{+4} nos ambientes redutores (BECEGATO e FERREIRA, 2005).

Os principais minerais que contêm Th (monazita e zircão) são estáveis durante o intemperismo e podem acumular-se. O tório liberado durante o intemperismo, pode ser retido em óxidos e hidróxidos de Fe e em argilas, podendo ser transportado por estes quando adsorvido (BECEGATO e FERREIRA, 2005). Dessa forma entende-se que o Th está mais presente na área de estudo por ser mais resistente as intempéries que o U. Rebello (2000) estudando rochas graníticas, concluiu que o Th devido a sua baixa mobilidade geoquímica, permanece fixo durante o intemperismo e a pedogênese.

Em geral se observa uma correlação positiva entre os teores de Th e K, ou seja, os locais com maiores valores para Th também apresentam maiores valores para K, sendo que as regiões com menores valores também coincidem.

Ao comparar os mapas geofísicos (Figura 70) com o modelo digital de elevação-MDE (Figura 71) percebe-se que os maiores teores de K são encontrados na parte mais alta e os menores estão na parte mais baixa, coincidente com a parte mais próxima ao rio, assim como o Th que apresenta o mesmo comportamento. Enquanto o U apresenta comportamento diferente, tendo maiores teores em altitudes intermediárias. Segundo Ferreira et al, 2009 o Th, geralmente contido como elemento menor em minerais acessórios, é de mobilização mais difícil e pode ficar retido, durante o intemperismo, junto com esses minerais no solo residual.

Não é o caso do U, cuja mobilidade frente ao intemperismo é maior. K, por outro lado, pode ser inteiramente lixiviado nas regiões de climas tropicais a subtropicais, razão pela qual as medidas gamaespectrométricas nessas regiões, em solos bem desenvolvidos e aluviões, podem registrar valores próximos de zero ou até zero. Ao se proceder a análise de correlação não linear, entre altitude e dados geofísicos, comprova-se a forte correlação existente entre estes (Tabela 45).

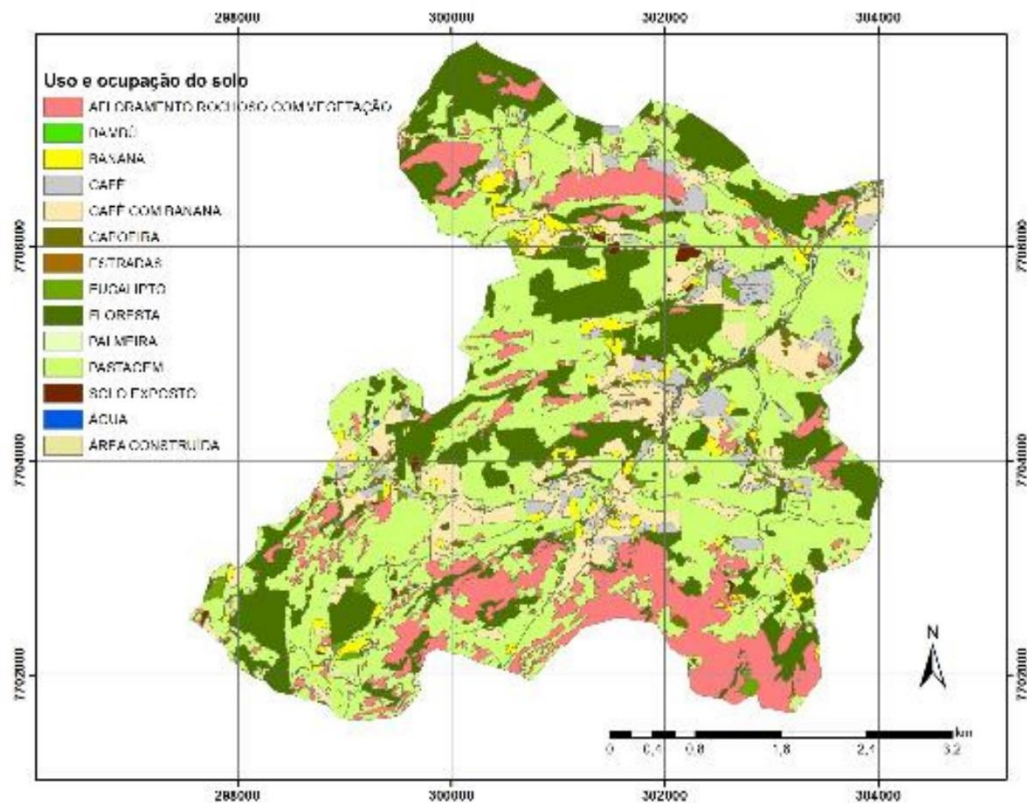


Figura 69- Mapa de uso e ocupação do solo.

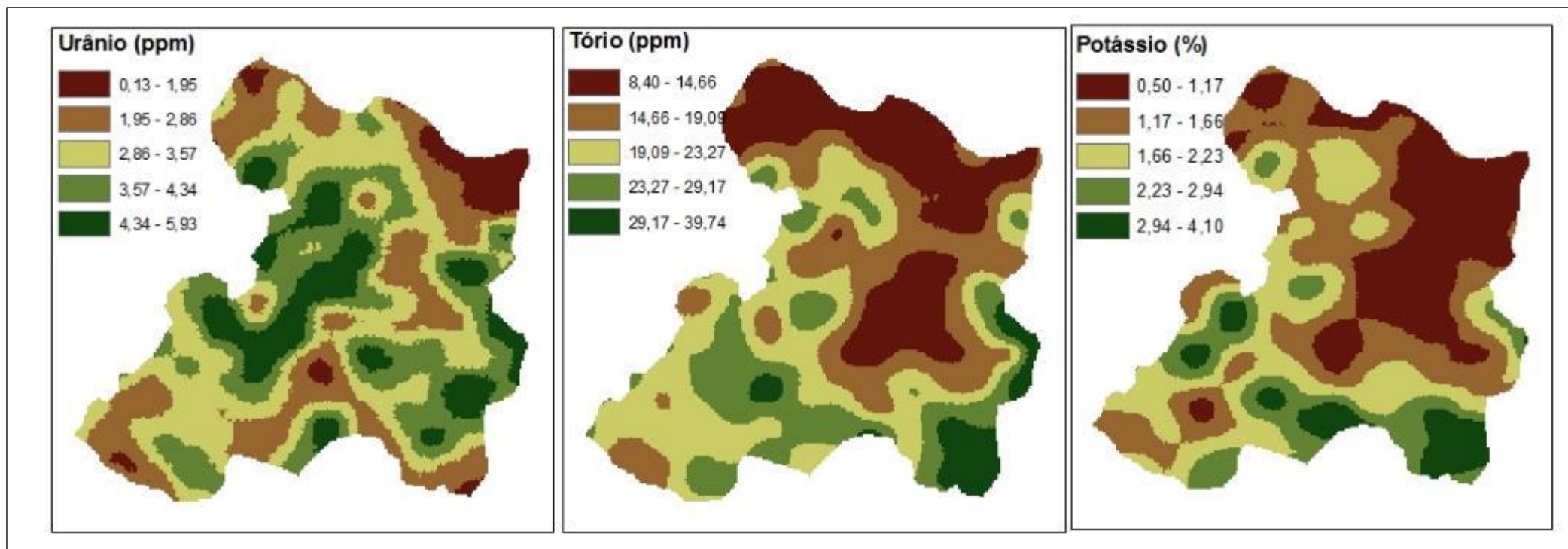


Figura 70- Mapa de Potássio (%), Tório (ppm) e Urânio (ppm).

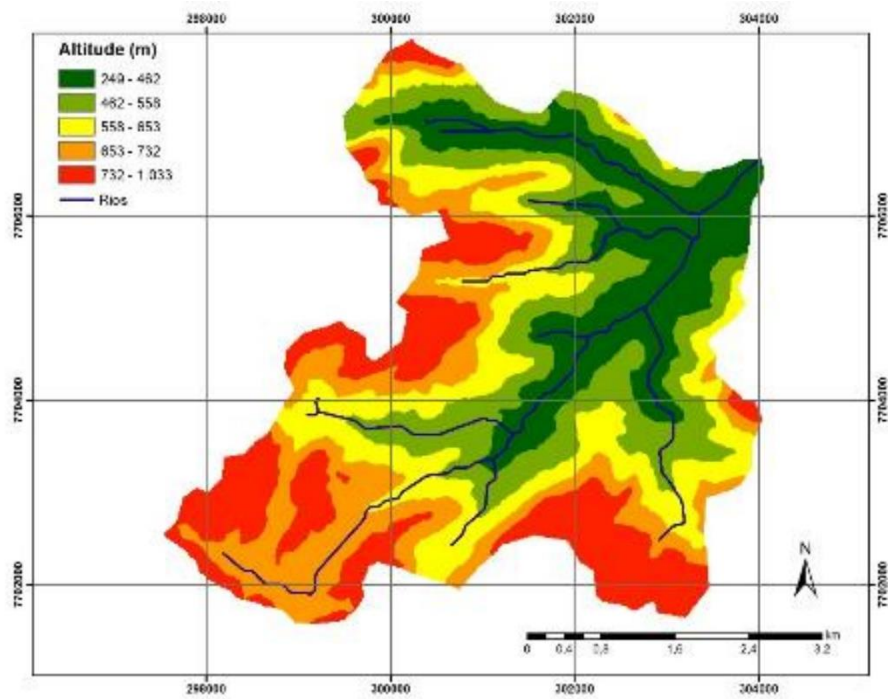


Figura 71- Modelo digital de elevação.

Tabela 45- Correlação não linear entre altitude e dados geofísicos

	Z	Th	U	K	Th/K	U/K	U/Th
Z	1,00	0,88	0,78	0,90	0,53	0,77	0,79
Th	0,87	1,00	0,39	0,85	0,54	0,56	0,68
U	0,78	0,40	1,00	0,54	0,37	0,43	0,60
K	0,85	0,85	0,54	1,00	0,81	0,81	0,56
Thkraz	0,51	0,54	0,49	0,80	1,00	0,68	0,42
Ukraz	0,75	0,56	0,45	0,81	0,68	1,00	0,74
Uthraz	0,81	0,69	0,63	0,52	0,42	0,73	1,00

Z-Altitude; Th – Tório; U-Urânio; Th/K- Razão Tório e K; U/K- Razão Urânio e Potássio; U/Th- Razão Urânio e Tório

Os dados de temperatura da superfície (Figura 72) obtido a partir das bandas 10 e 11 do sensor termal (TIRS) do satélite Landsat 8, apresentou uma amplitude de mais de 10 °C e correlação positiva entre a verdadeira temperatura cinética do alvo e o fluxo radiante emitido pelos mesmos (SOUSA e FERREIRA JUNIOR, 2012). Apesar da diferença de mais de 1.200 m entre a menor e maior altitudes, observa-se que o uso do solo foi o que mais influenciou a temperatura do solo (Tabela 46). Esses dados têm sido utilizados para se estimar mudanças no uso e ocupação do solo e alterações de índices de vegetação (GALLO e XIAN, 2014).

Dessa forma ao se comparar esses dados ao mapa de uso e ocupação do solo percebe-se que os locais onde se encontram as maiores temperaturas são ocupados por afloramento rochoso com vegetação, com máxima de 24,92 °C e média de 18,79°C ; capoeira, com máxima de 25,34° e mínima de 12,10° e principalmente por áreas de pastagem com valores máximos de 26,07°C e média de 18,42 °C, associadas a esses há também pequenas áreas de solo exposto, que apresentou temperatura média de 18,30 °C.

Já as áreas que possuem menor temperatura são aquelas próximas aos trechos de drenagem do rio, sendo ocupadas por cultivos de café, com mínima de 10,51 °C e média de 16,86 °C . Os valores mais baixos para temperatura da superfície independentemente da localização na bacia é para área de floresta, com valores médios de 15,38 °C. Silva et al. 2015 encontraram tendências semelhantes, onde áreas de florestas mais densas absorvem mais radiação que áreas desmatadas e pastagens e apresentam menor temperatura de superfície.

É importante ressaltar que existem muitos trabalhos onde é feita a comparação entre a temperatura da superfície e o uso e ocupação do solo na zona urbana (SOUZA et al, 2016; AMORIM e MONTEIRO, 2011; YOUNG, 2005) porém é raro encontrar trabalhos que façam comparação entre diferentes uso e ocupação do solo em toda uma bacia tipicamente rural.

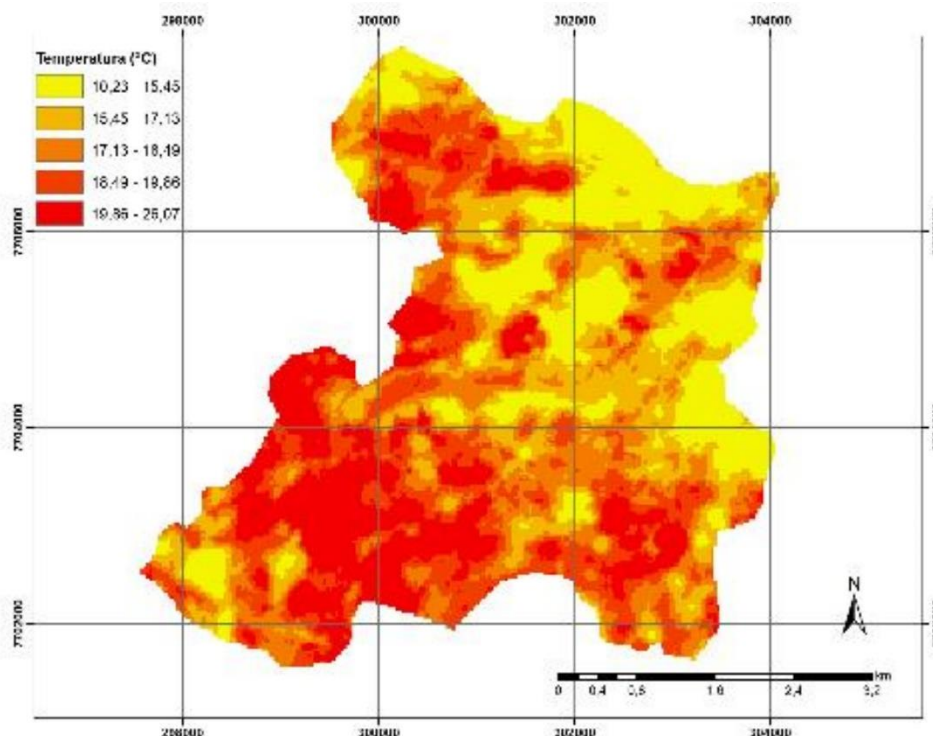


Figura 72- Mapa de temperatura da superfície para a bacia do Ribeirão Inhaúma, Iconha/ES.

Tabela 46- Valores de temperatura, altitude e área para os diferentes usos do solo na bacia do Ribeirão Inhaúma, Iconha/ES

Uso e ocupação do solo	Área	Temp min	Temp max	Temp média	Altitude min	Altitude max	Altitude média
	ha	°C			m		
Afloramento rochoso com vegetação	354,96	11,20	24,92	18,79	249	1033	685
Água	2,07	13,55	22,48	18,12	301	721	487
Área construída	8,82	14,22	22,41	17,96	249	795	451
Bambu	2,79	14,71	20,82	18,53	301	825	520
Banana	72,81	12,44	23,62	17,93	249	800	556
Café	91,35	10,51	23,13	16,86	249	791	469
Café com banana	236,97	11,54	24,38	17,66	249	811	515
Capoeira	26,82	12,10	25,34	17,14	249	860	519
Estradas	19,35	13,00	23,42	18,08	249	795	494
Eucalipto	16,83	14,65	23,75	17,52	390	874	655
Floresta	507,42	10,23	23,18	15,38	249	1004	632
Palmeira	0,27	16,14	20,46	18,17	346	588	385
Pastagem	1052,37	11,01	26,07	18,42	249	1021	594
Solo exposto	10,89	12,46	22,99	18,30	275	745	549

Temp min- Temperatura mínima; Temp max- temperatura máxima; Temp média-Temperatura média; Altitude min- Altitude mínima; Altitude max- Altitude máxima

Ao se proceder o teste de normalidade para profundidade 0-10 cm (Tabela 47), percebeu-se que Ca, Si, Ti, Cr, Cu, Ni, P, V, S não apresentaram distribuição normal mesmo após a transformação logarítmica e raiz quadrada dos dados. Dessa forma estes não foram considerados na análise de componentes principais para essa profundidade. Silva e Lima (2012), estudando a fertilidade de um Latossolo Vermelho-Amarelo não encontraram normalidade para os dados de P, K, Na, Ca, Mg, Fe, Cu.

Tabela 47- Teste de normalidade para a profundidade 0-10 cm

Kolmogorov-Smirnov					
Elemento	Nº de amostras em cada prof.	0-10 cm	Elemento	Nº de amostras em cada prof.	0-10 cm
Al	100	0,200*	Cu	100	0,003
Ca	100	0,000	Mn	100	0,200*
Fe	100	0,200*	Ni	100	0,001
K	100	0,200*	P	100	0,014
Mg	100	0,200*	Pb	100	0,200*
Na	100	0,200*	V	100	0,052
Si	100	0,039	Zn	100	0,200*
Ti	100	0,028	Zr	100	0,200*
Cr	100	0,014	S	100	0,005

*Distribuição normal pelo teste de Kolmogorov-Smirnov a 5 %

A análise de componentes principais apontou a separação em quatro grupos. O número de componentes principais foi definido de acordo com Johnson & Wichern (2002), onde se utilizou os componentes associados com valores maiores que 1 (um), sendo que estes explicaram juntos 75% da variação dos dados (Tabela 48), atendendo assim o pressuposto indicado por Kaiser (1960), onde considera-se a retenção das componentes que explicam mais de 70 % da variância. O componente principal 1 (PC1) explicou 34,5 % dos dados; o componente principal 2 (PC2) 18,1 %, o componente principal 3 (PC3) 12,1 % e o componente principal 4 (PC4) 10,3 % (Tabela 4). Segundo Johnson & Wichern (2002) os primeiros componentes são responsáveis pela maior parte da variabilidade contida no conjunto de dados.

Tabela 48- Variância total explicada pelos componentes principais

Componente	Componentes totais			Componentes principais		
	Total	% da variância	% da Variância Acumulada	Total	% da variância	% da Variância Acumulada
1	3,447	34,475	34,475	3,447	34,475	34,475
2	1,813	18,133	52,608	1,813	18,133	52,608
3	1,212	12,117	64,725	1,212	12,117	64,725
4	1,025	10,253	74,978	1,025	10,253	74,978
5	0,758	7,578	82,556			
6	0,593	5,933	88,488			
7	0,412	4,122	92,610			
8	0,274	2,745	95,355			
9	0,237	2,366	97,721			
10	0,228	2,279	100,000			

Para PC1 (Tabela 49) as variáveis que melhor explicaram a variação total dos dados foram Al, Fe, Pb, Cl; Para PC2 (Tabela 49) K, Mg, Zn e Zr; Para PC3 (Tabela 49) Mg e para PC4 (Tabela 49) Na. Foram considerados significativos valores acima de 0,5. Todas as correlações significativas foram positivas.

Tabela 49- Composição dos fatores após a rotação dos eixos dos dados de solo coletados na bacia do Ribeirão Inhaúma, Iconha/ES

	Componente			
	1	2	3	4
Al	0,536	-0,487	0,442	0,113
Fe	0,899	-0,029	-0,137	-0,020
K	-0,458	0,737	0,159	-0,029
Mg	-0,041	0,131	0,935	0,027
Na	0,005	-0,166	0,069	0,897
Mn	-0,066	0,785	0,115	-0,294
Pb	0,848	-0,124	-0,034	-0,048
Zn	-0,122	0,880	-0,016	-0,012
Zr	0,376	0,534	-0,279	0,401
Cl	0,678	-0,140	0,116	0,151

Reunindo-se informações provenientes do mapa de solos elaborado pela Embrapa e da descrição de perfis na área de estudo, percebe-se que os solos encontrados na região são Latossolos, Gleissolos, Neossolos, Cambissolos e afloramentos de rocha. Com isso é possível inferir maiores características sobre os componentes principais estabelecidos, somado a informações da geologia do local, formada basicamente por metatonalitos.

O PC1 é representado pelas variáveis Al, Fe, Pb e Cl; Al e Fe são variáveis ligadas ao material de origem e conseqüente processo de intemperização; metatonalitos segundo Reis (2015) são caracterizados por teores de Al_2O_3 situados entre 14 e 19% e de Fe_2O_3 entre 2,8 e 7,5%; nas análises de termofluorescência realizadas, as amostras possuem valor médio de 24,75 % Al_2O_3 e 10,37 % de Fe_2O_3 . De acordo Kämpf & Curi (2000), os óxidos e hidróxidos de ferro e de alumínio, são relevantes indicadores pedogenéticos, pois sua formação é influenciada pelas condições do ambiente, além de persistirem por longos períodos no solo. Dentre os elementos considerados na análise de componentes principais, o Al e Fe são os mais abundantes nas amostras, enquanto a presença do Pb pode estar relacionada a alta sorção deste a óxidos de Fe, o que lhe confere uma correlação positiva com este.

Suarez e Langmuir (1976) mostraram que os óxidos de ferro e de manganês apresentam uma alta sorção de metais pesados. Segundo Santos (2005) o chumbo pode encontrar-se associado a minerais de argila e a hidróxidos de Fe e Al. O chumbo ocorre naturalmente em todos os solos, variando entre 1 a 200 mg/kg (SANTOS, 2005), como é o caso dos valores apresentados no resultados dessa análise. O Cl encontra-se no solo, em sua maior parte, em sais solúveis como $CaCl_2$ e $MgCl_2$, sua concentração varia de 20 a 900 mg/kg (ABREU et al, 2007), os valores encontrados neste trabalho foram de 400 mg/kg.

O PC2 (K, Mn, Zn, Zr) constitui o grupo que possui os menores valores entre aqueles envolvidos na análise, K entre os macronutrientes; Mn e Zn entre os micronutrientes e Zr entre os demais metais, já que o Pb foi inserido em outro componente principal. Os valores encontrados nos solos para a concentração de Mn e Zn são em média

600 mg/kg para o primeiro, e de 10 a 300 mg/kg para o segundo (KRAUSKOPF, 1972). Esses estão na mesma faixa dos valores médios encontrados que foram de 639,4 mg/kg para Mn e 119,0 mg/kg para Zn (Tabela 50). O Zn ocorre como elemento acessório em alguns minerais primários como olivina, hornblenda, biotita e magnetita, minerais esses que ocorrem tanto em rochas ácidas como básicas (SANTOS, 2005). Os metatonalitos são consideradas rochas intermediárias, com teores de SiO₂ de aproximadamente 58 a 67% (Pinho, 2003). Segundo Reis (2015), o Mn em algumas áreas pode ser originado da desagregação do piroxênio; que é um mineral encontrado em rochas como metatonalito e xisto; Esse conjunto de minerais pode também ter seus teores relacionados a adubações nas áreas de cultivo.

A PC3 e PC4 são compostas por um único elemento, sendo Mg e Na, respectivamente. Ambos estão dentre os elementos derivados dos metatonalitos (KRAUSKOF, 1967). Segundo Havlin et al, (1999) o magnésio na forma não trocável é encontrado em minerais primários e secundários, como biotita, hornblenda, olivina, dentre outros. O teor médio de Mg na crosta terrestre gira em torno de 19 g/kg, e varia segundo a origem geológica do solo (DECHEN e NACHTIGALL, 2007), nas amostras foi encontrado um valor médio de 639,38 mg/kg. O Mg presente nas amostras pode ter o efeito das calagens realizadas nas áreas de agricultura e pastagem.

O metatonalito possui cerca de 3 a 5,7% de Na₂O (REIS 2015), o que é corroborado por Krauskof (1967), que cita ser está fonte de alguns elementos inclusive sódio. Segundo Dechen e Nachtigall (2007), o Na constitui 2,8% da crosta terrestre, e encontra-se no solo como cátion monovalente, adsorvido aos colóides de argila, sendo capaz, quando em altas quantidades, de deslocar Ca e K.

Tabela 50- Teores médios de nutrientes no solo, na camada 0-10 cm obtidos por termofluorescência

	Valores médios		Valores médios
Al ₂ O ₃ (%)	24,75	MnO (ppm)	639,38
CaO (%)	0,38	NiO (ppm)	17,79
Fe ₂ O ₃ (%)	10,37	P ₂ O ₅ (ppm)	2503,77
K ₂ O (%)	1,26	PbO (ppm)	203,19
MgO (%)	1,34	V ₂ O ₅ (ppm)	417,78
Na ₂ O (%)	1,71	ZnO (ppm)	119,01
SiO ₂ (%)	34,15	ZrO ₂ (ppm)	770,61
TiO ₂ (%)	2,12	SO ₃ (%)	0,12
Cr ₂ O ₃ (ppm)	266,55	Cl (%)	0,40
CuO (ppm)	93,56		

A análise de correlação de Pearson é apresentada na Tabela 51. A partir dela selecionou-se àquelas que obtiveram coeficiente maior que 0,5. Dessa maneira observou-se correlação positiva entre Zn com P e K; Mn com Ca; P, V e Zr com Ti; V com P e Zr;

Th com kperc; Uthra com V; Ukraz com Thkra e Uthra e Temp com Th. E correlação negativa entre Zn e Al; Si com Fe e PB; Th com Ukraz e Uthra.

Tanto o Kperc obtido através dos dados da CPRM, quanto o K obtido pela análise de termofluorescência representam valores totais no elemento no solo, dessa forma esperava-se que houvesse correlação entre eles, porém de acordo com a correlação de Pearson, o coeficiente foi de apenas 0,11.

A fim de checar a existência de mais correlações entre as variáveis, procedeu-se o teste de correlação não linear (Tabela 52). Essa apresentou apenas correlações positivas, com coeficientes mais elevados.

Os gráficos de dispersão (Figura 73), mostram as correlações não lineares maiores que 0,8.

Tabela 51- Correlação de Pearson entre dados obtidos por termofluorescência de raios-x, dados gamaespectrométricos e temperatura

	Al	Ca	Fe	K	Mg	Na	Si	Ti	Cr	Cu	Mn	Ni	P	Pb	V	Zn	Zr	S	Cl	Th	U	Kperc	Th/K	U/K	U/th	T
Al	1,00	-0,46	0,42	-0,49	0,22	0,23	-0,37	-0,12	0,17	0,21	-0,45	-0,11	-0,31	0,45	-0,13	-0,50	-0,10	-0,45	0,38	-0,17	0,01	-0,36	0,36	0,39	0,15	0,07
Ca	-0,46	1,00	-0,24	0,27	-0,26	-0,09	0,06	-0,01	-0,13	0,02	0,60	0,03	0,56	-0,22	-0,02	0,46	-0,01	0,12	-0,29	0,09	0,02	0,28	-0,22	-0,24	-0,10	-0,08
Fe	0,42	-0,24	1,00	-0,45	-0,15	0,09	-0,70	0,62	0,21	0,00	-0,17	-0,23	0,26	0,75	0,57	-0,13	0,16	-0,30	0,48	0,23	0,22	0,08	0,20	0,23	0,04	0,25
K	-0,49	0,27	-0,45	1,00	0,23	-0,28	0,44	0,05	-0,18	-0,12	0,41	0,15	0,12	-0,49	0,09	0,60	0,27	-0,02	-0,37	0,12	-0,15	0,11	-0,03	-0,26	-0,25	-0,30
Mg	0,22	-0,26	-0,15	0,23	1,00	0,07	0,24	-0,23	0,07	0,15	-0,05	0,23	-0,43	-0,11	-0,21	0,08	-0,12	-0,08	0,03	-0,09	-0,07	-0,12	0,07	0,02	0,03	-0,06
Na	0,23	-0,09	0,09	-0,28	0,07	1,00	0,03	0,01	0,11	-0,01	-0,14	0,00	-0,14	0,12	0,01	-0,23	0,00	-0,04	0,10	-0,06	0,08	-0,12	0,11	0,11	0,07	0,03
Si	-0,37	0,06	-0,70	0,44	0,24	0,03	1,00	-0,31	-0,09	0,14	0,00	0,23	-0,29	-0,56	-0,26	0,12	-0,01	0,00	-0,27	-0,21	-0,21	-0,17	-0,07	-0,11	0,00	-0,32
Ti	-0,12	-0,01	0,62	0,05	-0,23	0,01	-0,31	1,00	-0,10	-0,08	0,00	-0,13	0,55	0,44	0,99	0,30	0,64	-0,29	0,25	0,39	0,29	0,22	0,10	0,06	-0,11	0,18
Cr	0,17	-0,13	0,21	-0,18	0,07	0,11	-0,09	-0,10	1,00	0,15	-0,07	0,47	-0,09	0,24	-0,12	-0,14	-0,32	0,14	0,05	0,10	0,12	0,10	0,02	0,08	0,06	0,13
Cu	0,21	0,02	0,00	-0,12	0,15	-0,01	0,14	-0,08	0,15	1,00	0,10	0,22	-0,04	0,10	-0,08	0,13	-0,08	-0,22	0,12	-0,17	0,02	-0,12	-0,01	0,15	0,20	-0,05
Mn	-0,45	0,60	-0,17	0,41	-0,05	-0,14	0,00	0,00	-0,07	0,10	1,00	0,09	0,45	-0,21	-0,01	0,70	-0,05	0,06	-0,23	-0,07	0,02	0,15	-0,16	-0,11	0,03	-0,21
Ni	-0,11	0,03	-0,23	0,15	0,23	0,00	0,23	-0,13	0,47	0,22	0,09	1,00	-0,02	-0,10	-0,10	0,21	-0,08	0,05	-0,04	-0,13	0,01	-0,02	-0,18	0,03	0,16	-0,12
P	-0,31	0,56	0,26	0,12	-0,43	-0,14	-0,29	0,55	-0,09	-0,04	0,45	-0,02	1,00	0,15	0,55	0,56	0,22	-0,15	-0,02	0,27	0,16	0,25	-0,02	0,00	-0,10	0,05
Pb	0,45	-0,22	0,75	-0,49	-0,11	0,12	-0,56	0,44	0,24	0,10	-0,21	-0,10	0,15	1,00	0,41	-0,23	0,12	-0,26	0,40	0,19	0,26	0,07	0,16	0,23	0,07	0,25
V	-0,13	-0,02	0,57	0,09	-0,21	0,01	-0,26	0,99	-0,12	-0,08	-0,01	-0,10	0,55	0,41	1,00	0,33	0,66	-0,31	0,23	0,38	0,26	0,20	0,11	0,06	-0,11	0,15
Zn	-0,50	0,46	-0,13	0,60	0,08	-0,23	0,12	0,30	-0,14	0,13	0,70	0,21	0,56	-0,23	0,33	1,00	0,18	-0,06	-0,21	0,00	0,05	0,15	-0,13	-0,06	0,03	-0,26
Zr	-0,10	-0,01	0,16	0,27	-0,12	0,00	-0,01	0,64	-0,32	-0,08	-0,05	-0,08	0,22	0,12	0,66	0,18	1,00	-0,19	0,08	0,25	0,14	0,05	0,07	-0,03	-0,13	0,06
S	-0,45	0,12	-0,30	-0,02	-0,08	-0,04	0,00	-0,29	0,14	-0,22	0,06	0,05	-0,15	-0,26	-0,31	-0,06	-0,19	1,00	-0,23	0,10	0,08	0,23	-0,37	-0,30	-0,01	0,17
Cl	0,38	-0,29	0,48	-0,37	0,03	0,10	-0,27	0,25	0,05	0,12	-0,23	-0,04	-0,02	0,40	0,23	-0,21	0,08	-0,23	1,00	0,06	-0,03	-0,03	0,08	0,10	0,00	0,14
Th	-0,17	0,09	0,23	0,12	-0,09	-0,06	-0,21	0,39	0,10	-0,17	-0,07	-0,13	0,27	0,19	0,38	0,00	0,25	0,10	0,06	1,00	0,27	0,76	-0,06	-0,52	-0,60	0,55
U	0,01	0,02	0,22	-0,15	-0,07	0,08	-0,21	0,29	0,12	0,02	0,02	0,01	0,16	0,26	0,26	0,05	0,14	0,08	-0,03	0,27	1,00	0,39	-0,26	0,23	0,50	0,16
Kperc	-0,36	0,28	0,08	0,11	-0,12	-0,12	-0,17	0,22	0,10	-0,12	0,15	-0,02	0,25	0,07	0,20	0,15	0,05	0,23	-0,03	0,76	0,39	1,00	-0,56	-0,68	-0,30	0,38
Th/K	0,36	-0,22	0,20	-0,03	0,07	0,11	-0,07	0,10	0,02	-0,01	-0,16	-0,18	-0,02	0,16	0,11	-0,13	0,07	-0,37	0,08	-0,06	-0,26	-0,56	1,00	0,58	-0,26	-0,02
U/K	0,39	-0,24	0,23	-0,26	0,02	0,11	-0,11	0,06	0,08	0,15	-0,11	0,03	0,00	0,23	0,06	-0,06	-0,03	-0,30	0,10	-0,52	0,23	-0,68	0,58	1,00	0,60	-0,28
U/th	0,15	-0,10	0,04	-0,25	0,03	0,07	0,00	-0,11	0,06	0,20	0,03	0,16	-0,10	0,07	-0,11	0,03	-0,13	-0,01	0,00	-0,60	0,50	-0,30	-0,26	0,60	1,00	-0,33
T	0,07	-0,08	0,25	-0,30	-0,06	0,03	-0,32	0,18	0,13	-0,05	-0,21	-0,12	0,05	0,25	0,15	-0,26	0,06	0,17	0,14	0,55	0,16	0,38	-0,02	-0,28	-0,33	1,00

Th – Tório; U-Urânio; Kperc-K obtido da gamaespectrometria; Th/K- Razão Tório e K; U/K- Razão Urânio e Potássio; U/Th- Razão Urânio e Tório; T- temperatura

Tabela 52- Correlação não linear- RDC entre dados obtidos por termofluorescência de raios-x, dados gamaespectrométricos e temperatura

	Al	Ca	Fe	K	Mg	Na	Si	Ti	Cr	Cu	Mn	Ni	P	Pb	V	Zn	Zr	S	Cl	Th	U	Kperc	Th/K	U/K	U/Th	T
Al	1,00	0,77	0,48	0,70	0,34	0,35	0,41	0,54	0,31	0,25	0,57	0,28	0,31	0,47	0,54	0,69	0,32	0,57	0,41	0,31	0,38	0,39	0,45	0,53	0,32	0,21
Ca	0,77	1,00	0,46	0,87	0,23	0,29	0,39	0,31	0,27	0,23	0,68	0,24	0,35	0,53	0,33	0,77	0,34	0,37	0,44	0,14	0,36	0,32	0,34	0,31	0,24	0,24
Fe	0,45	0,45	1,00	0,44	0,24	0,19	0,73	0,73	0,31	0,21	0,44	0,33	0,57	0,77	0,71	0,42	0,40	0,50	0,51	0,42	0,32	0,38	0,35	0,39	0,27	0,40
K	0,70	0,87	0,47	1,00	0,33	0,36	0,50	0,38	0,37	0,32	0,75	0,26	0,34	0,50	0,40	0,86	0,35	0,52	0,42	0,34	0,41	0,30	0,37	0,36	0,41	0,34
Mg	0,36	0,25	0,24	0,30	1,00	0,33	0,35	0,38	0,28	0,25	0,22	0,33	0,46	0,19	0,34	0,18	0,30	0,32	0,24	0,39	0,31	0,49	0,44	0,36	0,24	0,28
Na	0,34	0,30	0,23	0,31	0,33	1,00	0,28	0,29	0,26	0,24	0,41	0,33	0,46	0,23	0,30	0,34	0,35	0,41	0,27	0,34	0,25	0,37	0,27	0,30	0,34	0,28
Si	0,39	0,40	0,73	0,50	0,36	0,29	1,00	0,46	0,25	0,27	0,41	0,31	0,50	0,65	0,43	0,52	0,24	0,26	0,38	0,38	0,30	0,24	0,29	0,32	0,29	0,54
Ti	0,54	0,31	0,72	0,38	0,37	0,25	0,47	1,00	0,32	0,23	0,43	0,21	0,76	0,51	0,99	0,43	0,72	0,35	0,36	0,57	0,39	0,38	0,29	0,36	0,42	0,32
Cr	0,30	0,27	0,30	0,41	0,30	0,31	0,27	0,31	1,00	0,37	0,18	0,59	0,30	0,31	0,33	0,31	0,40	0,41	0,25	0,31	0,46	0,22	0,27	0,24	0,33	0,34
Cu	0,25	0,22	0,23	0,31	0,26	0,33	0,26	0,22	0,41	1,00	0,38	0,36	0,28	0,26	0,25	0,23	0,28	0,41	0,27	0,34	0,27	0,27	0,21	0,23	0,27	0,23
Mn	0,58	0,68	0,42	0,75	0,28	0,38	0,41	0,40	0,23	0,32	1,00	0,34	0,49	0,30	0,41	0,81	0,29	0,35	0,32	0,18	0,28	0,23	0,33	0,27	0,22	0,27
Ni	0,26	0,21	0,33	0,26	0,33	0,23	0,31	0,22	0,60	0,32	0,34	1,00	0,27	0,23	0,18	0,24	0,22	0,20	0,46	0,31	0,19	0,29	0,28	0,28	0,27	0,29
P	0,32	0,39	0,57	0,35	0,48	0,34	0,52	0,76	0,31	0,31	0,49	0,26	1,00	0,40	0,76	0,61	0,58	0,45	0,28	0,49	0,31	0,37	0,28	0,39	0,26	0,26
Pb	0,47	0,52	0,76	0,50	0,18	0,26	0,65	0,51	0,31	0,30	0,33	0,26	0,41	1,00	0,49	0,36	0,21	0,45	0,49	0,27	0,48	0,31	0,25	0,37	0,39	0,36
V	0,55	0,35	0,70	0,39	0,33	0,22	0,44	0,99	0,35	0,29	0,46	0,22	0,75	0,49	1,00	0,47	0,73	0,37	0,35	0,57	0,36	0,37	0,29	0,38	0,43	0,34
Zn	0,70	0,77	0,43	0,86	0,35	0,28	0,56	0,43	0,31	0,30	0,81	0,23	0,61	0,38	0,46	1,00	0,34	0,38	0,39	0,19	0,34	0,27	0,37	0,33	0,23	0,41
Zr	0,29	0,30	0,39	0,42	0,33	0,30	0,23	0,70	0,39	0,39	0,32	0,20	0,56	0,28	0,72	0,35	1,00	0,23	0,36	0,46	0,34	0,29	0,43	0,26	0,28	0,30
S	0,58	0,40	0,47	0,45	0,34	0,34	0,25	0,35	0,40	0,39	0,39	0,21	0,45	0,39	0,36	0,45	0,27	1,00	0,33	0,44	0,35	0,45	0,43	0,34	0,24	0,26
Cl	0,41	0,42	0,50	0,40	0,25	0,26	0,39	0,37	0,24	0,34	0,32	0,47	0,31	0,49	0,35	0,35	0,31	0,34	1,00	0,29	0,26	0,24	0,28	0,28	0,27	0,32
Th	0,29	0,12	0,33	0,39	0,45	0,18	0,35	0,58	0,37	0,31	0,19	0,34	0,50	0,29	0,59	0,26	0,51	0,43	0,29	1,00	0,39	0,85	0,54	0,56	0,68	0,61
U	0,31	0,30	0,32	0,39	0,30	0,38	0,34	0,39	0,46	0,27	0,27	0,20	0,29	0,45	0,36	0,32	0,35	0,33	0,35	0,40	1,00	0,54	0,37	0,43	0,60	0,29
Kperc	0,41	0,25	0,40	0,31	0,35	0,36	0,25	0,41	0,16	0,27	0,22	0,26	0,29	0,31	0,40	0,23	0,39	0,49	0,21	0,85	0,54	1,00	0,81	0,81	0,56	0,47
Th/K	0,40	0,32	0,37	0,38	0,44	0,23	0,29	0,39	0,22	0,25	0,32	0,27	0,28	0,27	0,42	0,32	0,44	0,45	0,26	0,54	0,49	0,80	1,00	0,68	0,42	0,37
U/K	0,51	0,32	0,39	0,36	0,37	0,33	0,32	0,38	0,30	0,23	0,24	0,28	0,38	0,38	0,41	0,26	0,24	0,35	0,27	0,56	0,45	0,81	0,68	1,00	0,74	0,44
U/Th	0,32	0,17	0,24	0,40	0,26	0,33	0,20	0,39	0,29	0,23	0,33	0,29	0,27	0,39	0,41	0,21	0,28	0,29	0,27	0,69	0,63	0,52	0,42	0,73	1,00	0,40
T	0,22	0,26	0,39	0,34	0,36	0,29	0,54	0,32	0,33	0,20	0,26	0,32	0,35	0,35	0,33	0,44	0,30	0,31	0,32	0,58	0,29	0,48	0,34	0,38	0,38	1,00

Th – Tório; U-Urânio; Kperc-K obtido da gamaespectrometria; Th/K- Razão Tório e K; U/K- Razão Urânio e Potássio; U/Th- Razão Urânio e Tório; T- temperatura

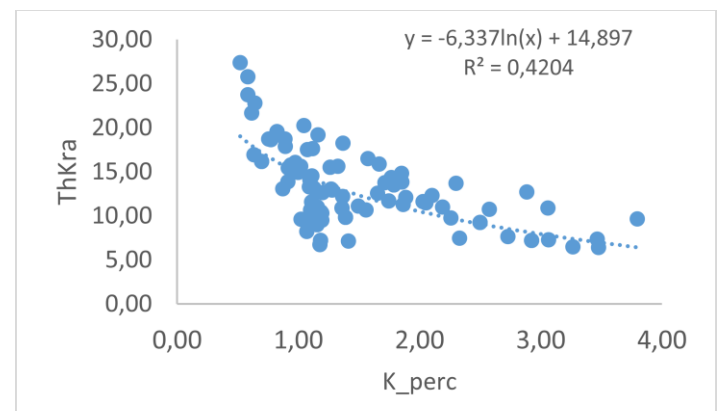
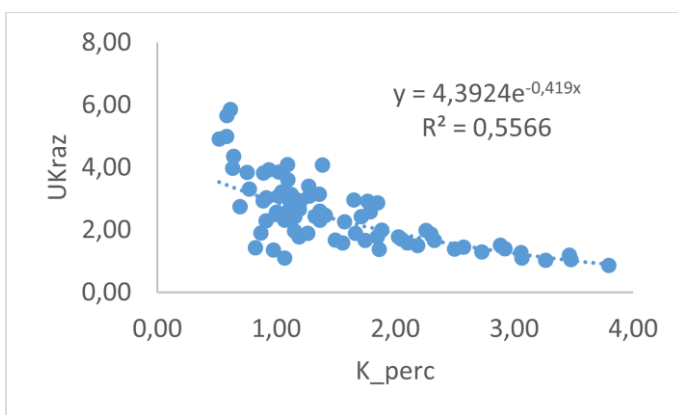
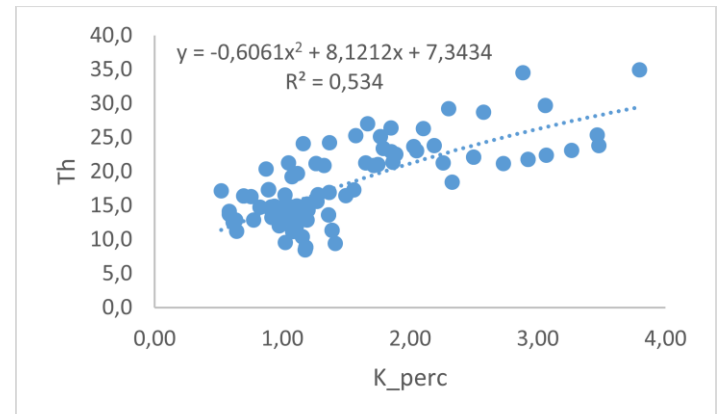
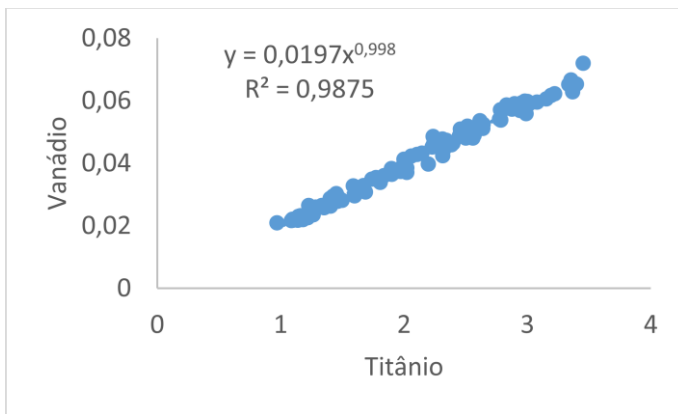
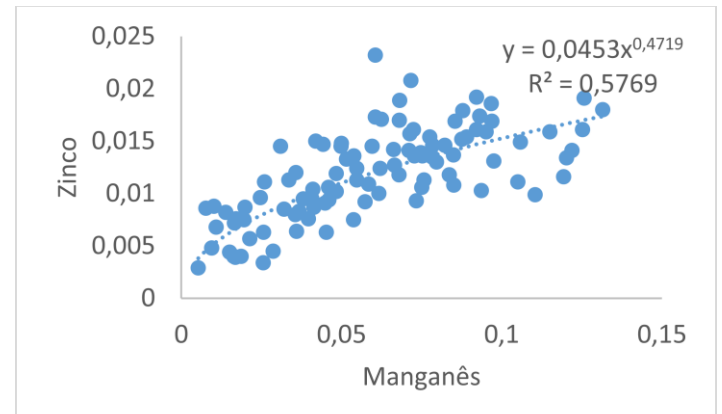
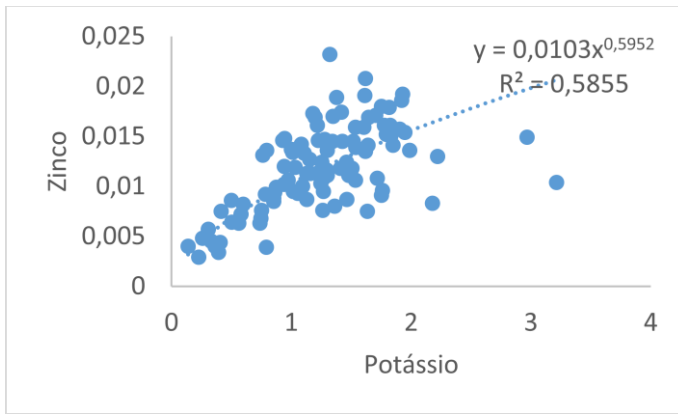
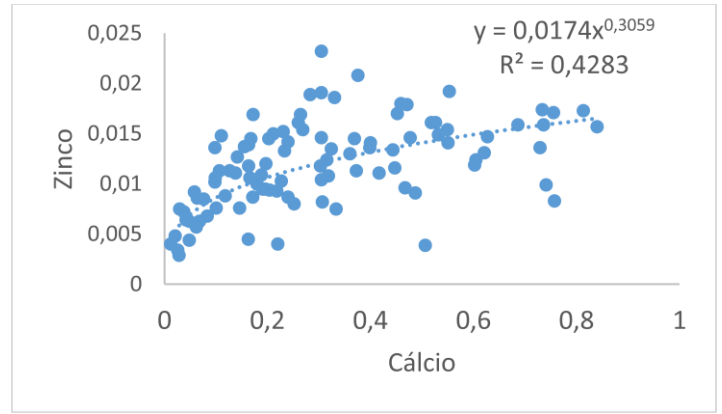
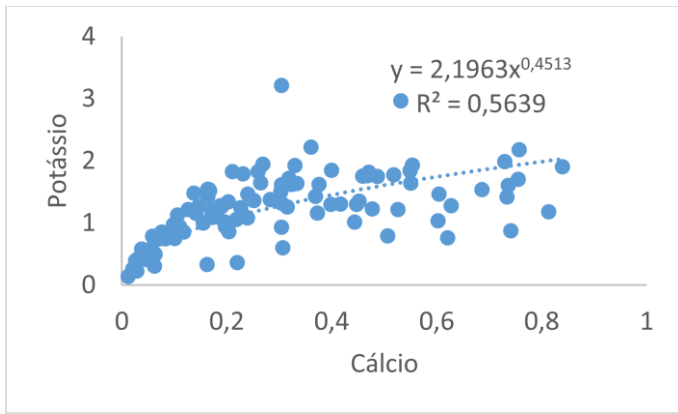


Figura 73- Gráficos de dispersão para correlações maiores que 0,8.

K_perc – Potássio medido por gamaespectrômetro; Th – Tório; UKraz- Razão Urânio e Tório; ThKra – Razão Tório e Potássio

O Vanádio ocorre, geralmente, em depósitos de sulfetos, associado a Pb, Zn, Cu e Mn (WEDEPOHL 1978); é conhecido por formar complexos catiônicos e aniônicos, apresentando vários comportamentos; podendo ser carregado por minerais titaníferos. As concentrações de Ti encontradas nos solos variam de 1.000 a 10.000 mg/kg (KABATA - PENDIAS & PENDIAS 2001). São encontradas na literatura a associação de Ti e V principalmente em depósitos minerais (Fischer, 1975; Taylor et al., 2006). Os minerais de Ti são muito resistentes a intempéries enquanto o V é distribuído em perfis de solo de forma bastante uniforme e a variação no conteúdo de V do solo é herdada dos materiais originais (KABATA-PENDIAS 2010). Assim, concentrações mais altas de V (até 500 mg / kg) são relatadas para Cambissolos (KABATA-PENDIAS 2010).

Zinco apresentou alta correlação com Ca, K e Mn, essas também mostraram-se altas pela correlação de Pearson, sendo todas positivas, indicando que onde houve um incremento da concentração de Zn, houve também um aumento nos valores de Ca, K e Mn . K também apresentou correlação com Ca, sendo que a correlação linear foi positiva, porém baixa (0,27); enquanto a não linear apresentou valores de 0,87.

Conforme discutido anteriormente o Th e K, apresentaram tendências de distribuição e comportamento semelhante na área estudada, enfatizando a correlação existente entre estes. O K apresentou correlação com a razão entre Tório e K, e Urânio e K; justamente devido a sua participação nessas razões.

5 CONCLUSÕES

1. Foi possível identificar a correlação existente entre o uso e ocupação dos solos com os dados de gamaespectrometria e de temperatura, indicando que o uso e ocupação do solo interfere na concentração de elementos e na temperatura do solo, o que influencia diretamente no estado de conservação deste.
2. Em relação aos dados aerogeofísicos de gamaespectrometria foram encontradas altas correlações com altitude e baixa correlação com dados obtidos pela análise de termofluorescência, indicando que dados provenientes da análise de termofluorescência não podem substituir àqueles advindos da gamaespectrometria e vice-versa, para fins de análise de características do terreno.
3. Análises de termofluorescência e de gamaespectrometria não são eficientes para determinação de K no solo, não podendo servir como base para recomendações que visem a adição desse nutriente ao solo.

6 REFERÊNCIAS BIBLIOGRÁFICAS

ABDEL-HALEEM, A.S.; SCROON, A.; EL-BAHI, S.M.; ZOHNY, E. Heavy metals and rare earth elements in phosphate fertilizer components using instrumental neutron activation analysis. Applied Radiation and Isotopes, v.55, p.569-573, 2001.

- ABREU, C. A. DE; LOPES, S.L.; SANTOS, G. Micronutrientes. In: NOVAIS, R. F.; ALVAREZ V. V. H.; Barros, N.F.; Fontes, R.L.F.; Cantarutti, R.B.; Neves, J.C. (Ed.) Fertilidade do Solo. Viçosa. Sociedade Brasileira de Ciência do Solo. v. 1, p.645-736, 2007.
- ADAMS, J. A. S.; GASPARINI, P. Gamma-ray spectrometry of rocks. 1 ed. New York: Elsevier, 1970.
- ADAMS, J. A. S.; GASPARINI, P. Gamma-ray spectrometry of rocks. 1 ed. New York: Elsevier, 280 p., 1970.
- ALLOWAY, B.J. Heavy metals in soils. 2 ed. London: Blackie Academic. 368 p. 1995.
- AMORIM, M.C. DE C.T.;MONTEIRO,A. As temperaturas intraurbanas: exemplos do Brasil e de Portugal , *Confins* [Online], v.13, 2011. Disponível em < <http://confins.revues.org/7284> > Acesso em Maio, 2016.
- BARBOSA, J.S.F. & FONTEILLES, M. Caracterização dos protólitos da região granulítica do sul da Bahia. *Rev.Bras.Geoc.*, v.19, p.3-12. 1989.
- BEA, F. Residence of REE, Y, Th and U in granites and crustal protoliths; implications for the chemistry of crustal melts. *Journal of Petrology*. v. 37, p.521-552, 1996.
- BECEGATO VA, FERREIRA FJF. Gamaespectrometria, resistividade elétrica e susceptibilidade magnética de solos agrícolas no nordeste do estado do Paraná. *Rev Bras de Geofísica*. 2005; 23: 371-405.
- BECEGATO, V.A; FERREIRA, F.J.F. Gamaespectrometria, resistividade elétrica e susceptibilidade magnética de solos agrícolas no nordeste do estado do Paraná. *Rev Bras de Geofísica*. v.23, p. 371-405, 2005.
- BÖGUEL NETO, G. Análise da Confiabilidade do ensaio de identificação de ligas através de espectrometria por fluorescência de Raios X. Trabalho de conclusão de curso. Guaratinguetá: Universidade Estadual Paulista. 50 p., 2012.
- BONOTTO D.M. Radioatividade nas águas: da Inglaterra ao Guarani.1nd Ed. São Paulo: Editora UNESP; 2004.
- CAMPOS, M.L; SILVA, F.N. DA; FURTINI NETO, A.E.; GUILHERME, L.R.G.; MARQUES, J.J.; ANTUNES, A.S. Determinação de cádmio, cobre, cromo, níquel, chumbo e zinco em fosfatos de rocha. *Pesq. agropec. bras.*, Brasília, v.40, n.4, p.361-367,2005.
- DECHEN, A. R. E NACHTIGALL, G.R. Elementos requeridos à nutrição de plantas. In: ALVAREZ V. V. H.; BARROS, N.F.; FONTES, R.L.F.; CANTARUTTI, R.B.; NEVES, J.C. (Ed.) Fertilidade do Solo. Viçosa. Sociedade Brasileira de Ciência do Solo. v. 1, p.92-132, 2007.
- Dickson BL e Scott KM. Interpretation of aerial gamma-ray surveys – adding the geochemical factors. *AGSO Journal of Australian Geology & Geophysics*. 1997; 17: 187–200
- DICKSON, B.L.; E SCOTT, K.M. Interpretation of aerial gamma-ray surveys – adding the geochemical factors. *AGSO Journal of Australian Geology & Geophysics*. v.17, p. 187–200,1997.

- FERREIRA, F.J.F.; CANDIDO, A.G; ROSTIROLLA, S.P. Correlação gamaesctrométrica de afloramentos e poços: Estudo de caso na formação Ponta Grossa (Bacia do Paraná, Brasil). v.28, p. 371-396, 2010.
- FISCHER, R.P. Vanadium Resources in Titaniferous Magnetite Deposits. Geological survey professional paper 926-B. U.S. Government Printing Office. v. 18; p.677-340, 1975.
- GALLO, K.; XIAN, G. Application of Spatially Gridded Temperature and Land Cover Data Sets For Urban Heat Island Analysis. *Urban Climate*, v. 8, n. 1, p. 1-10, 2014.
- HAVLIN, J.L.; BEATON, J.D.; TISDALE, S.L.; NELSON, W.L. Soil fertility and fertilizers: An introduction to nutrient management. 6 ed. New Jersey, Prentice Hall, 499 p. 1999.
- IBGE - EMBRAPA - Mapa de Solos do Brasil. Rio de Janeiro: IBGE, 2001.
- JENSEN, J. R. Introductory Digital Image Processing: A Remote Sensing Perspective. Englewood Cliffs, New Jersey: Prentice-Hall, 318 p., 1996.
- JOHNSON, RA & WICHERN, DW Aplicada análise estatística multivariada. 5 ed. New Jersey, Prentice Hall, 767 p. , 2002.
- KABATA-PENDIAS, A.; PENDIAS, H. Trace elements in soil and plants. 3 ed. Boca Raton: CRC Press, 331 p. , 2001.
- KAISER, H. F. The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, v 20, p 141-151, 1960.
- KAMPF, N.; CURI, N. Óxidos de ferro: indicadores de atributos de ambientes pedogenéticos e geoquímicos. In: NOVAIS, R. F.; ALVAREZ V.. V. H.; SCHAEFER, C. E. R. (Ed.) Tópicos em Ciência do Solo. Viçosa. Sociedade Brasileira de Ciência do Solo. v. 1, p.107-138, 2000.
- KRAUSKOF, K. B. Geochemistry of micronutrients. *In* Micronutrients in agriculture (J.J. Mortvedt, P.M. Giordano & W.L Lindsay, eds.). Soil Science Society of America, Madison. p.7-40. 1972.
- KRAUSKOF, K. B. Introduction to geochemistry. New York: McGraw-Hill, 1967.
- LANDIS, J. R. & KOCH, G. G. The measurement of observer agreement for categorical data. *Biometrics*, v. 33, p. 159-174, 1997.
- LOPEZ-PAZ, D.; HENNIG, P; SCHOLKOPF, B. The Randomized Dependence Coefficient. 2013. Disponível em < <https://papers.nips.cc/paper/5138-the-randomized-dependence-coefficient.pdf>>; Acesso em Fev, 2016.
- MINTY BRS. Fundamentals of airborne gamma-ray spectrometry. *AGSO Journal of Australian Geology & Geophysics*. v. 17, p. 39-50, 1997.
- OLIVEIRA, D.P. de. Gênese de Cambissolos desenvolvidos de rocha calcária na Chapada do Apodi – CE. Dissertação (mestrado) – Universidade Federal do Ceará-UFC, Fortaleza, 103 p. 2012.
- PFISTER H, PHILIPP G & PAUL YH. Population dose from natural radionuclides in phosphate fertilizers. *Radiation and Environmental Biophysics*. v.13, p.247-261,1976.

- PINHO I.C.A. *Os tonalitos/trondhjemitos das regiões de Ipiaú- Ubaitaba-Ibicaraí-Itabuna, Bahia*. Dissertação de Mestrado. Instituto de Geociências, Universidade Federal da Bahia, Bahia, 193 p. 2000.
- PINHO, I.C.DE A.; BARBOSA, J.S.F.; LEITE, C.M.M. Petrografia e litogeoquímica dos metatonalitos e seus enclaves de granulitos básicos da parte sul do Cinturão Itabuna - Salvador-Curaçá, Bahia.. *Revista Brasileira de Geociências* . 33(2):189-200, junho de 2003.
- PINHO,I.C.A. *Geologia dos Metatonalitos/Metatrandjemitos e granulitos básicos das regiões de Camamu-Ubaitaba-Itabuna,Bahia*. Tese (Doutorado). Salvador: Universidade Federal da Bahia, 176 p., 2005.
- POTTS P; WEST, M. *Portable X-ray Fluorescence Spectrometry: Capabilities for In Situ Analysis*. 1 ed. Cambridge: RSCPublishing, 2008.
- QUEIRÓS, I.B. *Contribuição da fluorescência de Raios-X na avaliação e modelação de Georrecursos*. Dissertação (Mestrado). Porto: Instituto Superior de Engenharia do Porto; 2011.
- REIS, D. A. dos. *Estudo ambiental sobre a bacia hidrográfica do Rio Matipó*. Dissertação (Mestrado). Universidade Federal de Ouro Preto-UFOP. Ouro Preto. 171 p. , 2015.
- RIGHI S, BETTI M, BRUZZI L & MAZZOTTI G. *Monitoring of natural radioactivity in working places*. *Microchemical Journal*. v. 67, p. 119–126 . 2000.
- SANTOS ACL, MENEZES PTL, NASCIMENTO CTC. *Gamaespectrometria aplicada em solo agrícola no noroeste do estado do Rio de Janeiro*. *Rev Bras de Geofísica*. v. 26, p. 181-194. 2008.
- SANTOS, A. L.C.; SANTOS, F. *Mapeamento das classes de uso e cobertura do solo da bacia hidrográfica do rio Vaza-Barris, Sergipe*. *Revista Multidisciplinar da Uniesp: Saber Acadêmico*. v. 10, p. 57-67, 2010.
- SANTOS, G.C.G. dos. *Comportamento de B, Zn, Cu, Mn, PB em solo contaminado sob cultivo de plantas e adição de fontes de matéria orgânica como amenizantes do efeito tóxico*. Tese (Doutorado). Escola Superior de Agricultura Luiz de Queiróz- ESALQ. Piracicaba. 2005.
- SAPUCAIA, N.S; ARGOLLO, R.M; BARBOSA, J.S.F. *Teores de potássio, uranio, tório e taxa de produção de calor radiogênico no embasamento adjacente` as bacias sedimentares de Camamu e Almada, Bahia, Brasil*. *Revista Brasileira de Geofísica*. v. 23, p. 453-47, 2005.
- SILVA S.M.P, CROSTA A.P, FERREIRA H.B, SILVA A.M, SILVA M.R.R. *Identificação gamaespectrometrica de placeres rutilo-monazíticos neoproterozóicos no sul da faixa Seridó, nordeste do Brasil*. *Rev Bras de Geofísica*. v. 28, p.61-77, 2010.
- SILVA, S. DE A.; LIMA, J.S DE. *Multivariate analysis and Geostatistics of the fertility of a humic rhodic hapludox under coffee cultivation*. *R. Bras. Ci. Solo*. v.36, p.467-474, 2012.
- SILVA; L.C.; CUNHA, J.M.; MACHADO, N.G.; CAMPOS, M.C.C.; BIUDES, M.S. *Estimativa do balanço de radiação por sensoriamento remoto de diferentes usos de solo no sudoeste da Amazônia brasileira*. *Sociedade & Natureza (UFU. Online)*, v. 27, p. 341-356, 2015.

SOBRINO, J. A.; JIMÉNEZ-MUÑOZ, J.C. Minimum configuration of thermal infrared bands for land surface temperature and emissivity estimation in the context of potential future missions. *Remote Sensing of Environment* . v. 148, p. 158-167, 2014.

SOUSA E FERREIRA JUNIOR. Relação entre temperatura de superfície terrestre, índices espectrais e classes de cobertura da terra no município de Goiânia (GO). *Raega: O espaço geográfico em análise*. v. 26, p 75-79, 2012.

SOUZA, J. F. DE; SILVA, R. M.; SILVA, A. M. Influência do uso e ocupação do solo na temperatura da superfície: o estudo de caso de João Pessoa – PB. *Ambiente Construído*, Porto Alegre, v. 16, n. 1, p. 21-37, 2016.

STRECK, E.V.; KÄMPF, N.; DALMOLIN, R.S.D.; KLAMT, E.; NASCIMENTO, P.C.; SCHNEIDER, P.; GIASSON, E. & PINTO, L.F.S. *Solos do Rio Grande do Sul*. 2.ed. Porto Alegre, Universidade federal do Rio Grande do Sul. 222p. 2008.

SUAREZ, D.L.; LANGMUIR, D. Heavy metal relationships in a Pennsylvania soil. *Geochim. Cosmochim. Acta*, v. 40, p. 589-598, 1976.

TAYLOR, P.R; SHUEY, S A; VIDAL, E E; GOMEZ, J C. Extractive Metallurgy of Vanadium-Containing Titaniferous Magnetite Ores: a Review. *Redorbit*. 2006. Disponível em < http://www.redorbit.com/news/entertainment/528669/extractive_metallurgy_of_vanadiumcontaining_titaniferous_magnetite_ores_a_review/#g3qDoX9JpcoXwJkO.99 > Acesso em Mai, 2016.

ULBRICH, H.H.G.J.; ULBRICH, M.N.C.; FERREIRA, F.J.F; ALVES, L.S., GUIMARÃES, G.B., FRUCHTING, A. Levantamentos Gamaespectrométricos em Granitos Diferenciados. I: Revisão da Metodologia e do Comportamento Geoquímico dos Elementos K, Th e U. *Geol. USP Sér. Cient.* v. 9, p.33-53, 2009.

WEDEPOHL K. H. *Handbook of Geochemistry*. Springer Verlag, Berlin, Heidelberg, v. II. 1978.

WEIJDEN CH & WEIJDEN RD. Mobility of major, minor and some redox-sensitive trace elements and rare-earth elements during weathering of four granitoids in central Portugal. *Chemical Geology*. v. 125, p.149–167, 1995

WILFORD JR., BIERWIRTH PN.; CRAIG MA. Application of airborne gamma-ray spectrometry in soil/regolith mapping and applied geomorphology. *AGSO Journal of Australian Geology and Geophysics*. v. 2, p. 201-216, 1997.

YOUNG, A.F. Aplicação de índices relativos de vegetação e temperatura para estudo das mudanças do uso e ocupação do solo: estudo de caso de Curitiba (PR), 1986 A 2002. Tese (Doutorado). Universidade Estadual De Campinas-UNICAMP. 112p. 2005.

CAPÍTULO VI

CORRELAÇÃO DE TEORES DE SILÍCIO, ALUMÍNIO, TITÂNIO E FERRO OBTIDOS POR ATAQUE TOTAL E FLUORESCÊNCIA DE RAIOS-X

RESUMO

Os teores de SiO_2 , Al_2O_3 , TiO_2 e Fe_2O_3 encontrados no solo, ajudam a determinar diferentes fatores como o grau de intemperização e desenvolvimento destes. Tradicionalmente, a análise dos teores desses óxidos é feita em laboratório pelo método denominado ataque total, porém este é um método dispensioso quando comparado a outras opções como a análise de fluorescência. Dessa forma, seria interessante um estudo que permitisse avaliar e comparar os resultados de ambos os métodos, para se analisar alternativas para laboratórios de rotina que trabalham com grande volume de análises diárias, facilitando assim este serviço. Sendo assim, este estudo propõe um estudo inicial sobre a possibilidade do uso da fluorescência como ferramenta na determinação de óxidos de Fe, Al, Si e Ti. O estudo foi realizado em Iconha/ES, na bacia do Ribeirão Inhaúma. Para análise de fluorescência foram utilizados o horizonte B de 10 perfis de solo coletados ao longo de toda bacia. As amostras foram secas ao ar, destorroadas e passadas em peneira com malha de 2 mm. No laboratório foram maceradas em almofariz de ágata e novamente peneiradas em malha 1 mm, posteriormente colocadas em molde padrão e prensadas manualmente, a fim de se formarem pastilhas. Essas pastilhas foram usadas para leitura no analisador de fluorescência de raios-X do Laboratório de Solos. O ataque sulfúrico das amostras foi feito conforme Embrapa (2011). Foram calculados as relações moleculares k_i e k_r e feita a análise de correlação entre os métodos. Os teores de SiO_2 e Al_2O_3 compuseram a maior parte da amostra nas duas análises. As diferenças de teores encontrados nas duas análises levam a diferentes valores dos índices K_i e K_r . Os maiores valores de Al_2O_3 , em relação a SiO_2 indicam o alto grau de intemperismo dos solos estudados. O índice k_r mostrou que há grande discrepância na interpretação dos resultados utilizando os dois métodos, já que o ataque sulfúrico apresentou valores bastante inferiores em relação a aqueles obtidos na análise de fluorescência. O índice k_i proporcionou mesma interpretação dos dados, independentemente da forma de análise utilizada. Verificou-se bons valores de R^2 para Al, Ti e Fe, com destaque para Ti com r^2 de 0,917 e ferro de 0,891. Al obteve valor de 0,636 e silício o pior desempenho com 0,269. Esse resultado demonstra a boa correlação existente entre os resultados obtidos pelas duas análises, indicando o potencial da análise de raios x ser utilizada como ferramenta na determinação desses óxidos. É necessário o estudo aprofundado dessas correlações, avaliando-se uma quantidade maior de amostras, que sirvam de subsídio para análises estatísticas que possam comprovar a eficiência do método, porém percebe-se que esses são resultados promissores para maior agilidade e eficiência na realização dessas análises.

Palavras-chave: ataque sulfúrico, fluorescência, K_i , K_r

ABSTRACT

The levels of SiO_2 , Al_2O_3 , TiO_2 and Fe_2O_3 found in the soil, help to determine different factors such as the degree of weathering and development of these. Traditionally, the analysis of the oxides contents is done in the laboratory by the method called total attack, however this is a dispensing method when compared to other options such as thermofluorescence analysis. Thus, it would be interesting to study the results of both methods in order to analyze alternatives to routine laboratories that work with a large volume of daily analyzes, thus facilitating this service. Thus, this study proposes an initial study about the possibility of using

thermofluorescence as a tool in the determination of Fe, Al, Si and Ti oxides. The study was carried out in Iconha / ES, in the Ribeirão Inhaúma basin. For thermofluorescence analysis, the B horizon of 10 soil profiles collected along the entire basin was used. The samples were air-dried, stripped and sieved with a 2 mm mesh. In the laboratory, they were macerated in agate mortar and sieved again in a 1 mm mesh, later placed in standard mold and manually pressed in order to form pellets. These pellets were used for reading in the soil laboratory's X-ray fluorescence analyzer. The sulfuric attack of the samples was done according to Embrapa (2011). The molecular relations k_i and k_r were calculated and the correlation analysis between the methods was done. The SiO₂ and Al₂O₃ contents made up the majority of the sample in the two analyzes. The differences in contents found in the two analyzes lead to different values of K_i and K_r indexes. The higher values of Al₂O₃ in relation to SiO₂ indicate the high degree of weathering of the studied soils. The k_r index showed that there is a great discrepancy in the interpretation of the results using the two methods, since the sulfuric attack showed values much lower than those obtained in the thermofluorescence analysis. The k_i index provided the same interpretation of the data, regardless of the form of analysis used. Good values of R² for Al, Ti and Fe were verified, with Ti with r^2 of 0.917 and iron of 0.891. Al obtained a value of 0.636 and silicon the worst performance with 0.269. This result demonstrates the good correlation between the results obtained by the two analyzes, indicating that the potential of x-ray analysis is used as a tool in the determination of these oxides. It is necessary to study in depth these correlations, evaluating a larger number of samples, which serve as a subsidy for statistical analysis that can prove the efficiency of the method, but it is perceived that these are promising results for greater agility and efficiency in carrying out these analyzes.

Keywords: sulfuric attack, fluorescence, K_i , K_r

1 INTRODUÇÃO

Dados relativos aos teores de SiO₂, Al₂O₃, TiO₂ e Fe₂O₃ encontrados no solo, são de grande valia para se determinar vários fatores, como o grau de desenvolvimento deste. O método mais utilizado atualmente para essa quantificação é o ataque total. Porém este é um método dispensioso e potencial poluidor do meio ambiente, pelo uso de reagentes químicos, quando comparado a outras possíveis opções como a fluorescência, que necessita apenas de preparo físico da amostra e apresenta como vantagens analisar as mesmas de forma não destrutiva, rápida e econômica.

Em 2001, Nagata et al., ponderavam que a utilização da fluorescência de raios-X em rotinas de análise química, já deveria ter um uso maior e mais frequente do era observado na literatura. Atualmente, observa-se uma tendência de uso desse método para avaliação de diferentes atributos do solo (TIGHE et al., 2018; CHAKRABORTY et al, 2017; KANIU e ANGEYO, 2015; KANIU et al., 2012).

A análise por Fluorescência de Raios X é um método quali-quantitativo baseado na medida das intensidades dos Raios X característicos emitidos pelos elementos que constituem a amostra (POTTS, 2008). Segundo Jenkins (1988), a espectrometria de fluorescência de raios-X permite identificar elementos através da medição do comprimento de onda ou da energia de uma emissão de raios-X.

É uma análise rápida, feita em aparelho de fluorescência de raios-X, que não exige uso de reagentes químicos. Dessa forma, apresenta-se como uma forma mais sustentável e segura de análise em relação ao meio ambiente. Ao contrário, o ataque sulfúrico, exige a manipulação de ácidos e necessita de maior tempo para análise. Nanni e Demattê (2006) apontam que a análise

de laboratório, usada para quantificação de atributos do solo, apresentam custo elevado (BASHAGALUKE et al., 2015) e possibilidade de geração de impactos ambientais pelo uso de reagentes químicos na análise, além de dispender muito tempo entre a amostragem e a aquisição dos resultados.

Dessa forma, seria interessante um estudo que permitisse avaliar e comparar os resultados de ambos os métodos, para analisar alternativas para laboratórios de rotina que trabalham com grande volume de análises diárias, facilitando assim este serviço.

Com isso, este trabalho propõe um estudo inicial sobre a possibilidade do uso da fluorescência como ferramenta na determinação de óxidos de Fe, Al, Si e Ti.

2 REVISÃO BIBLIOGRÁFICA

2.1 Ataque Sulfúrico

A determinação de teores totais de elementos do solo, pode ser feita segundo diferentes métodos, a eficiência de cada um deles varia conforme o tipo de elemento que se deseja determinar e também com a composição do solo (SALDANHA et al, 1997).

Dentre esses métodos, o ataque sulfúrico é utilizado na determinação de Fe, Al, Si, Ti, Mn e P. Esses elementos são determinados na forma de óxidos, sendo Fe₂O₃, Al₂O₃, SiO₂, TiO₂, MnO₂ e P₂O₅.

Em 1969, Vettori publicou boletim técnico sobre métodos de análise do solo, cuja determinação de Si, Al, Fe, T e P era feita por meio do ataque sulfúrico.

Nesse método usa-se o ácido sulfúrico como reagente, pois, segundo Saldanha (1997), seu poder oxidante é praticamente nulo, apenas aquele proveniente dos íons hidrogênio têm esse efeito. Porém, estes são facilmente reduzidos a hidrogênio gasoso, pela ação dos metais em solução. Entretanto, Bock (1979) ressalta que quando aquecido o ácido sulfúrico, causa oxidação praticamente total dos compostos orgânicos, podendo ainda outros elementos serem oxidados, inclusive o próprio ácido sulfúrico em SO₃, S ou H₂S (SALDANHA, et al., 1997).

A determinação dos elementos Si, Al e Fe, permitem a determinação de relações moleculares, como o Ki e Kr que são índices, que permitem inferir sobre o grau de intemperização dos solos (EMBRAPA, 2011).

2.2 Índices Ki e Kr

Existem diferentes formas criadas para quantificar o grau de intemperismo do solo, estas incluem métodos químicos, físicos, petrográficos, e com base em difração de raios-X e microscopia eletrônica (OLIVEIRA, 2006).

Harrassowitz (1926 apud VIEIRA, 1975; FALCÃO, 1984) propôs os índices Ki e Kr. Estes índices são definidos por:

$$K_i = \frac{SiO_2}{Al_2O_3} \quad (1)$$

(2)

$$K_r = \frac{SiO_2}{(Al_2O_3 + Fe_2O_3)}$$

Onde: SiO₂ é o teor de Dióxido de Silício,
Al₂O₃ é o teor de Óxido de Alumínio
Fe₂O₃ é o teor de Óxido de Ferro

Para se utilizar o teor percentual no cálculo de Ki e Kr, emprega-se as seguintes equações (REATTO et al., 1998):

$$K_i = 1,7 \times \frac{\% SiO_2}{\% Al_2O_3} \quad (3)$$

$$K_r = 1,7 \times \frac{\% SiO_2}{\% Al_2O_3 + (0,6325 \times \% Fe_2O_3)} \quad (4)$$

Valores de $K_i \geq 2,2$ pressupõe solos jovens, e valores $< 2,2$ solos maduros; enquanto valores de $K_r > 0,75$ determina solos cauliníticos, e quando o $K_r \leq 0,75$, trata-se de solos oxídicos (RESENDE & SANTANA, 1988). Segundo o SBCS (2006), solos cauliníticos são aqueles com predominância de argilominerais do grupo da caulinita, enquanto solos oxídicos têm predominância de óxidos de ferro e alumínio, podendo ser divididos em hematíticos e goetíticos.

2.3 Fluorescência De Raios-X

A fluorescência de raios X é uma técnica desenvolvida ao longo do séc. XX, presente desde 1913, quando demonstrou-se a relação entre a energia de raios-X característica de um elemento e o seu número atômico, tendo ao longo dos anos sofrido diversas melhorias nos detectores e tubos de raios-X (FELICZAKI e MELQUIADES, 2014).

Os raios X emitidos por tubos de raios X, ou raios X ou gama por uma fonte radioativa, excitam os elementos que constituírem, os quais, por sua vez, emitem linhas espectrais com energias características do elemento e cujas intensidades estão relacionadas com a concentração do elemento na amostra (NASCIMENTO FILHO, 1999). Ou seja, a espectrometria de fluorescência de raios-X realiza análises da composição química de uma amostra através da detecção dos raios-X característicos dos seus elementos (URBANO E BRANDÃO, 2013).

Uma revisão mais detalhada sobre a fluorescência de raios-x é apresentada no capítulo V.

Santos et al. (2010), analisaram Silício, Alumínio e Ferro por fluorescência, calculando os índices Ki e Kr a partir desses teores. Wastowski et al. (2010) determinaram os valores de Al, Fe, dentre outros elementos utilizando o método de fluorescência. Weindorf et al. (2012), caracterizaram diferentes solos utilizando fluorescência de raios-x, analisando as concentrações dos elementos presentes. Esses estudos demonstram a importância da técnica no estudo do solo.

3 MATERIAL E MÉTODOS

O estudo foi realizado em Iconha/ES, na bacia do Ribeirão Inhaúma com área de 2403,9 ha. Essa encontra-se sob as coordenadas 21°10'58.82"S e 41°00'08.87"O, possui relevo acidentado, clima Aw segundo a classificação de Koppen e predominância de Cambissolos.

Para análise de fluorescência foram utilizados o horizonte B de 10 perfis de solo coletados ao longo de toda bacia. A descrição morfológica desses foi realizada conforme Santos et al. (2015) e classificados de acordo com Embrapa (2013).

Os locais de amostragem foram determinados considerando as possibilidades de acesso devido às dificuldades de deslocamento na área em função do relevo acentuado e buscando ser o mais representativo possível em relação à paisagem da bacia (Figura 74).

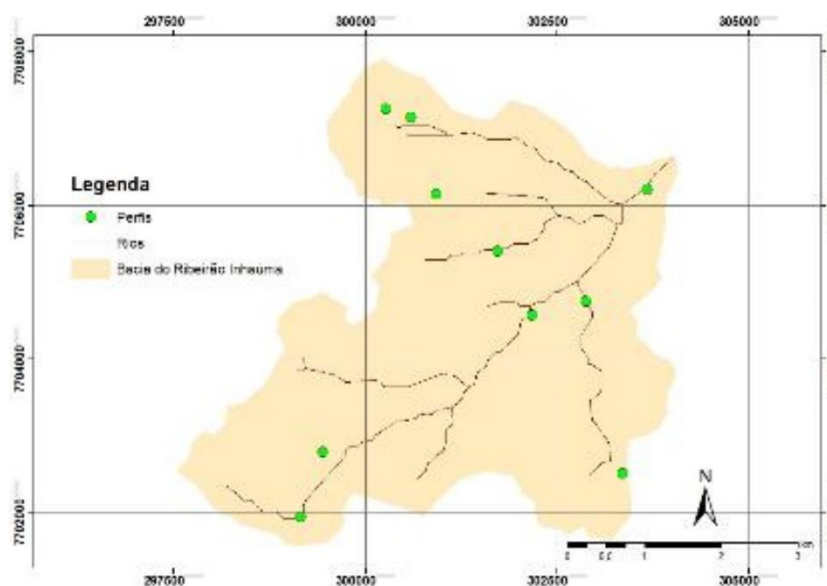


Figura 74- Malha de amostragem para análise de Fe, Al, Si e Ti.

As amostras foram secas ao ar, destorroadas e passadas em peneira com malha de 2 mm. No laboratório foram maceradas em almofariz de ágata e novamente peneiradas em malha 1 mm, posteriormente colocadas em molde padrão e prensadas manualmente, a fim de se formarem pastilhas.

Essas pastilhas foram usadas para leitura no analisador de fluorescência de raios-X do Laboratório de Solos, da Universidade Federal de Viçosa-UFV. Foi utilizado para tal o modelo Micro-EDX-1300 da Shimadzu, foram usadas as voltagens padrão do equipamento. As curvas de calibração foram obtidas com amostras padrão.

O ataque sulfúrico das amostras foi feito conforme Embrapa (2011). Primeiramente foi preparado o ácido sulfúrico diluído de 1:1, adicionando 500 mL de H₂SO₄ concentrado d=1,84 g/cm³ a um bécher de dois litros contendo 500 mL de água.

Para a análise, inicialmente se pesou 1 g de solo e este foi colocado em erlenmeyer 500 mL, adicionou-se 20 mL de ácido sulfúrico diluído de 1:1 e foi feita à fervura por meia hora, usando condensador de refluxo para evitar evaporação. Após a solução esfriar, foi adicionado

50 mL de água e filtrado para balão aferido de 250 mL, lavando o resíduo com água até completar o volume. No filtrado foi determinado ferro, alumínio, titânio e no resíduo silício.

Segundo Embrapa (2011), o extrato sulfúrico produzido com tal força iônica sugere que apenas os argilominerais (minerais secundários) são dissolvidos, dessa forma garante-se que os teores obtidos de Fe, Al, Si e Ti são próximos aos da fração coloidal do solo.

Foram calculados as relações moleculares k_i e k_r , segundo Camargo et al. (2009), a partir das equações citadas anteriormente e posteriormente foi feita a análise de correlação entre os métodos.

4 RESULTADOS E DISCUSSÃO

Dos dez perfis analisados, sete (perfis 1, 2, 3, 4, 5, 8 e 9) correspondem a classe dos Cambissolos, enquanto que três (perfis 6, 7 e 10) correspondem aos Latossolos. A Tabela 53 mostra os teores de SiO_2 , Al_2O_3 , TiO_2 e Fe_2O_3 obtidos pela análise de fluorescência, enquanto a Tabela 54 os mesmos obtidos pelo ataque total. Os índices K_i e K_r foram calculados em ambos métodos.

Para todos os elementos espera-se que os teores obtidos pelo método e fluorescência sejam superiores, já que o método permite a leitura até mesmo dos elementos estruturais e não apenas dos disponíveis. Essa diferença ocorre também pelo fato da análise do ataque total ter sido realizada somente utilizando a argila, enquanto que na fluorescência foi feita no solo. A influência das frações silte e, principalmente a areia, com presença de quartzo, alteram de maneira significativa os valores de elementos como a sílica (Tabela 55).

Os teores de SiO_2 e Al_2O_3 compõem a maior parte da amostra nas duas análises, quanto a análise de fluorescência os valores variam de 48 a 67 %, enquanto no ataque total de 38 a 72 %, sendo essa última percentagem discrepante em relação aos demais, que se mantiveram em torno de 48 %.

As diferenças de teores encontrados nas duas análises levam a diferentes valores dos índices K_i e K_r . Porém, partindo-se da interpretação sugerida para Latossolos e Cambissolos (Embrapa, 2013), onde valores de K_i inferiores a 2,2 são normalmente encontrados para Latossolos, enquanto valores acima destes são comuns para os Cambissolos, a discrepância não afetou a análise dos resultados.

Os maiores valores de Al_2O_3 , em relação a SiO_2 indicam o alto grau de intemperismo dos solos estudados. Esse fato pode ser corroborado pelos baixos valores de K_i encontrados para ambos solos. Neste trabalho não foi encontrado nenhum valor de K_i superior a 1,96 (fluorescência) e 1,62 (ataque sulfúrico), indicando que a composição dos Cambissolos pode ter exercido influência no resultado. Ker (1997) mostra que solos pouco evoluídos, como os Cambissolos, podem apresentar baixos valores desses índices se desenvolvidos a partir de materiais ricos em ferro.

Marques et al. (2007) estudando Cambissolos encontraram K_i variando de 0,64 a 1,81 nos horizontes A e B, e atribuíram esse fato a presença de gibbsita. Alves et al. (2008) trabalhando com Latossolos encontrou valores de k_i entre 0,19 e 1,67. Rolim Neto et al. (2009) encontraram valores de K_i entre 1,71 e 7,88 para o horizonte Bi de Cambissolos, e 0,37 a 1,08 para Latossolos. Oliveira (2001), menciona que solos com K_i entre 1 e 2 são predominantemente ácidos e com argila de atividade baixa. Alves et al. (2008) utilizando espectroscopia de fluorescência de raios-X para determinação dos teores de Si e Al, encontraram valores de k_i entre 0,19 e 1,67 em Latossolos.

Santos et al (2010) encontraram valores de Ki obtidos por ataque sulfúrico inferiores à aqueles determinados por fluorescência de raios-X, porém assim como nesse trabalho encontraram o mesmo padrão de comportamento entre os elementos analisados.

O índice Kr sugere o grau de intemperização, onde valores maiores que 0,75 indicam solos caulíníticos e valores menores ou igual solos oxídicos. Para os índice calculado com os dados de fluerecência apenas a amostra 7 apresentou valor menor que 0,75 ; enquanto no ataque sulfúrico a maior parte das amostras (3,6,7,8,9 e 10) apresentou esse comportamento. Costa (2003), estudando Latossolos encontrou valores baixíssimos para esses índices em sua área de estudo. Rolim Neto et al. (2009) encontraram valores de Kr variando entre 0,60 a 1,31 para horizonte B de Cambissolos e de 0,30 a 0,46 para Latossolos.

Segundo Melo et al. (1995) para valores de Kr superiores a 0,75 é possível classificar o material do perfil como muito intemperizado, com presença de matéria mineral de natureza caulínítica, porém não oxídica.

Analisando esse índice (Kr), ao contrário de Ki, é possível perceber que há grande discrepância na interpretação utilizando os dois métodos, já que o ataque sulfúrico apresentou valores bastante inferiores em relação a aqueles obtidos na análise de fluorescência.

Tabela 53- Teores obtidos para SiO₂, Al₂O₃, TiO₂ e Fe₂O₃ através da análise de fluorescência, e índices Ki e Kr no horizonte B de latossolos e cambissolos

Amostra	Elementos				Índices	
	SiO ₂	Al ₂ O ₃	TiO ₂	Fe ₂ O ₃	ki	kr
	%					
1	31,60	27,39	1,67	9,19	1,96	1,62
2	25,53	33,03	3,84	18,57	1,31	0,97
3	23,25	30,08	2,83	10,27	1,31	1,08
4	24,82	33,64	1,96	13,64	1,25	1,00
5	26,12	27,27	3,06	15,37	1,63	1,20
6	24,09	30,97	2,02	13,70	1,32	1,03
7	17,38	33,47	2,73	16,93	0,88	0,67
8	21,46	26,92	4,00	15,83	1,36	0,99
9	21,02	34,04	1,63	10,98	1,05	0,87
10	29,72	37,63	1,73	13,63	1,34	1,09

Tabela 54- Teores obtidos para SiO₂, Al₂O₃, TiO₂ e Fe₂O₃ através da análise de ataque total, índices Ki e Kr no horizonte B de latossolos e cambissolos

Amostra	Elementos				Índices	
	SiO ₂	Al ₂ O ₃	TiO ₂	Fe ₂ O ₃	ki	kr
	%					
1	21,31	27,19	1,52	9,78	1,33	1,09
2	15,11	22,40	2,41	16,95	1,15	0,78
3	15,61	29,28	3,03	16,31	0,91	0,67
4	19,95	27,58	2,57	10,71	1,23	0,99
5	35,23	36,90	1,75	14,04	1,62	1,31
6	20,88	30,79	2,79	16,62	1,15	0,86
7	14,44	31,14	1,71	13,75	0,79	0,62
8	12,78	25,34	3,61	16,31	0,86	0,61
9	11,73	35,95	1,43	11,32	0,55	0,46
10	14,43	34,44	1,27	11,63	0,71	0,59

Tabela 55- Análise granulométrica do horizonte B dos perfis amostrados no Ribeirão Inhaúma, Iconha/ES

Amostra	Argila total	Areia fina	Areia grossa	Areia total	Silte
1	31,20	12,20	35,60	47,80	21,00
2	42,20	14,70	20,00	34,70	23,10
3	36,20	19,90	32,90	52,80	11,00
4	46,90	4,50	18,90	23,40	29,70
5	51,90	10,60	9,10	19,70	28,40
6	27,00	11,70	24,70	36,40	36,60
7	43,50	10,90	23,60	34,50	22,00
8	28,70	15,00	29,70	44,70	26,60
9	41,80	15,20	33,00	41,10	17,10
10	41,80	11,10	20,80	31,90	26,30

A Figura 75 apresenta a correlação existente entre os valores encontrados na análise dos teores de silício, alumínio, titânio e ferro por fluorescência e ataque total. Para se proceder a correlação eliminou-se os pontos considerados *outliers*.

Verifica-se bons valores de r² para Al, Ti e Fe, com destaque para Ti com r² de 0,917. Já ferro apresentou r² de 0,891 e o Al obteve valor de 0,636. O silício foi o pior desempenho, com 0,269 ; provavelmente devido ao feito da sílica do quartzo na fração areia. Esse resultado demonstra a boa correlação existente entre os resultados obtidos pelas duas análises, indicando o potencial da análise de raios-X se utilizada como ferramenta na determinação desses óxidos.

É necessário o estudo aprofundado dessas correlações, avaliando-se uma quantidade maior de amostras, que sirvam de subsídio para análises estatísticas que possam comprovar a

eficiência do método, porém percebe-se que esses são resultados promissores para maior agilidade e eficiência na realização dessas análises.

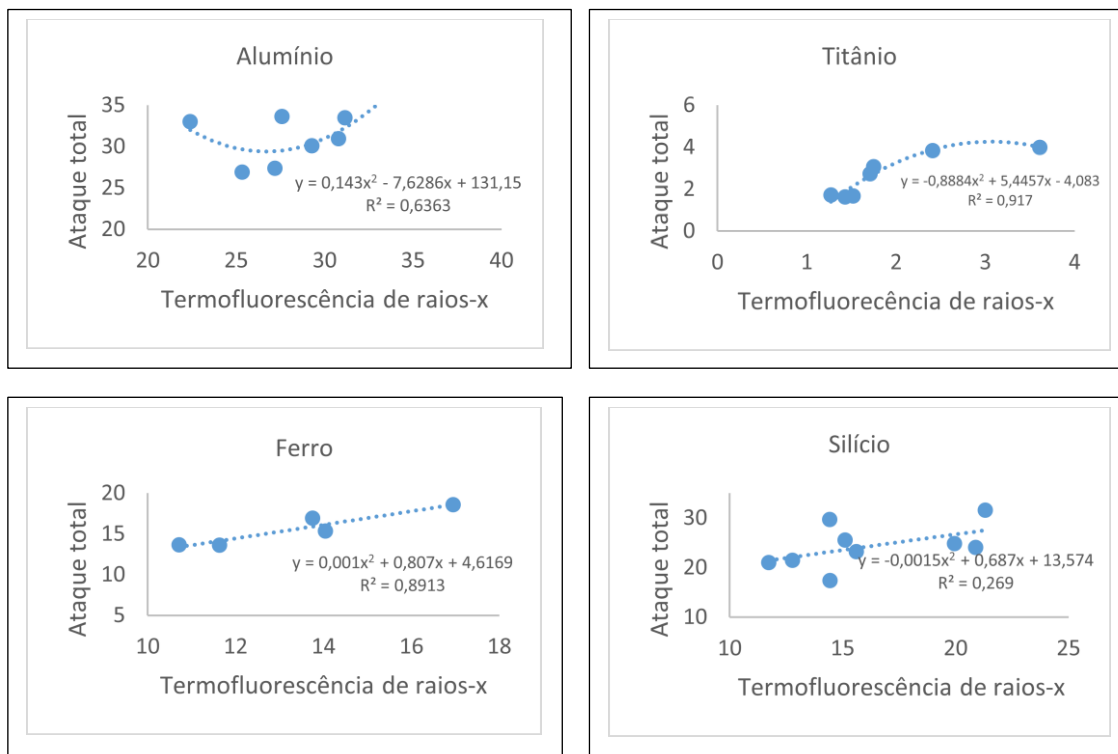


Figura 75- Correlação entre teores obtidos de Alumínio, Titânio, Ferro e Silício, por ataque total e fluorescência de raios-x.

5 CONCLUSÕES

1. As análises de fluorescência de raios-X mostram-se promissoras para determinação dos teores de óxidos de Al, Ti, Si e Fe de Cambissolos e Latossolos;
2. O índice Ki proporcionou mesma interpretação dos dados, independentemente da forma de análise utilizada. Kr apresentou valores mais discrepantes influenciando na interpretação de acordo com o método de análise utilizado;
3. São necessárias análises de uma quantidade maior de amostras utilizando os dois métodos, afim de comprovar a correlação existente entre os teores encontrados.

6 REFERÊNCIAS BIBLIOGRÁFICAS

ALVES, Marcelo Eduardo; MASCARENHAS, Yvonne Primerano; VAZ, Carlos Manoel Pedro. Comparação de procedimentos de quantificação de caulinita e gibbsita na fração argila desferrificada de Latossolos. *Revista Brasileira de Ciência do Solo*, [s.l.], v. 32, n. 2, p.569-577, abr. 2008. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0100-06832008000200012>.

BASHAGALUKE JANVIER; NSHOBOLE NICOLE; FATAKI DIDIER, MOCHOGE BENSON, MUGWE JAYNE, WALANGULULU JEAN. Application of infrared technique in soil properties characterization in South Kivu province of DR Congo. *African Journal of Food Science and Technology*. Vol. 6(2) pp. 58-67, Feb 2015.

CAMARGO, O.A.; MONIZ, A.C.; JORGE, J.A. & VALADARES, J.M.A.S. Métodos de análises química, mineralógica e física de solos do Instituto Agrônomo de Campinas. Campinas, Instituto Agrônomo de Campinas, 1986. 94p. (Boletim Técnico, 106)

COSTA, SAD da. Caracterização química, física, mineralógica e classificação de solos ricos em ferro do Quadrilátero Ferrífero. [Dissertação]. Universidade Federal de Viçosa-UFV. Viçosa/MG. 83 p. 2003.

CHAKRABORTY, S; MAN T.; PAULETTE L.; DEB S.; LI B.;WEINDORF D.C.; FRAZIER M. Rapid assessment of smelter/mining soil contamination via portable X-ray fluorescence spectrometry and indicator kriging. *Geoderma*, [s.l.], v. 306, p.108-119, nov. 2017. Elsevier BV. <http://dx.doi.org/10.1016/j.geoderma.2017.07.003>.

EMBRAPA – CENTRO NACIONAL DE PESQUISA DE SOLO. **Manual de Métodos de Análise de Solo**. 2ed. Rio de Janeiro, RJ: EMBRAPA Solos, 230p. - Documentos/Embrapa Solos, ISSN 1517-2627:132, 2011. EMBRAPA.

EMBRAPA. **Sistema Brasileiro de Classificação de Solos**. 3ed. Brasília, DF: EMBRAPA Informação Tecnológica, 353p, 2013.

FALCÃO, M.F.G. Fatores Geológicos e Características de Intemperismo Condicionantes das Propriedades Mecânicas de Solos Residuais. Dissertação de Mestrado, DEC, PUC-Rio, RJ, 1984. 131p.

FELICZAKI, RAFAEL MATEUS; MELQUIADES, FABIO LUIZ. Simulação de fluorescência de raios x em amostras de solo com influência de umidade. **Revista Brasileira de Física Tecnológica Aplicada**, [s.l.], v. 1, n. 1, p.1-8, 1 maio 2014. Universidade Tecnológica Federal do Parana (UTFPR). <http://dx.doi.org/10.3895/s2358-00892014000100001>.

KANIU, M.I.; ANGEYO, K.H.. Challenges in rapid soil quality assessment and opportunities presented by multivariate chemometric energy dispersive X-ray fluorescence and scattering spectroscopy. *Geoderma*, [s.l.], v. 241-242, p.32-40, mar. 2015. Elsevier BV. <http://dx.doi.org/10.1016/j.geoderma.2014.10.014>.

KANIU, M.I.; ANGEYO K.H.; MWALA; A.K.; MANGALA M.J. Direct rapid analysis of trace bioavailable soil macronutrients by chemometrics-assisted energy dispersive X-ray

fluorescence and scattering spectrometry. *Analytica Chimica Acta*, [s.l.], v. 729, p.21-25, jun. 2012. Elsevier BV. <http://dx.doi.org/10.1016/j.aca.2012.04.007>.

KER, JC. Latossolos do Brasil: Uma revisão. *Geonomos*, v.5, n. 1, p.17-40, 1997.

MARQUES, FA; RIBEIRO MR; BITTAR SMB; LIMA NETO J de A; LIMA JFWF. Caracterização e classificação de Cambissolos do arquipélago de Fernando de Noronha, Pernambuco. *Revista Brasileira de Ciência do Solo*, [s.l.], v. 31, n. 5, p.1023-1034, out. 2007. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0100-06832007000500019>.

MELO V. F. et al. Reserva mineral e caracterização mineralógica de alguns solos do Rio Grande do Sul. *Revista Brasileira de Ciência do Solo*, Viçosa, v.19, p. 159-164, 1995.

NAGATA, N; BUENO, M.I.M.S; E PERALTA-ZAMORA, P.G. Métodos matemáticos para correção de interferências espectrais e efeitos interelementos na análise quantitativa por fluorescência de raios-X *Quim. Nova*, Vol. 24, No. 4, 531-539, 2001.

OLIVEIRA, E.P. Caracterização bio-físico-químico-mineralógica e micromorfológica de um perfil de alteração de granitognaisse de Curitiba, PR. MS Dissertation. Departamento de Engenharia Civil, Pontifícia Universidade Católica, 197 p. 2006.

REATTO, A., FONTES, M. P. F.; ALVAREZ V, V. H.; RESENDE, M.; KER, J. C.; COSTA. L. M. Caracterização mineralógica, potencial de reserva e sustentabilidade agrícola de alguns sítios florestais de eucalipto da região do Vale do Rio Doce (MG). *R. Bras. Ci. Solo*, 22:255-266, 1998.

RESENDE, M. & SANTANA, D.P. Uso das relações Ki e Kr na estimativa da mineralogia para a classificação dos Latossolos. In: REUNIÃO DE CLASSIFICAÇÃO, CORRELAÇÃO DE SOLOS E INTERPRETAÇÃO DE APTIDÃO AGRÍCOLA, 3., Rio de Janeiro, 1988. Anais. Rio de Janeiro, Embrapa – SNLCS, SBCS, 1988. p.225- 232.

ROLIM NETO, Fernando Cartaxo et al. Topolitossequências de solos do Alto Paranaíba: atributos físicos, químicos e mineralógicos. *Revista Brasileira de Ciência do Solo*, [s.l.], v. 33, n. 6, p.1795-1809, dez. 2009. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0100-06832009000600028>.

SANTOS, A do C; PEREIRA MG; ANJOS LHC dos; BERNINI T de A; COOPER M; NUMMER AR; FRANCELINO MR. Gênese e classificação de solos numa topossequência no ambiente de mar de morros do médio Vale do Paraíba do Sul, RJ. *Revista Brasileira de Ciência do Solo*, [s.l.], v. 34, n. 4, p.1297-1314, ago. 2010. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0100-06832010000400027>.

TIGHE, M; ROGAN G.; WILSON S.C.; GRAVE P.; KEALHOFER L.; YUKONGDI P. The potential for portable X-ray fluorescence determination of soil copper at ancient metallurgy sites, and considerations beyond measurements of total concentrations. *Journal Of Environmental Management*, [s.l.], v. 206, p.373-382, jan. 2018. Elsevier BV. <http://dx.doi.org/10.1016/j.jenvman.2017.10.052>.

VIEIRA, L.S. Manual de Ciência do Solo. 1ed, São Paulo, Editora Agronômica Ceres, 1975.

WASTOWSKI ARCI DIRCEU, ROSA GENESIO MARIO DA, CHERUBIN MAURÍCIO ROBERTO; RIGON JOÃO PAULO GONSIORKIEWICZ. Caracterização dos níveis de

elementos químicos em solo, submetido a diferentes sistemas de uso e manejo, utilizando espectrometria de fluorescência de raios-x por energia dispersiva (EDXRF). *Quim. Nova*, Vol. 33, No. 7, 1449-1452, 2010.

WEINDORF, DAVID C; ZHU YUANDA; MCDANIEL PAUL; VALERIO MITCHELL; LYNN LORENE; MICHAELSON GARY; CLARK MARK; PING CHIEN LU. Characterizing soils via portable x-ray fluorescence spectrometer: 2. Spodic and Albic horizons. *Geoderma*, [s.l.], v. 189-190, p.268-277, nov. 2012. Elsevier BV. <http://dx.doi.org/10.1016/j.geoderma.2012.06.034>.

CONCLUSÕES GERAIS

- É possível o mapeamento de classes e atributos do solo da bacia hidrográfica do Ribeirão Inhaúma no município de Iconha/ES utilizando-se diferentes técnicas com boa acurácia;
- Lógica *fuzzy* é um método eficiente para mapeamento de classes de solo na área estudada;
- O mapeamento digital é uma técnica eficiente a ser indicada para mapeamento de solos no estado do Espírito Santo.
- As análises de fluorescência de raios-X mostram-se promissoras para determinação dos teores de óxidos de Al, Ti, Si e Fe de Cambissolos e Latossolos;

ANEXOS

ANEXO 1 – PROCESSAMENTO DOS DADOS AEROGEOFÍSICOS (CPRM, 2010, p. 29 a 34):

Correção do Tempo Morto

A correção do “tempo morto” consiste na divisão das contagens dos canais radiométricos pelo valor do *live time* registrado pelo aparelho, normalizando, assim, os valores brutos, dos canais da contagem total, potássio, urânio, tório e urânio *upward* para contagem por segundo.

Aplicação de Filtragem

Este processo se aplica somente aos dados afetados por variações de alta frequência quais sejam: dados do radar altímetro, do canal de radiação cósmica e do canal de urânio *up*, utilizando no cálculo da influência do radônio nas medições realizadas.

Dependendo do comportamento dos dados, são aplicados dois tipos de filtragem:

Filtragem não linear, que permite a remoção de *spikes* nos dados e a compensação de variações abruptas do radar altímetro;

Filtragem do tipo passa-baixa, que reduz o erro estatístico nos dados da radiação cósmica, suaviza o comportamento do radônio. Opcionalmente, é aplicado aos demais canais radiométricos com objetivos específicos, como o cálculo das razões radiométricas.

Remoção do *background* da Aeronave e Cósmico

O *background* é obtido através do somatório das contribuições do *background* da aeronave e da radiação cósmica em cada uma das janelas do gamaespectrômetro.

O cálculo das contribuições da aeronave e da radiação cósmica é conduzido através da fórmula (IAEA, 1991):

$N = a + bC$, onde:

N – somatório das duas contribuições (em cps),

a – *background* da aeronave em cada janela do gamaespectrômetro,

C – canal de radiação cósmica.

b – razão entre a contagem em determinada janela e a contagem no canal do cósmico.

Remoção do *Background* do Radônio

O efeito do *background* do radônio, por sua vez, é determinado a partir das medições realizadas na janela do urânio pelo detector *upward looking*. A expressão que define a parcela de radônio influenciando no canal do urânio é a seguinte (IAEA, 1991):

$$U_r = (u - a_1U - a_2Th) + (a_2 b_t - b_u) / (a_u - a_1 - a_2 a_t),$$

Onde:

U_r – *background* do radônio medido no canal *downward* do urânio,

u – contagem medida no canal *upward* do urânio,

U – contagem medida no canal *downward* do urânio,

Th – contagem medida no canal *downward* do tório,

a1, a2, au, at, bu, bt – coeficientes de proporcionalidade, sendo que bu e bt são zerados e a1 e

a2 os coeficientes *skyshine*.

As contagens relativas ao urânio, tório e urânio *up* devem ser corrigidas previamente dos efeitos de *backgrounds* da aeronave e cósmico.

A relação entre as contagens atribuídas ao radônio observadas na janela do urânio, com as demais janelas de canais *downward*, detectadas nos cristais voltados para baixo, pode ser determinada através de regressão linear aplicada sobre um conjunto de dados que reflita as variações decorrentes da presença de radônio nos dados.

O procedimento de cálculo dos coeficientes que expressam a relação entre os detectores (*upward* e *downward*) utiliza as seguintes fórmulas (IAEA, 1991):

$$U_r = a_u U_u + b_u, K_r = a_k U_r + b_k, T_r = a_t U_r + b_t,$$

$$T_{C_r} = a_{t_c} U_r + b_{t_c}, \text{ onde:}$$

ur é a componente do radônio no urânio *up*, Ur, Kr, Tc e TCr são as contribuições do radônio nas demais janelas associadas ao detector *downward*. Se os componentes dos *backgrounds* da aeronave e cósmico são perfeitamente removidos, as constantes “b’s” (bu, bk, bt e btc) devem ser zeradas (IAEA, 1991, p. 27).

No cálculo dos coeficientes atc, ak, au e at (Tabela 1) foram utilizados os valores calculados a partir dos testes de *high level*.

Tabela 1 – Constantes de Calibração do Radônio.

Canal	Coefficiente	PR-PEC	PR-PRS
Contagem Total	atc	15,72	15,35
Potássio	ak	0,86	0,83
Urânio <i>up</i>	au	0,17	0,15
Tório	at	0,14	0,16

Estimativa dos Coeficientes *Skyshine* (a1 e a2)

Estes coeficientes relacionam a contribuição das radiações de urânio e tório provenientes do terreno, que influenciam as contagens do urânio no detector *upward*. Admitindo-se que tais contribuições variem linearmente com as contagens destas mesmas radiações nas janelas de urânio e tório, nos detectores voltados para baixo, a estimativa destes coeficientes emprega a expressão geral (GRASTY & MINTY, 1995):

$$u_g = a_1 U_g + a_2 T_g, \text{ onde:}$$

ug – contribuição do solo na janela do urânio *up*;

Ug – contribuição do solo na janela do urânio *down*; ug – contribuição do solo na janela do urânio *down*; a1 e a2 – constantes de calibração requeridas.

A partir de uma série de valores de u_g , U_g e T_g os fatores de calibração a_1 e a_2 podem ser determinados pelo método dos mínimos quadrados. Isso pode ser feito resolvendo as duas equações simultâneas abaixo:

$$a_1 \sum (U_g)^2 + a_2 \sum U_g T_g = \sum u_g U_g$$

$$a_1 \sum U_g T_g + a_2 \sum (U_g)^2 = \sum u_g T_g$$

Este processo foi efetuado automaticamente utilizando todos os dados do levantamento. Os valores obtidos no levantamento em pauta estão demonstrados na Tabela 2 abaixo.

Tabela 2 – Coeficientes de *Skyshine*.

Coeficiente	PR-PEC	PR-PRS
a1	0,0563	0,0456
a2	0,0157	0,0170

Correção do Efeito *Compton*

É aplicada com objetivo principal de eliminar a influência das radiações atribuídas aos canais de mais alta energia que penetram nos canais de baixa energia, quais sejam: contribuições do tório no urânio e no potássio, assim como a contribuição do urânio no potássio. Nos sistemas de alta resolução, em uso atualmente, são também consideradas as influências de radiações de baixa energia nas janelas de energia mais alta, resultando, desta forma, nos seis coeficientes abaixo:

α – radiações de tório no urânio,

β – radiações de tório no potássio, γ – radiações de urânio no potássio, a – radiações de urânio no tório,

b – radiações de potássio no tório, g – radiações de potássio no urânio.

Os valores adotados para correção do efeito *Compton* são os descritos na Tabela 12– Coeficientes de Espalhamento *Compton*, apresentada no item 2.3.6. Maior detalhamento deste teste está presente no Anexo I-b.

Correção Altimétrica (Coeficiente de Atenuação Atmosférica)

A correção altimétrica tem por objetivo referenciar os valores radiométricos à altura nominal do aerolevantamento (100 m), eliminando falsas anomalias ocasionadas por elevações no terreno.

A atenuação das radiações gama em relação ao afastamento da fonte, pode ser expressa matematicamente, de forma aproximada, pela fórmula (IAEA, 1991):

$$N_H = N_0 \cdot e^{-\mu H} \quad (1), \text{ onde:}$$

N_H é a radiação a distância H da fonte,

N_0 é a radiação na superfície do terreno ($H=0$),

μ é o coeficiente de atenuação atmosférica.

Extraindo-se o logaritmo neperiano na relação acima, tem-se:

$$\ln(NH) = -\mu H + \ln(N_0)$$

que é a equação de uma reta de coeficiente angular $-\mu$ e o coeficiente linear $\ln(N_0)$. Na determinação dos coeficientes de atenuação atmosférica (μ) para cada um dos canais radiométricos foram utilizados os valores apresentados na Tabela 3, obtidos durante os testes realizados em Maricá-RJ. Os gráficos apresentando a correlação logarítmica entre as contagens e a altura de vôo encontram-se no Anexo I-f.

Tabela 3 – Coeficiente de Atenuação Atmosférica.

Canal Radiométrico	μ em m^{-1}	
	PR-PEC (20/02/2009)	PR-PRS (23/05/2009)
Contagem Total	-0,0067	-0,0069
Potássio	-0,0085	-0,0090
Urânio	-0,0069	-0,0084
Tório	-0,0067	-0,0067

Conversão para Concentração de Elementos

As sensibilidades dos detectores das aeronaves para as janelas do potássio, urânio e tório foram determinadas com base nas razões entre as medições efetuadas a bordo (N) e em terra (C), com aplicação da expressão:

$$S = N/C,$$

Onde:

S corresponde à sensibilidade para cada janela,

N é a média das contagens corrigidas (em cps) para cada canal referente à altura do levantamento (100 m) e situada no trecho de interesse das estações terrestres utilizadas,

C é a média das concentrações para cada canal das estações terrestres de interesse.

A Tabela 4 a seguir, corresponde à sensibilidade dos detectores analisados (volume total de 42 litros e dimensões de 10 cm x 10 cm x 40 cm cada), tomando por base a altura de vôo de 100 m sobre o terreno.

Tabela 4 – Coeficientes de Sensibilidade.

Canal Radiométrico	Sensibilidade	
	PR-PEC (20/02/2009)	PR-PRS (23/05/2009)
Contagem Total	191,09 cps/ μ R/h CT	240,55 cps/ μ R/h
Potássio	62,46 cps / %K	80,23 cps / %K
Urânio	10,17 cps / ppm eU	11,90 cps / ppm eU
Tório	3,69 cps /ppm eTh	4,51 cps /ppm eTh

Para calcular a taxa de exposição (*Exposure Rate*) do canal de contagem total (em μ R/h) utiliza-se a fórmula abaixo (IAEA, 1991):

$E = 1,505K + 0,653eU + 0,287eTh$, onde:

K, eU e eTh correspondem às concentrações aparentes destes elementos definidas em terra, quando do levantamento da pista de calibração dinâmica. Os valores da concentração determinados para o canal de contagem total foram 15,22 $\mu R/h$ (PR-PEC) e 13,97 $\mu R/h$ (PR- PRS).

Os gráficos comparativos dos perfis coletados durante a calibração dinâmica, após a conversão da amostragem para concentração aparente, acompanham o texto descritivo do procedimento de calibração apresentado no Anexo I-e.

Determinação das Razões eU/eTh, eU/K e eTh/K

As razões foram calculadas a partir dos valores radiométricos corrigidos. Para a eliminação de indeterminações, ou incorreções devidas a valores anormalmente abaixo, foi convencionada a fixação dos valores mínimos de K, U e Th em 10% da média destes canais na área de levantamento e aplicado também um filtro passa-baixa de 10 amostras. A Tabela 5 a seguir, resume os valores utilizados.

Tabela 5 – Valores Mínimos para Cálculo das Razões Radiométricas.

Canal Radiométrico	Valor Mínimo
K (%)	0,18
eU (ppm)	0,19
eTh (ppm)	1,40

APÊNDICES

APÊNCICE A- DESCRIÇÃO GERAL E MORFOLÓGICA DOS PERFIS

Perfil 1

DESCRIÇÃO GERAL

Classificação SiBCS: Cambissolo Háptico tb distrófico

Localização, município, estado e coordenadas: Estrada principal, sentido Inhaúma, no município de Iconha/ES, coordenadas 303675E e 7706204N

Situação, declive e cobertura vegetal sobre o perfil: Descrito e coletado em barranco de corte de estrada, terço médio, forte ondulado, sob consórcio de café e banana.

Altitude: 297 m

Litologia: Metatonalito

Material originário: Intemperismo do material de origem

Pedregosidade: Pedregosa

Rochosidade: Rochosa

Relevo local: Forte ondulado

Relevo regional: Montanhoso

Erosão: Forte- laminar

Drenagem: Bem drenado

Vegetação primária: Floresta subperenifólia

Uso atual: Cultivo de café e banana

Clima: Aw, da classificação de Koppen

Descrito por: Márcio Rocha Francelino, Patrícia Morais da Matta Campbell



2Ap 0-5 cm, **preto (7,5YR 2,5/1, úmida e 7,5YR 3/1, seca)**; franco-argiloarenosa; fraca pequena, granular; consistência firme, plástica e não pegajosa; transição ondulada e clara.

2Bi1 5-35 cm, **preto-brunado (7,5YR 3/1, úmida e 7,5YR 3/2, seca)**; franco-argiloarenosa; fraca pequena, blocos subangulares, consistência firme, ligeiramente plástica e ligeiramente pegajosa; transição ondulada e clara.

3Ab 35-65 cm, **bruno acinzentado (7,5YR 4/2, úmida e 7,5YR 4/3, seca)**; franco-argiloarenosa; fraca pequena, blocos subangulares, consistência friável, ligeiramente plástica e ligeiramente pegajosa; transição ondulada e gradual.

3Ab2 65-90 cm, **bruno (7,5YR 4/6, úmida e 7,5YR 4/4, seca)**; argilossiltosa; forte média, blocos angulares, consistência friável, não plástica e não pegajosa; transição plana e clara.

3Bi2 90-130 cm, **bruno (7,5YR 4/6, úmida e 7,5YR 5/6, seca)**; franco-argiloarenosa; cascalhenta, moderada média grande muito grande, blocos angulares, consistência firme, não plástica e não pegajosa; transição ondulada e clara.

Cr 130-180+

OBSERVAÇÕES: Não foi coletado e caracterizado o horizonte Cr, presença de concreções de Mn no 3Bi2; na transição 3Ab2 para 3Bi2 há presença de fragmento rochoso de 14-20 cm.

Perfil 2

DESCRIÇÃO GERAL

Classificação SiBCS: Cambissolo Háplico tb distrófico

Localização, município, estado e coordenadas: Localidade de Nova Esperança, no município de Iconha/ES, coordenadas 302871E e 7704748N.

Situação, declive e cobertura vegetal sobre o perfil: Descrito e coletado em barranco de corte de estrada, sob pastagem.

Altitude: 396 m

Litologia: Metatonalito

Material originário: Intemperismo do material de origem

Pedregosidade: Não pedregosa

Rochosidade: Não rochosa

Relevo local: Forte ondulado

Relevo regional: Montanhoso

Erosão: Moderada- laminar

Drenagem: Bem drenado

Vegetação primária: Floresta subperenifólia

Uso atual: Pastagem

Clima: Aw, da classificação de Koppen

Descrito por: Márcio Rocha Francelino, Patrícia Morais da Matta Campbell



DESCRIÇÃO MORFOLÓGICA

Colúvio 0-15 cm, **bruno (7,5YR 4/6, úmida e 7,5YR 5/6, seca)**; franco-argiloarenosa; fraca pequena, granular; consistência firme, ligeiramente plástica e não pegajosa; transição plana e abrupta.

Ap 15-40 cm, **bruno (7,5YR 4/6, úmida e 7,5YR 5/6, seca)**; franco-argilossiltosa; fraca pequena, blocos subangulares, consistência friável, não plástica e ligeiramente pegajosa; transição plana e abrupta.

- 2B1 40-55 cm, **bruno (7,5YR 4/6, úmida e 7,5YR 5/6, seca)**; argiloarenosa; fraca média, blocos subangulares, consistência friável, não plástica e não pegajosa; transição plana e abrupta.
- 2E1 55-60 cm, **bruno (7,5YR 4/4, úmida e 7,5YR 6/3 seca)**; franco siltosa; maciça, consistência solta, não plástica e não pegajosa; transição plana e abrupta.
- 2B2 60-70 cm, **bruno (5YR 4/6, úmida e 5YR 4/6, seca)**; franco-argilossiltosa; moderada média, blocos subangulares, consistência firme, não plástica e ligeiramente pegajosa; transição plana e abrupta.
- 2E2 70-75 cm, **bruno (7,5YR 4/4, úmida e 7,5YR 6/3, seca)**; franco-siltosa; maciça, consistência solta, não plástica e não pegajosa; transição plana e quebrada.
- 2B3 75-110 cm, **bruno (5YR 4/6, úmida e 5YR 5/6, seca)**; franco-argilossiltosa; fraca pequena, blocos subangulares, consistência friável, ligeiramente plástica e ligeiramente pegajosa; transição plana e abrupta.
- C 110-190 cm

RAÍZES: Comuns médias no horizonte 2B3, comuns finas no horizonte Ap; Raras médias nos demais

OBSERVAÇÕES: Não foi coletado e caracterizado o horizonte C.

Perfil 3

DESCRIÇÃO GERAL

Classificação SiBCS: Cambissolo Háplico tb distrófico

Localização, município, estado e coordenadas: Localidade de Nova Esperança, no município de Iconha/ES, coordenadas 303348E e 7702503N.

Situação, declive e cobertura vegetal sobre o perfil: Descrito e coletado em barranco de corte de estrada, sob pastagem.

Altitude: 640 m

Litologia: Metatonalito

Material originário: Intemperismo do material de origem

Pedregosidade: Ligeiramente pedregosa

Rochosidade: Não rochosa

Relevo local: Forte ondulado-Montanhoso

Relevo regional: Montanhoso-Escarpado

Erosão: Ligeira- laminar

Drenagem: Bem drenado

Vegetação primária: Floresta subperenifólia

Uso atual: Pastagem

Clima: Aw, da classificação de Koppen

Descrito por: Márcio Rocha Francelino, Patrícia Moraes da Matta Campbell



DESCRIÇÃO MORFOLÓGICA

- 2Ap 0-25 cm, 0-15 cm, **preto (7,5YR 2,5/1, úmida e 7,5YR 4/2, seca)**; argiloarenosa; fraca pequena, granular; consistência firme, plástica e não pegajosa; transição ondulada e abrupta.
- 2Bi 25-65 cm, 15-80 cm, **cinzento-brunado (7,5YR 4/6, úmida e 7,5YR 6/6, seca)**; argiloarenosa; pouco cascalhenta, fraca pequena média, blocos subangulares, consistência muito friável, não plástica e não pegajosa; transição ondulada e abrupta.
- 2Cr 65-80 até 160 cm, **cinzento-brunado (7,5YR 4/1, úmida e 7,5YR 8/1, seca)**; areia franca; cascalhenta, maciça, consistência solta, não plástica e não pegajosa.

RAÍZES: Presença de raízes comuns e finas no horizonte 2Bi; comuns, médias e finas no horizonte Ap; raras, finas comuns e médias no horizonte Cr.

OBSERVAÇÕES: Contato lítico a 160 cm; presença de matacões com mais de 25 cm de diâmetro, a 70 cm.

Perfil 4

DESCRIÇÃO GERAL

Classificação SiBCS: Cambissolo Háptico tb distrófico

Localização, município, estado e coordenadas: Localidade de São José, no município de Iconha/ES, coordenadas 301927E e 7706030N.

Situação, declive e cobertura vegetal sobre o perfil: Descrito e coletado em barranco de corte de estrada, sob pastagem.

Altitude: 490 m

Litologia: Metatonalito

Material originário: Intemperismo do material de origem

Pedregosidade: Ligeiramente Pedregosa

Rochosidade: Não Rochosa

Relevo local: Forte ondulado

Relevo regional: Montanhoso

Erosão: Ligeira- laminar

Drenagem: Bem drenado

Vegetação primária: Floresta subperenifólia

Uso atual: Pastagem

Clima: Aw, da classificação de Koppen

Descrito por: Márcio Rocha Francelino, Patrícia Morais da Matta Campbell



DESCRIÇÃO MORFOLÓGICA

Não foi feita descrição morfológica. Horizontes observados: Ap, AB, Bi1, Bi2, Bi3

RAÍZES: Comuns, finas, médias no Ap; Raras e finas nos demais

Perfil 5

DESCRIÇÃO GERAL

Classificação SiBCS: Cambissolo Háplico tb distrófico

Localização, município, estado e coordenadas: Localidade de São José, no município de Iconha/ES, coordenadas 300920E e 7706147N

Situação, declive e cobertura vegetal sobre o perfil: Descrito e coletado em barranco de corte de estrada, sob consórcio de café e banana

Altitude: 396 m

Litologia: Metatonalito

Material originário: Intemperismo do material de origem

Pedregosidade: Ligeiramente pedregosa

Rochosidade: Não rochosa

Relevo local: Forte ondulado

Relevo regional: Montanhoso

Erosão: Forte- laminar

Drenagem: Bem drenado

Vegetação primária: Floresta subperenifólia, tropical

Uso atual: Consórcio de café e banana

Clima: Aw, da classificação de Koppen

Descrito por: Márcio Rocha Francelino, Patrícia Morais da Matta Campbell



DESCRIÇÃO MORFOLÓGICA

Ap 0-15 cm, **preto-brunado (7,5YR 3/1, úmida e 7,5YR 5/1, seca)**; argiloarenosa; fraca pequena, granular; consistência friável, não plástica e não pegajosa; transição plana e abrupta.

BA 0-25 cm, **preto-brunado (7,5YR 3/1, úmida e 7,5YR 5/1, seca)**; argiloarenosa; fraca pequena, granular, consistência friável, não plástica e não pegajosa; transição ondulada e clara.

Bi1 25-65 cm, **bruno-forte (7,5YR 5/6, úmida e 7,5YR 4/6 seca)**; argilosiltosa; fraca pequena média, blocos subangulares, consistência firme, não plástica e pegajosa; transição ondulada e gradual.

Bi2 65-100 cm, **bruno-forte (7,5YR 5/6, úmida e 7,5YR 4/6 seca)**; franco argilosiltosa; pouco cascalhenta, fraca pequena, blocos subangulares, consistência firme, não plástica e ligeiramente pegajosa; transição plana e gradual.

Bi3 100-180+ cm, **bruno-forte (7,5YR 5/6, úmida e 7,5YR 4/6 seca)**; franco-argiloarenosa; fraca muito pequena, blocos subangulares, consistência firme, não plástica e ligeiramente pegajosa.

RAÍZES: Raras e finas no horizonte Bi3; Poucas e finas no horizonte Bi2 e raras, finas nos horizontes BA e Bi1.

OBSERVAÇÕES: O horizonte Ap foi coletado a 10 m do perfil.

Perfil 6

DESCRIÇÃO GERAL

Classificação SiBCS: Latossolo Amarelo

Localização, município, estado e coordenadas:

Estrada principal, sentido Inhaúma, município de Iconha/ES, coordenadas 302166E e 7704572N

Situação, declive e cobertura vegetal sobre o perfil: Descrito e coletado em barranco de corte de estrada, terço médio, sob consórcio de café e banana.

Altitude: 380 m

Litologia: Metatonalito

Material originário: Intemperismo do material de origem

Pedregosidade: Não pedregosa

Rochosidade: Não rochosa

Relevo local: Forte ondulado

Relevo regional: Montanhoso

Erosão: Forte- laminar

Drenagem: Bem drenado

Vegetação primária: Floresta subperenifólia, tropical

Uso atual: Consórcio de café e banana

Clima: Aw, da classificação de Koppen

Descrito por: Márcio Rocha Francelino, Patrícia Morais da Matta Campbell



DESCRIÇÃO MORFOLÓGICA

BAp 0-25 cm, **bruno-forte (7,5YR 3/4, úmida e 7,5YR 5/4 seca)**; argiloarenosa; moderada pequena média, blocos subangulares; consistência friável, plástica e ligeiramente pegajosa; transição plana e gradual.

- Bw1 25-55 cm, **bruno (7,5YR 4/6, úmida e 7,5YR 5/6 seca)**; franco argiloarenosa; fraca média, blocos subangulares, consistência firme, ligeiramente plástica e não pegajosa; transição plana e gradual.
- Bw2 55-130 cm, **bruno (7,5YR 4/6, úmida e 7,5YR 5/8 seca)**; argilosiltosa; fraca pequena média, blocos subangulares, consistência firme, ligeiramente plástica e ligeiramente pegajosa; transição plana e difusa.
- Bw3 130-200+ cm, **bruno (7,5YR 4/6, úmida e 7,5YR 5/8 seca)**; argilosiltosa; fraca média grande, blocos subangulares, consistência firme, ligeiramente plástica e ligeiramente pegajosa.

RAÍZES: Raras e grossas no horizonte BA; Médias e raras nos horizontes B1 e B2; Raras e finas em todos exceto no B3.

OBSERVAÇÕES: Pedoatividade moderada (mesofauna).

Perfil 7

DESCRIÇÃO GERAL

Classificação SiBCS: Latossolo Amarelo

Localização, município, estado e coordenadas:

Localidade de São José, no município de Iconha/ES, coordenadas 301721E e 7705405N.

Situação, declive e cobertura vegetal sobre o perfil: Descrito e coletado em barranco de corte de estrada, terço superior, sob interface pastagem/mata.

Altitude: 570 m

Litologia: Metatonalito

Material originário: Intemperismo do material de origem

Pedregosidade: Não pedregosa

Rochosidade: Não rochosa

Relevo local: Forte ondulado

Relevo regional: Montanhoso

Erosão: Ligeira- laminar

Drenagem: Bem drenado

Vegetação primária: Floresta subperenifólia

Uso atual: Pastagem/mata

Clima: Aw, da classificação de Köppen

Descrito por: Márcio Rocha Francelino, Patrícia Morais da Matta Campbell



DESCRIÇÃO MORFOLÓGICA

Colúvio 0-15 cm, **bruno (7,5YR 4/6, úmida e 7,5YR 5/6 seca)**; franco-argiloarenosa; fraca pequena muito pequena, blocos subangulares; consistência friável, não plástica e não pegajosa; transição irregular e abrupta.

2Bw1 15-70 cm, **bruno (7,5YR 4/6, úmida e 7,5YR 5/8 seca)**; franco-argilossiltosa; moderada pequena, blocos subangulares, consistência firme, ligeiramente plástica e ligeiramente pegajosa; transição irregular e abrupta.

2Ab1/2Bw2 70-85 cm, **bruno-forte (7,5YR 3/4, úmida e 7,5YR 5/4 seca)**; argiloarenosa; fraca pequena, blocos subangulares, consistência friável, não plástica e ligeiramente pegajosa; transição irregular e abrupta.

3Bw2 85-140 cm, **bruno (7,5YR 4/6, úmida e 7,5YR 5/8 seca)**; franco argilosiltosa; fraca pequena que se desfaz em muito pequena, consistência friável, plástica e ligeiramente pegajosa; transição irregular e clara.

4Ab2 140-195+ cm, **bruno (7,5YR 4/4, úmida e 7,5YR 6/4 seca)**; argiloarenosa; fraca, pequena, blocos subangulares, consistência friável, plástica e ligeiramente pegajosa.

OBSERVAÇÕES: Sem horizonte A, no 4Ab2 presença de grande quantidade fragmento de carvão vegetal.

Perfil 8

DESCRIÇÃO GERAL

Classificação SiBCS: Cambissolo Háptico tb distrófico

Localização, município, estado e coordenadas: Município de Iconha/ES, coordenadas 294418E e 7703763N.

Situação, declive e cobertura vegetal sobre o perfil: Descrito e coletado em barranco de corte de estrada, sob consórcio de café com banana.

Altitude: 581 m

Litologia: Metatonalito

Material originário: Intemperismo do material de origem

Pedregosidade: Pedregosa

Rochosidade: Não Rochosa

Relevo local: Ondulado

Relevo regional: Montanhoso

Erosão: Forte- laminar

Drenagem: Bem drenado

Vegetação primária: Floresta subperenifólia

Uso atual: Pastagem

Clima: Aw, da classificação de Koppen

Descrito por: Márcio Rocha Francelino, Patrícia Morais da Matta Campbell

SEM IMAGEM

DESCRIÇÃO MORFOLÓGICA

Não foi feita descrição morfológica. Contato lítico a 140 cm.

RAÍZES: Raras e médias

Perfil 9

DESCRIÇÃO GERAL

Classificação SiBCS: Cambissolo Háplico

Localização, município, estado e coordenadas:

Localidade de Alto Inhaúma, no município de Iconha/ES, coordenadas 299440E e 7702784N

Situação, declive e cobertura vegetal sobre o perfil: Descrito e coletado em barranco de corte de estrada, terço médio, sob pastagem

Altitude: 735 m

Litologia: Metatonalito

Material originário: Intemperismo do material de origem

Pedregosidade: Ligeiramente pedregosa

Rochosidade: Não rochosa

Relevo local: Forte ondulado

Relevo regional: Montanhoso/Escarpado

Erosão: Moderado- laminar

Drenagem: Bem drenado

Vegetação primária: Floresta subperenifólia

Uso atual: Pastagem

Clima: Aw, da classificação de Koppen

Descrito por: Márcio Rocha Francelino, Patrícia Morais da Matta Campbell



DESCRIÇÃO MORFOLÓGICA

- Ap 0-17 cm, **preto (7,5YR 2,5/1 úmida e 7,5YR 4/1 seca)**; franco-argiloarenosa; moderada pequena, granular; consistência firme, ligeiramente plástica e não pegajosa; transição plana e clara.
- AB 17-39 cm, **bruno-escuro (7,5YR 3/3, úmida e 7,5YR 4/2 seca)**; franco-argiloarenosa; fraca média que se desfaz em granular média, blocos subangulares, consistência firme, ligeiramente plástica e ligeiramente pegajosa; transição plana e gradual.
- BA 39-65 cm, **bruno (7,5YR 4/6, úmida e 7,5YR 5/6 seca)**; franco-argilosiltosa; pouco cascalhente, moderada pequena média, blocos subangulares, consistência firme, plástica e ligeiramente pegajosa; transição ondulada e gradual.

- Bi1 65-110 cm, **bruno (7,5YR 4/6, úmida e 7,5YR 5/6 seca)**; franco argilosiltosa; pouco cascalhenta, fraca pequena media, blocos angulares/subangulares, consistência firme, plástica e ligeiramente pegajosa; transição ondulada e gradual.
- Bi2 110-200+ cm, **bruno (7,5YR 4/6, úmida e 7,5YR 5/6 seca)**; argilosiltosa; pouco cascalhenta, fraca media, blocos subangulares, consistência friável, ligeiramente plástica e ligeiramente pegajosa.

RAÍZES: Raras e finas nos horizontes Bi1 e Bi2; Comuns e finas nos horizontes AB e BA; Finas, comuns e médias no horizonte Ap.

OBSERVAÇÕES: Presença comum de material primário na massa do solo (>4%)

Perfil 10

DESCRIÇÃO GERAL

Classificação SiBCS: Latossolo Amarelo distrófico
Localização, município, estado e coordenadas: Localidade de Alto Inhaúma, no município de Iconha/ES, coordenadas 299153E e 7701944N
Situação, declive e cobertura vegetal sobre o perfil: Descrito e coletado em barranco de corte de estrada, terço superior, sob capineira.
Altitude: 720 m
Litologia: Metatonalito
Material originário: Intemperismo do material de origem
Pedregosidade: Ligeiramente pedregosa
Rochosidade: Não rochosa
Relevo local: Suave ondulado
Relevo regional: Montanhoso
Erosão: Moderado- laminar
Drenagem: Bem drenado
Vegetação primária: Floresta subperenifólia
Uso atual: Capineira
Clima: Aw, da classificação de Koppen
Descrito por: Márcio Rocha Francelino, Patrícia Morais da Matta Campbell



DESCRIÇÃO MORFOLÓGICA

- Ap 0-20 cm, **preto (7,5YR 2,5/1, úmida e 7,5YR 5/1 seca)**; franco-arenosa; fraca pequena, granular; consistência firme, não plástica e não pegajosa; transição plana e gradual.

- AB 20-30 cm, **preto (7,5YR 2,5/1, úmida e 7,5YR 4/6 seca)**; argiloarenosa; fraca pequena, blocos subangulares, consistência friável, não plástica e não pegajosa; transição ondulada e gradual.
- Bw1 30-80 cm, **bruno (7,5YR 4/6, úmida e 7,5YR 5/6 seca)**; argilosiltosa; moderada média, blocos subangulares, consistência firme, plástica e ligeiramente pegajosa; transição plana e gradual.
- Bw2 80-120 cm, **bruno (7,5YR 4/6, úmida e 7,5YR 5/6 seca)**; franco argilosiltosa; fraca pequena, blocos subangulares, consistência friável, ligeiramente plástica e ligeiramente pegajosa; transição ondulada e clara.
- BC 120-130 cm, **bruno (7,5YR 4/6, úmida e 7,5YR 5/6 seca)**; franco-argiloarenosa;, fraca muito pequena, blocos subangulares, consistência friável, ligeiramente plástica e não pegajosa, transição ondulada e gradual.
- C 130-200 cm, **cinzento-brunado (7,5YR 4/1, úmida e 7,5YR 8/1 seca)**; areia franca; maciça, consistência muito friável.

RAÍZES: Comuns e finas nos horizontes C, AB,A ; Poucas e finas nos horizontes B1, B2 e AB

OBSERVAÇÕES: O horizonte BC se desfaz em grão simples, presença de pequenos fragmentos de pedra, cascalhos de 5 cm. Os horizontes B2 e Ap se desfazem em muito pequeno.

Perfil 11

DESCRIÇÃO GERAL

Classificação SiBCS: Cambissolo Háptico tb distrófico

Localização, município, estado e coordenadas: Município de Iconha/ES, coordenadas 300589E e 7707148N

Situação, declive e cobertura vegetal sobre o perfil: Descrito e coletado em barranco de corte de estrada, terço médio, sob pastagem.

Altitude: 435 m

Litologia: Metatonalito

Material originário: Intemperismo do material de origem

Pedregosidade: Ligeiramente pedregosa

Rochosidade: Não rochosa

Relevo local: Ondulado

Relevo regional: Montanhoso

Erosão: Moderado- laminar

Drenagem: Bem drenado

Vegetação primária: Floresta subperenifólia

Uso atual: Pastagem

Clima: Aw, da classificação de Köppen

Descrito por: Márcio Rocha Francelino, Patrícia Morais da Matta Campbell



DESCRIÇÃO MORFOLÓGICA

- A 0-15 cm, **bruno-escuro (7,5YR 3/3, úmida e 7,5YR 6/3 seca)**; franco-arenosa; fraca média, granular; consistência firme, não plástica e não pegajosa; transição plana e clara.
- AB 15-35 cm, **bruno (7,5YR 4/4, úmida e 7,5YR 5/4 seca)**; argiloarenosa; fraca media, blocos subangulares, consistência muito firme, ligeiramente plástica e ligeiramente pegajosa; transição plana e gradual.
- Bi1 35-80 cm, **bruno (7,5YR 4/6, úmida e 7,5YR 5/6 seca)**; franco-argilosiltosa; moderada média, blocos subangulares, consistência firme, ligeiramente plástica e ligeiramente pegajosa; transição plana e clara.
- Bi2 80-130 cm, **bruno-escuro (7,5YR 5/8, úmida e 7,5YR 6/8 seca)**; franco argiloarenosa; moderada média, blocos angulares, consistência firme, ligeiramente plástica e ligeiramente pegajosa; transição plana e clara.
- C 130-170+ cm

OBSERVAÇÕES: Aterro em função de terraceamento, pedregoso nos horizontes C e B2, não foi descrito o horizonte C.

Perfil 12

DESCRIÇÃO GERAL

Classificação SiBCS: Latossolo vermelho

Localização, município, estado e coordenadas: Município de Iconha/ES, coordenadas 300268E e 7707259N

Situação, declive e cobertura vegetal sobre o perfil: Descrito e coletado em barranco de corte de estrada, terço médio, sob floresta secundária.

Altitude: 500 m

Litologia: Metatonalito

Material originário: Intemperismo do material de origem

Pedregosidade: Não pedregosa

Rochosidade: Não rochosa

Relevo local: Montanhoso

Relevo regional: Montanhoso

Erosão: Ligeira- laminar

Drenagem: Bem drenado

Vegetação primária: Floresta subperenifólia

Uso atual: Floresta secundária

Clima: Aw, da classificação de Koppen

Descrito por: Márcio Rocha Francelino, Patrícia Morais da Matta Campbell



DESCRIÇÃO MORFOLÓGICA

- A 0-5 cm, **bruno-escuro (7,5YR 3/3, úmida e 7,5YR 6/3 seca)**; franco-arenosa; fraca pequena, granular; consistência firme, não plástica e não pegajosa; transição ondulada e clara.
- BA 5-35 cm, **bruno-avermelhado (5YR 4/6, úmida e 5YR 6/6 seca)**; argiloarenosa; fraca pequena, blocos subangulares, consistência muito firme, ligeiramente plástica e ligeiramente pegajosa; transição ondulada e gradual.
- Bw1 35-90 cm, **bruno-avermelhado (5YR 4/6, úmida e 5YR 5/6 seca)**; argilosiltosa; fraca pequena média, blocos angulares, consistência firme, não plástica e ligeiramente pegajosa; transição plana e gradual.
- Bw2 90-160 cm, **bruno-avermelhado (5YR 4/6, úmida e 5YR 5/6 seca)**; argilosiltosa; fraca média, blocos angulares, consistência firme, não plástica e ligeiramente pegajosa; transição plana e difusa.
- Bw3 160-185+ cm, **bruno-avermelhado (5YR 4/6, úmida e 5YR 5/6 seca)**; argilosiltosa; fraca pequena, blocos angulares, consistência firme, não plástica e pegajosa.

Amostra extra 1

DESCRIÇÃO GERAL

Classificação SiBCS: GleissoloHáplico

Localização, município, estado e coordenadas:

Localidade de Alto Inhaúma, Município de Iconha/ES, coordenadas 299127E e 7701849N

Situação, declive e cobertura vegetal sobre o perfil: Descrito e coletado em mini-trincheira, fundo de vale, sob pastagem.

Altitude: 500 m

Litologia: Metatonalito

Material originário: Colúvio

Pedregosidade: Não pedregosa

Rochosidade: Não rochosa

Relevo local: Plano

Relevo regional: Forte ondulado

Erosão: Moderada- sulcos

Drenagem: Mal drenado

Vegetação primária: Floresta subperenifólia

Uso atual: Floresta secundária

Clima: Aw, da classificação de Koppen

Descrito por: Márcio Rocha Francelino, Patrícia Morais da Matta Campbell

SEM IMAGEM

DESCRIÇÃO MORFOLÓGICA

Não foi realizada.

Amostra extra 2

DESCRIÇÃO GERAL

Classificação SiBCS: Neossolo Litólico

Localização, município, estado e coordenadas: Município de Iconha/ES, coordenadas 299328E e 7703782N

Situação, declive e cobertura vegetal sobre o perfil: Descrito e coletado em barranco de corte de estrada, terço médio, sob consórcio de café e banana.

Altitude: 596 m

Litologia: Metatonalito

Material originário: Intemperismo do material de origem

Pedregosidade: Extremamente pedregosa

Rochosidade: Extremamente rochosa

Relevo local: Montanhoso

Relevo regional: Ondulado

Erosão: Muito forte- laminar

Drenagem: Bem drenado

Vegetação primária: Floresta subperenifólia

Uso atual: Café e banana.

Clima: Aw, da classificação de Koppen

Descrito por: Márcio Rocha Francelino, Patrícia Morais da Matta Campbell



DESCRIÇÃO MORFOLÓGICA

- 2C 0-25 cm, **bruno-escuro (7,5YR 3/3, úmida e 7,5YR 5/3, seca)**; argilosiltosa; fraca muito pequena, blocos subangulares; consistência friável, ligeiramente plástica e ligeiramente pegajosa; transição quebrada e abrupta.
- 3C 50-110 cm, **bruno-escuro (7,5YR 3/4, úmida e 7,5YR 4/2, seca)**; franco-argiloarenosa; fraca pequena, blocos subangulares, consistência firme, ligeiramente plástica e ligeiramente pegajosa.

Amostra extra 3

DESCRIÇÃO GERAL

DESCRIÇÃO GERAL

Classificação SiBCS: Cambissolo Háplico

Localização, município, estado e coordenadas:
Município de Iconha/ES, coordenadas 300735E e 7703619N

Situação, declive e cobertura vegetal sobre o perfil: Descrito e coletado em barranco de corte de estrada, terço médio, sob e café.

Altitude: 502 m

Litologia: Metatonalito

Material originário: Intemperismo do material de origem

Pedregosidade: Ligeiramente pedregosa

Rochosidade: Extremamente rochosa

Relevo local: Forte ondulado

Relevo regional: Montanhoso

Erosão: Forte- laminar

Drenagem: Bem drenado

Vegetação primária: Floresta subperenifólia

Uso atual: Café.

Clima: Aw, da classificação de Koppen

Descrito por: Márcio Rocha Francelino, Patrícia Morais da Matta Campbell



DESCRIÇÃO MORFOLÓGICA

Não foi realizada.