

GUSTAVO VIEIRA VELOSO

**DISTRIBUIÇÃO ESPACIAL DE METAIS E METALÓIDES NOS SOLOS DO
ESTADO DE MINAS GERAIS UTILIZANDO GEOESTATÍSTICA E MÉTODOS
DE APRENDIZADO DE MÁQUINA**

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

VIÇOSA
MINAS GERAIS – BRASIL
2017

**Ficha catalográfica preparada pela Biblioteca Central da Universidade
Federal de Viçosa - Câmpus Viçosa**

T

V443d
2017

Veloso, Gustavo Vieira, 1986-

Distribuição espacial de metais e metalóides nos solos do estado de Minas Gerais utilizando geoestatística e métodos de aprendizado de máquina / Gustavo Vieira Veloso. – Viçosa, MG, 2017.

xi, 82f. : il. (algumas color.) ; 29 cm.

Orientador: Jaime Wilson Vargas de Mello.

Tese (doutorado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. Química do solo. 2. Solos - Metais pesados - Minas Gerais. 3. Solos - Composição. 4. Solos - Análise.

I. Universidade Federal de Viçosa. Departamento de Solos e Nutrição de Plantas. Programa de Pós-Graduação em Solos e Nutrição de Plantas. II. Título.


CDD 22 ed. 631.41

GUSTAVO VIEIRA VELOSO

**DISTRIBUIÇÃO ESPACIAL DE METAIS E METALÓIDES NOS SOLOS DO
ESTADO DE MINAS GERAIS UTILIZANDO GEOESTATÍSTICA E MÉTODOS
DE APRENDIZADO DE MÁQUINA**

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

APROVADA: 13 de julho de 2017.



Gerson Rodrigues dos Santos



José João Lelis Leal de Souza



Elpídio Inácio Fernandes Filho



Walter Antônio Pereira Abrahão
(Coorientador)



Jaime Wilson Vargas de Mello
(Orientador)

À minha família e aos meus amigos, especialmente, meu pai, minha mãe e meu irmão.

“Se o conhecimento pode criar problemas, não é
através da ignorância que podemos solucioná-los”
Isaac Asimov

AGRADECIMENTOS

A Deus.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) a Fundação de Meio Ambiente do Estado de Minas Gerais (Feam), que permitiram a execução desde trabalho.

Aos meus pais pela ajuda durante toda essa fase de capacitação profissional.

À minha família, especialmente meu irmão, pelos ensinamentos e apoio durante essa jornada no curso de graduação e de pós-graduação.

Ao Orientador Jaime Wilson Vargas de Mello pela orientação, confiança, apoio e ajuda.

Ao Coorientador. Walter Antônio Pereira Abrahão, pelo apoio e ajuda nos trabalhos e pela orientação.

Aos Coorientador Nerilson Terra Santos pelo apoio e ideias para melhoria dessa tese.

Aos Professores Elpídio Inácio Fernandes Filho e Jose João Lelis Leal de Souza (Jota Jota) por todo apoio e ensinamentos, sem ao apoio senhores não existiria essa tese e futuros artigos.

Ao Departamento de Solos e Nutrição de Plantas e a Universidade Federal de Viçosa, pela formação acadêmica e pelo crescimento intelectual e profissional e apoio na realização das análises a todos os funcionários do Departamento de Solos e Nutrição de Plantas pelo apoio durante as todas as fases dessa jornada.

Aos meus amigos do Laboratório de Geoprocessamento pelo apoio durante esta fase final do Doutorado.

Ao meu estagiário Yang, pelo pela ajuda durante as análises laboratoriais.

Aos meus amigos Lucas, Juliana, Marconi, Raiza, Fernanda, Rosana, Eliana, Cleverson, Gabriela, Raphael, Marcus, Humberto, Cristhian, Geice pela ajuda e apoio durante o doutorado.

Aos meus colegas do Laboratório de Mecanização pelas conversas e conhecimento adquirido.

Enfim, a todas aquelas pessoas que, direta ou indiretamente, contribuíram para que eu alcançasse este sonho. A todos o meu muito obrigado.

BIOGRAFIA

GUSTAVO VIEIRA VELOSO, filho de Marcos Sebastião Veloso e Maria Madalena Vieira Veloso, nasceu em Montes Claros, estado de Minas Gerais, no dia 29 de Outubro de 1986.

Em maio de 2006, iniciou o curso de Engenharia Agrícola e Ambiental na Universidade Federal de Viçosa (UFV), graduou-se em julho de 2011.

Em Agosto de 2011 iniciou o curso de Mestrado em Engenharia Agrícola na Universidade Federal de Viçosa (UFV), finalizando em julho de 2013.

Em Agosto de 2013 iniciou o curso de Doutorado em Solos e Nutrição de Plantas na Universidade Federal de Viçosa (UFV), finalizando em julho de 2017.

SUMÁRIO

RESUMO	viii
ABSTRACT	x
CAPÍTULO 1	
INTRODUÇÃO GERAL	1
CAPÍTULO 2	
USO DE ALGORITMOS DE APRENDIZAGEM DE MÁQUINAS PARA PREDIÇÃO DE METAIS E METALÓIDES NO ESTADO DE MINAS GERAIS.	4
CAPÍTULO 3	
COMPARAÇÃO DE SOFTWARES PARA ESPACIALIZAÇÃO DE METAIS E METALÓIDES NO ESTADO DE MINAS GERAIS PELA KRIGAGEM ORDINÁRIA	36
CAPÍTULO 4	
COMPARAÇÃO ENTRE MÉTODOS DE ESPACIALIZAÇÃO DE DADOS DE METAIS E METALOIDES PARA ESTADO DE MINAS GERAIS.....	62
CAPÍTULO 5	
CONCLUSÃO GERAL	82

RESUMO

VELOSO, Gustavo Vieira, D.Sc., Universidade Federal de Viçosa, julho de 2017. **Distribuição espacial de metais e metalóides nos solos do estado de Minas Gerais utilizando geoestatística e métodos de aprendizado de máquina.** Orientador: Jaime Wilson Vargas de Mello. Coorientadores: Nerilson Terra Santos e Walter Antônio Pereira Abrahão.

O Estado de Minas Gerais tem uma área de 588.384 km² que corresponde a um percentual de 6,9% do território brasileiro e de 63,5% da região sudeste do Brasil. Minas Gerais apresenta uma grande diversidade geológica. Essa variedade geológica sugere que os solos derivados desses materiais possuem elementos químicos nos mais diferentes teores no Estado. Nesse sentido, o objetivo desta pesquisa foi gerar mapas dos teores de metais e metalóides na camada superficial dos solos do Estado de Minas Gerais, utilizando-se de técnicas de aprendizado de máquina e krigagem Ordinária e comparar os resultados dos métodos. Os mapas de teores de 13 elementos traço (Al, As, B, Ba, Cd, Co, Cr, Cu, Fe, Hg, Mn, Mo, Ni, Pb, Sr, V e Zn) foram gerados a partir da Krigagem Ordinária desses elementos e de 5 outros constituintes do solo. Foi utilizado neste estudo um total de 648 amostras de solos georreferenciadas obtidas do 'Banco de Solos' do Estado de Minas Gerais. A seleção das variáveis para a predição do teor de metais e metalóides seguiu procedimentos que visaram identificar as covariáveis com maior relevância, sendo usados métodos correlação linear e não linear. Para identificar a dependência espacial dos dados, foram realizados os seguintes testes de modelos teóricos de semivariograma: esférico, exponencial, estável e gaussiano. Seis algoritmos de predição dos teores foram usados neste estudo para o aprendizado de máquina, sendo estes: Cubist, SVMRadialSigma, Random Forest, ExtraTree, Ranger e Rborist. A verificação do desempenho desses algoritmos foi executada pela validação cruzada. Os resultados da seleção de covariáveis para o aprendizado de máquinas mostram coeficiente de determinação do algoritmo de treinamento (R^2) utilizado na seleção de covariáveis variaram de 0,03 a 0,43. As covariáveis categóricas litologia, geomorfologia e classe de solo apresentaram a maior importância na predição de 17 elementos. Covariáveis bioclimáticas foram importantes na predição dos teores de 16 elementos. Enquanto que, As covariáveis relacionadas à gamaespectrometria auxiliaram na predição dos teores de dez dos elementos analisados. Por sua vez, a magnetometria apresentou baixa importância para a predição do teor da maioria dos elementos nos solos, sendo importante somente para predição do Sr. Em geral, a transformação dos dados não promoveu um aumento na capacidade de predição dos elementos pelos modelos de aprendizado de máquinas, exceto para três elementos (Al, Sr e Zn). Na Krigagem

Ordinária houver aumento do número de elementos com boa capacidade preditiva em alguns dos métodos de krigagem (Al, As, Co, Hg, Mn, Sb, Se, Sr e Zn). OS resultados de nRMSE variam entre 10% e 20%, obtendo resultados considerados bons, exceto por B e Mo. Os melhores resultados de nRMSE foram obtidos na predição com os dados sem transformações. A transformação dos dados em log1p gerou melhores resultados de treinamento apenas para Al e Zn. Para os elementos com os melhores resultados no método de aprendizado de máquinas foram: *Ranger*, *SVMRadialSigma*, *RandomFlorest*, *Extratree*. Os algoritmos *Cubist* e *Rborist* não alcançaram desempenho satisfatório na predição para nenhum elemento analisado. Os mapas espacializados pelos métodos de krigagem e aprendizado de máquinas apresentaram características similaridades para os elementos avaliados, com estimativas dos teores próximos. Os mapas especializados mostram maiores teores de As, Cd, Cr e Ni estão localizados na região do Quadrilátero Ferrífero, enquanto os elementos Cu, Fe, Mn, V e Zn apresentaram os maiores teores na região do triângulo mineiro. O elemento Co apresentou teores altos nas duas regiões, Quadrilátero Ferrífero e Triângulo mineiro. Os resultados mostraram que 11 elementos demonstraram melhores pelos algoritmos de aprendizado de máquinas (Al, As, B, Co, Fe, Hg, Mn, Ni, Pb, Se e Zn). A espacialização dos teores dos metais e metaloides pelos algoritmos de aprendizado de máquinas, gerou mapas de 11 elementos com desempenhos superiores à krigagem. Os mapas gerados pelo método de espacialização por krigagem apresentam teores máximos superiores aos apresentados todos os elementos tirando o Pb. Os métodos de aprendizado de máquinas geram mapas com maior nível de detalhamento quando comparados com mapas gerados pela krigagem.

ABSTRACT

VELOSO, Gustavo Vieira, D.Sc., Universidade Federal de Viçosa, July, 2017. **Space distribution of metals and metalloids in the soils of the state of minas gerais using geostatistics and machine learning methods.** Adviser: Jaime Wilson Vargas de Mello. Co-Advisers: Nerilson Terra Santos and Walter Antônio Pereira Abrahão.

The State of Minas Gerais has an area of 588,384 km² which corresponds to a percentage of 6.9% of Brazilian territory and 63.5% of the southeastern region of Brazil. Minas Gerais presents a wide lithological diversity, which suggests that the soils derived from these materials have large variation of chemical elements content in this State. In this sense, the goal of this research was to generate prediction maps of metals and metalloids contents in the surface layer of the soils of the State of Minas Gerais, using techniques of machine learning and Ordinary kriging and compare the results of these methods. The maps of 13 trace elements (Al, As, B, Ba, Cd, Co, Cr, Cu, Fe, Hg, Mn, Mo, Ni, Pb, Sr, V, Zn) was generated from Ordinary kriging of these elements and other soil constituents. A total of 648 geo-referenced soil samples obtained from the 'Soil data bank' of the State of Minas Gerais, were used in this study. The selection of variables for the prediction of metal and metalloid contents followed procedures that aimed to identify covariates with great relevance, using linear and nonlinear correlation methods for removal of collinearity. After confirmation of data spatial dependence, the following tests of theoretical semivariogram models, such as: spherical, exponential, stable and Gaussian were performed. Six predicting algorithms of machine learning were used in this study: Cubist, SVMRadialSigma, Random Forest, ExtraTree, Ranger and Rborist. The algorithms performance was verified by cross validation. The results of the covariates selection by machine learning showed coefficient of the training algorithm (R²) used in the selection of covariates ranged from 0.03 to 0.43. The categorical covariates lithology, geomorphology and soil class presented the highest importance in the prediction of 17 elements. Bioclimatic covariates were important in predicting the contents of 16 elements. Meanwhile, covariates related to gamma spectrometry helped to predict the contents of ten elements analyzed. In addition, magnetometry was low importance for the content prediction of most elements in soils, being important only for the prediction of Sr. In general, the transformation of the data did not promote an increase in the prediction capacity of the elements at using machines learning, except for three elements (Al, Sr, Zn). With Ordinary kriging there was an increase in the number of elements with "good" predictive capacity in some of the kriging methods (Al, As, Co, Hg, Mn, Sb, Se, Sr and Zn). In general, the best results (nRMSE) range from 10% to 20%. These results are

considered good, except for B and Mo. The best results of nRMSE were obtained for data, without transformations. The best results using machine learning method were obtained with the algorithms: Ranger, SVMRadialSigma, Random Forest and Extratree. The algorithms Cubist and Rborist did not reach satisfactory performance in the prediction for any element analyzed. The spatialized maps by kriging and machine learning methods presented similar characteristics for the evaluated elements, with estimates of the next levels. The Cu, Fe, Mn, V and Zn elements presented the highest levels in the “Triângulo Mineiro” region. The element Co, presented high levels in the two regions, “Quadrilátero Ferrífero” and “Triângulo Mineiro”. The spatialization of metal and metalloid contents by machine learning algorithms generated 11 elements with higher performance than kriging. The maps generated by the kriging spatialisation method have higher maximum levels than all the elements showing Pb. Machine learning methods generate maps with a greater level of detail when compared to maps generated by kriging.

CAPÍTULO 1

INTRODUÇÃO GERAL

O Estado de Minas Gerais tem uma área de 588.384 km² que corresponde a um percentual de 6,9% do território brasileiro e de 63,5% da região sudeste do Brasil (FJP, 2013). Minas Gerais apresenta uma grande diversidade geológica, representada por uma sequência de idades que vão desde a arqueana à cenozoica, que ocorreram em condições tectônicas e metamórficas das mais variadas (MACHADO e SILVA; 2010). Essa variedade sugere que os solos derivados desses materiais possuem elementos químicos nos mais diferentes teores no Estado. Com a finalidade de avaliar o teor e distribuição em solos no Estado, De Souza (2013), gerou mapas dos teores dos principais metais e semimetais de ocorrência natural em solos superficiais, a partir de coletas e aplicação de estatística espacial. Este trabalho surgiu da necessidade espacializar os teores naturais e entender mais sobre distribuição de metais e metaloides pelo Estado.

O background geoquímico pode ser definido como o teor de um elemento químico encontrado em uma área não antropizada. O background geoquímico dos solos reflete a dinâmica dos fatores e processos de formação do solo, e por isso varia no tempo e no espaço (MATSCHULLAT et al., 2000). A espacialização dos teores dos elementos químicos em áreas diversas como os Estado de Minas Gerais pode ser utilizada para definição de regiões homogêneas e entender melhor sobre a distribuição do elemento em todo o Estado, além de prevenir e identificar áreas de concentração acima do limite para segurança dos seres vivos. Atualmente, a espacialização dos dados é feita utilizando técnicas de krigagem ou regressão krigagem (ZHENG et al., 2008; LIU et al., 2006). Outra metodologia de espacialização de variáveis pedológicas são os métodos de aprendizado de máquinas (HENGL et al., 2017; HENDERSON et al., 2005; AHMAD et al., 2010; KOVAČEVIĆ et al., 2010; FORKUOR et al, 2017; BRUNGARD et al., 2015). O aprendizado de máquinas necessita de covariáveis para que seja feito o treinamento e espacialização dos dados e, atualmente, ainda existem poucos trabalhos na espacialização de metais e metaloides em solos (WILFORD et al., 2016; HAO et al., 2016; AYOTTE et al, 2016; KIRKWOOD et al, 2016; PENG Yi et al, 2016). As covariáveis especializadas podem se encontradas gratuitamente junto empresas e órgãos públicos nacionais e internacionais com a Companhia de Pesquisa de Recursos Naturais (CPRM), Empresa Brasileira de Pesquisa Agropecuária (Embrapa), Nacional Aeronautics and Space Administration (NASA), Companhia de

Desenvolvimento Econômico de Minas gerais (CODEMIG), etc. Essas covariáveis apresentam correlação com a formação de solos e com os teores de metais e metalóides.

O estudo objetivou gerar mapas de predição dos teores de metais e metalóides na camada superficial dos solos do Estado de Minas Gerais, por meio do uso de técnicas de aprendizado de máquina e krigagem ordinária. Os objetivos específicos foram:

- a) Selecionar melhores covariáveis para predição de metais e metalóides em solos no estado utilizando métodos de máquinas do tipo árvore de decisão e suporte vetor máquina.
- b) Avaliar os dois softwares de krigagem ordinária para predição dos teores de metais e metalóides em solos do Estado de Minas Gerais.
- c) Avaliar o melhor preditor entre as técnicas de krigagem e os métodos de aprendizado de máquinas para todos os elementos traço em solos do Estado.

Estes objetivos foram separados em três capítulos que seguem a ordem a cima.

REFERENCIA BIBIOGRAFICA

AHMAD, S.; KALRA, A.; STEPHEN, H. .Estimating soil moisture using remote sensing data: A machine learning approach. **Advances in Water Resources**, v. 33, n. 1, p. 69-80, 2010.

AYOTTE, J. D.; NOLAN, B. T.; GRONBERG, J. A.. Predicting Arsenic in Drinking Water Wells of the Central Valley, California. **Environmental Science & Technology**, v. 50, n. 14, p. 7555-7563, 2016.

BRUNGARD, C. W. et al. Machine learning for predicting soil classes in three semi-arid landscapes. **Geoderma**, v. 239, p. 68-83, 2015.

DE SOUZA, J. J. L. L. et al. Geochemistry and spatial variability of metal (loid) concentrations in soils of the state of Minas Gerais, Brazil. **Science of the Total Environment**, v. 505, p. 338-349, 2015.

FUNDAÇÃO JOÃO PINHEIRO. *Perfil demográfico do Estado de Minas Gerais - 2002*. Belo Horizonte: Centro de Estatística e Informações, 2003.

FORKUOR, G. et al. High Resolution Mapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso: A Comparison of Machine Learning and Multiple Linear Regression Models. **PloS one**, v. 12, n. 1, p. e0170478, 2017.

HAO, L. et al. Spatial distribution and sources of trace elements in surface soils, Changchun, China: Insights from stochastic models and geostatistical analyses. **Geoderma**, v. 273, p. 54-63, 2016.

HENGL, T. et al. SoilGrids250m: Global gridded soil information based on machine learning. **PLoS one**, v. 12, n. 2, p. e0169748, 2017.

HENDERSON, B. L. et al. Australia-wide predictions of soil properties using decision trees. **Geoderma**, v. 124, n. 3, p. 383-398, 2005.

KIRKWOOD, C. et al. A machine learning approach to geochemical mapping. **Journal of Geochemical Exploration**, v. 167, p. 49-61, 2016.

KOVAČEVIĆ, M.; BAJAT, B.; GAJIĆ, B. Soil type classification and estimation of soil properties using support vector machines. **Geoderma**, v. 154, n. 3, p. 340-347, 2010.

LIU, X.; WU, J.; XU, J.. Characterizing the risk assessment of heavy metals and sampling uncertainty analysis in paddy field by geostatistics and GIS. **Environmental Pollution**, v. 141, n. 2, p. 257-264, 2006.

MATSCHULLAT, Jörg; OTTENSTEIN, R.; REIMANN, Clemens. Geochemical background—can we calculate it? **Environmental geology**, v. 39, n. 9, p. 990-1000, 2000.

MACHADO, M. F.; SILVA, S. F. **Geodiversidade do estado de Minas Gerais**. 2010.

PENG, Y. et al. Digital Mapping of Toxic Metals in Qatari Soils Using Remote Sensing and Ancillary Data. **Remote Sensing**, v. 8, n. 12, p. 1003, 2016.

WILFORD, J.; DE CARITAT, P.; BUI, E.. Predictive geochemical mapping using environmental correlation. **Applied Geochemistry**, v. 66, p. 275-288, 2016.

ZHENG, Y.; CHEN, .; HE, J.. Multivariate geostatistical analysis of heavy metals in topsoils from Beijing, China. **Journal of Soils and Sediments**, v. 8, n. 1, p. 51-58, 2008.

CAPÍTULO 2

USO DE ALGORITMOS DE APRENDIZAGEM DE MÁQUINAS PARA PREDIÇÃO DE METAIS E METALÓIDES NO ESTADO DE MINAS GERAIS.

INTRODUÇÃO

O Estado de Minas Gerais possui uma área de 588.384 km², o que corresponde a um percentual de 6,9% do território brasileiro e de 63,5% da região sudeste do Brasil (FJP, 2013). Este Estado apresenta uma grande diversidade geológica, com idades que vão desde o Arqueano, cerca de 3,8 bilhões anos atrás, até o Holoceno, compreendendo os últimos 11,7 mil anos, os quais incorreram em condições tectônicas e metamórficas variadas (MACHADO e SILVA, 2010).

Os elementos químicos As, Ag, Cd, Cr, Pb, Hg, Se, Sb são considerados altamente tóxicos, mesmo quando presentes em teores relativamente baixos. Por sua vez, os elementos B, Co, Cr, Cu, Fe, Mn, Mo, Ni, e Zn são essenciais para plantas e animais em baixas concentrações (CEMPEL e NIKEL, 2006; GÖHRE e PASZKOWSKI, 2006; MALAFAIA et al., 2014).

A política nacional de gerenciamento de áreas contaminadas, Resolução nº 420 de 28 de dezembro de 2009, tem o objetivo de promover a conservação da qualidade dos solos. Para tanto, o Conselho Nacional do Meio Ambiente (CONAMA) definiu os critérios para obtenção dos valores orientadores da qualidade de solo, quanto à presença de substâncias químicas potencialmente tóxicas. Nesta resolução, ficou estabelecido que cada Estado é responsável pela determinação dos seus Valores de Referência de Qualidade (VRQ) para solos.

O VRQ é o teor de uma substância ou elemento químico potencialmente tóxico que ocorre nos solos sem contribuição antrópica (CONAMA, 2009). No Estado de Minas Gerais, os VRQ's dos solos foram determinados e encontram-se divulgados na Resolução nº 166 de 2011 do Conselho Estadual de Política Ambiental de Minas Gerais (COPAM).

Atualmente, o programa “Solos de Minas” contém 648 amostras de solos georreferenciadas. A partir dessas amostras foram criados mapas de teores para

elementos traço, que funcionam como referência para estudos futuros nas áreas de Geoquímica, Engenharia Ambiental, Biologia entre outras ciências (DE SOUZA et al., 2015). Nesse sentido, estudos utilizaram a técnica da krigagem ordinária estimaram 18 variáveis no Estado de Minas Gerais, sendo 13 elementos (Al, As, B, Ba, Cd, Co, Cr, Cu, Fe, Hg, Mn, Mo, Ni, Pb, Sr, V e Zn) e 5 outros constituintes do solo (fósforo extraível, capacidade de troca catiônica, matéria orgânica do solo, soma de base e teor de argila) para determinar os padrões de concentração e distribuição espacial desses elementos.

A determinação dos teores naturais de metais e metaloides em solos de áreas não antropizadas é de grande importância para o gerenciamento ambiental de áreas contaminadas por esses elementos. Parte destes elementos presentes nos solos está associada à constituição de rochas, sendo que, solos com mesmo material de origem podem apresentar diferentes teores de metais pesados. (FADIGAS et al., 2002; OLIVEIRA e COSTA, 2004; OLIVEIRA e JUCÁ, 2004; BIONDI et al., 2011; PELOZATO et al., 2011). Os teores de metais e metaloides são dependentes de uma variedade de fatores, tais como: o grau de intemperismo e da lixiviação (MCBRIDE, 1994)

O solo é o produto da interação entre material parental, clima, relevo, organismos e tempo (JENNY, 1941), conseqüentemente estes fatores influenciam nos seus teores de metais e metaloides. Diferentes materiais parentais configuram solos com diferentes teores de elementos (GRAY et al., 2016; WILFORD, et al., 2016). O sistema radicular de espécies vegetais é outro importante fator que altera a distribuição espacial de nutrientes e metais pesados em solo (JOBÁGY & JACKSON, 2004). A intensidade dos processos morfogenéticos é ponderada pela exposição da litologia e o aprofundamento do manto de intemperismo.

Os métodos atuais de aprendizagem de máquinas têm se apresentado como importantes instrumentos para estudos pedológicos (HENGL et al., 2017; HENDERSON et al., 2005; AHMAD et al., 2010; KOVAČEVIĆ et al., 2010; FORKUOR et al., 2017; BRUNGARD et al., 2015). Nesse sentido, o método de aprendizagem de máquinas necessita de covariáveis para o aprendizado e predição dos resultados, enquanto que a krigagem requer a dependência espacial das variáveis para realizar a predição. Apesar do ganho do poder preditivo da possibilidade de explorar variáveis categóricas e numéricas que refletem fatores e processos de formação de solos, ainda são poucos os trabalhos desenvolvidos para predição de teores de metais e metaloides em solos que utilizam o método de aprendizado de máquinas (WILFORD

et al., 2016; HAO et al., 2016; AYOTTE et al., 2016; KIRKWOOD et al., 2016; PENG. Y et al., 2016). Este trabalho objetivou selecionar covariáveis e avaliar os algoritmos de aprendizagem de máquinas Cubist, SVMRadialSigma, RandomFlorest, ExtraTree, Ranger e Rborist nas predições de metais e metaloides para os solos do Estado de Minas Gerais.

MATERIAL E MÉTODOS

Banco de Dados

Foram utilizadas 648 amostras de solos georreferenciadas obtidas do Banco de Solos do Estado de Minas Gerais (Figura 1). Estas amostras foram coletadas entre 0-20 cm de profundidade. O processo de amostragem e quantificação dos elementos está descrito em De Souza et al (2015) e Abrahão e Marques (2013). Na Tabela 1, encontra-se a estatística descritiva dos elementos avaliados:

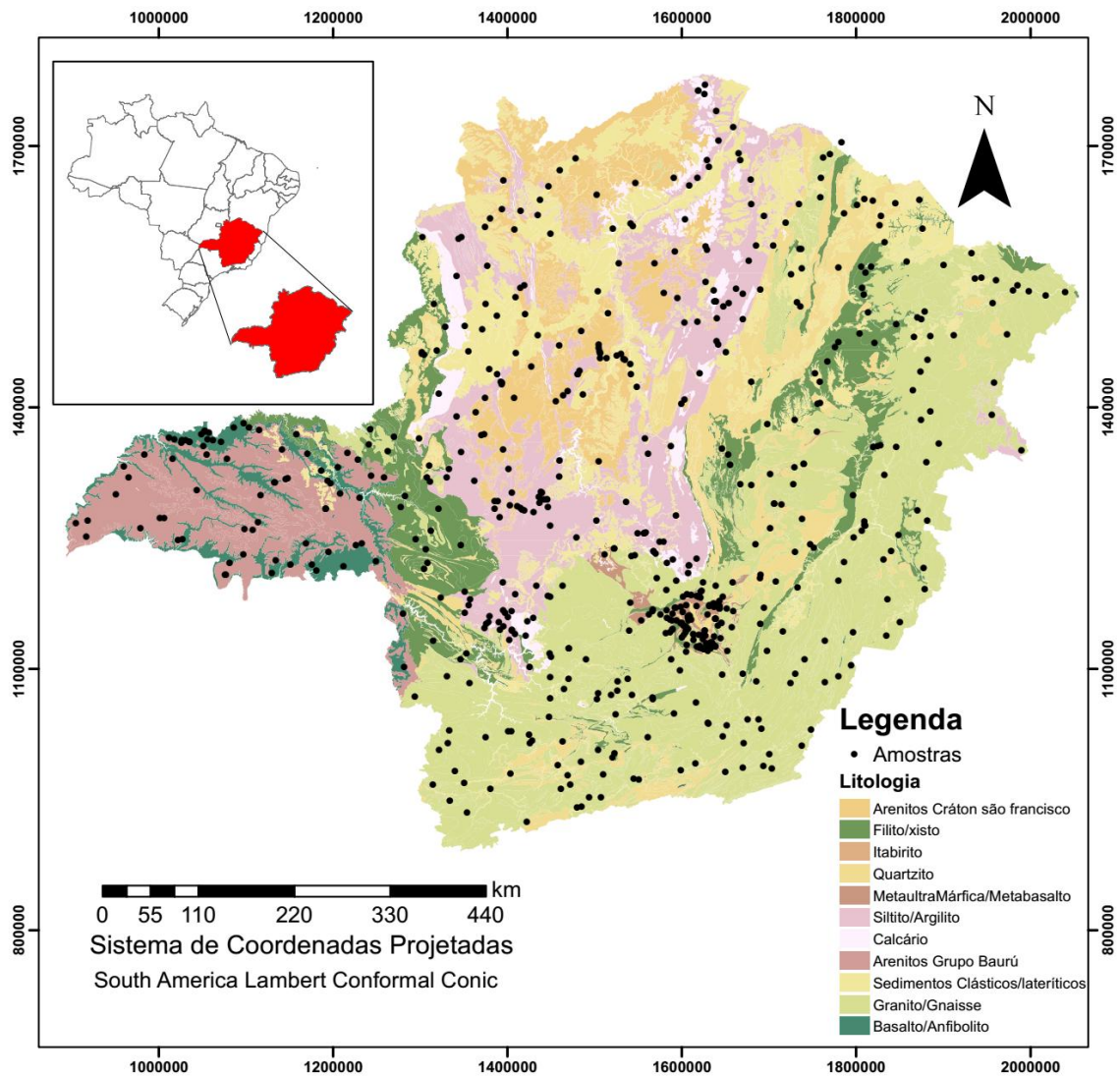


Figura 1. Distribuição de pontos coletados e mapa litológico simplificado da área estudo
(Fonte: Modificado de Codemig 2014).

O distribuição litológica de Minas Gérias apresenta algumas peculiaridades. Na região norte e noroeste apresentam maior proporção de argilitos, arenitos e siltitos e sedimentos. No sul, sudoeste e nordeste contem predomínio de rochas acidades com granito e ganisses com algumas extrusões de rochas metamórficas (filito e xisto). A região oeste predominam as rochas areníticas, com a presença de basalto e anfibolito nos vales dos rios. A região central possui a maior variação litológica dentre as regiões avaliadas, apresentando no centro leste principalmente, contendo proporção maior de rochas metamórficas com itabirito, filito, xisto e quartzito (Figura 1).

Tabela 1 – Estatísticas descritivas das amostras de solos na cama de 0-20 cm , para os teores de metais e metaloides;

	Número de amostras	Média	Desvio Padrão	CV**	Mediana	Máximo	Mínimo
Elemento	*	mg/kg	mg/kg	%	mg/kg	mg/kg	mg/kg
Al	647	5,55x10 ⁴	4.11x10 ⁴	74,02	4.70x10 ⁴	2,51x10 ⁵	850,99
As	576	16,17	58,13	359,49	4,26	854,51	0,13
B	647	7,71	6,74	87,42	6,67	43,64	0,28
Ba	647	99,82	272,09	272,58	46,62	4472,50	0,72
Cd	576	2,64	4,35	164,50	1,06	56,82	0,18
Co	647	11,52	19,97	173,36	5,27	227,12	0,221
Cr	647	115,38	192,01	166,41	71,13	2940,16	33,59
Cu	647	47,34	51,02	107,77	32,31	388,05	0,20
Fe	647	4,62x10 ⁴	3,40x10 ⁴	73,64	39377,91	2,11x10 ⁵	506,23
Hg	576	0,22	0,66	299,20	0,05	8,59	0,01
Mn	647	418,33	974,46	232,94	172,60	1,56x10 ⁴	0,77
Mo	647	2,47	2,23	90,33	2,36	16,91	0,46
Ni	647	29,04	42,51	146,40	18,12	591,13	0,50
Pb	576	20,19	17,68	87,58	16,70	183,33	1,30
Sb	559	1,22	4,85	399,43	0,31	81,86	0,00
Se	559	0,45	0,56	124,97	0,28	4,62	0,02
Sr	629	17,62	55,85	317,02	6,54	815,88	0,02
V	647	123,79	118,25	95,53	81,73	705,00	2,45
Zn	647	43,96	38,07	86,60	33,90	392,70	1,00

*Adimensional

** Coeficiente de Variação

Seleção de Variáveis

Em função do enfoque exploratório do trabalho, as análises de precisão se iniciaram com 36 covariáveis no formato matricial com resolução (célula) de 1 km², sendo 33 covariáveis numéricas (Tabela 2) e três categóricas.

A escolha das covariáveis levou em consideração os tipos de solos presentes na área de estudo, além dos principais fatores responsáveis pelo processo pedogenético, sendo estes: o material de origem, a morfologia local, os fatores climáticos e a vegetação. Para essa última covariável, foi escolhido mapa de vegetação do mês de agosto por apresentarem a menor influência de sombreamento por nuvens e maior diferenciação entre vegetações que sofrem ou não o processo de senescência. O mapa de geomorfologia foi confeccionado seguindo Silva e Rodrigues (2009), modificado para se adequar as características e tamanho da área de estudo.

Tabela 2 – Covariáveis utilizadas neste trabalho

Covariáveis	
Covariáveis Numéricas (33)	
Grupo/ Fonte	Covariáveis
Bioclimáticas (HIJMANS et al., 2005)	(Bio 1) – Temperatura anual (Bio 2) – Variação da temperatura diurna (Bio 3) – Isotermia (Bio 4) – Sazonalidade da temperatura (Bio 5) - Temperatura máxima do mês mais quente (Bio 6) – Temperatura do mês mais frio (Bio 7) – Variação da temperatura anual (Bio 8) – Temperatura média do triênio mais úmido (Bio 9) – Temperatura média do triênio mais seco (Bio 10) – Temperatura média do triênio mais quente (Bio 11) – Temperatura média do triênio mais frio (Bio 12) – Precipitação anual (Bio 13) – Precipitação mês mais úmido (Bio 14) – Precipitação mês mais seco (Bio 15) – Variação da precipitação anual (Bio 16) – Precipitação do trimestre mais úmido (Bio 17) – Precipitação do trimestre mais seco (Bio 18) – Precipitação do trimestre mais quente (Bio 19) – Precipitação do trimestre mais frio
Gamaespectrometria (CODEMIG,2014)	Gama U equivalente (eU) Gama Th equivalente (eTh) Gama K Gama Th equivalente/K (eTh/K) Gama U equivalente/Th equivalente (eTh/K) Gama U equivalente/K (eU/K) Gama ternário
Índice de vegetação da diferença normalizada do mês de agosto (CARROLL, 2004),	NVDI
Magnetometria (CODEMIG, 2014)	Magnetometria Primeira derivada vertical Magnetometria Intensidade total Magnetometria Sinal Analítico
Gravimetria (CODEMIG, 2014).	Gravimetria
Modelo Digital de Elevação (FARR et al , 2007).	MDE
Covariáveis categóricas (3)	
Litologia simplificada (CODEMIG, 2014),(Figura 1)	Litologia
Classes de solo (UFV et al., 2010, adaptado),	Solos
Geomorfologia (Figura 2)	Geomorfologia

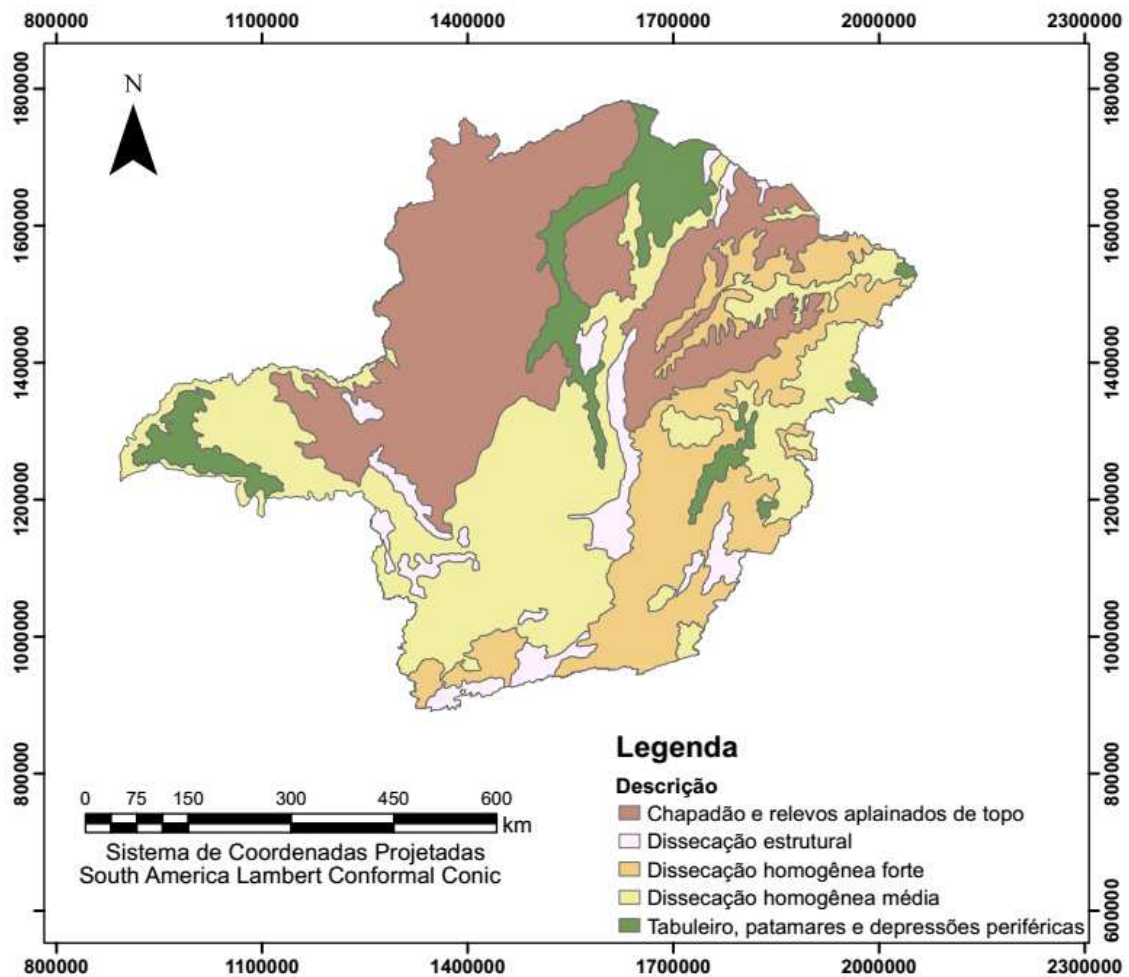


Figura 2. Mapa de geomorfologia

A seleção das variáveis para a predição do teor de metais e metalóides seguiu procedimentos que visaram identificar as covariáveis com maior relevância (Figura 3). Primeiramente, a redundância foi minimizada pela retirada das covariáveis numéricas com alta correlação. Caso duas covariáveis apresentassem correlação acima de 0,95, foi removida a variável com menor valor de importância para prever o elemento pelo algoritmo *Random Forest*. Este processo foi aplicado a todos os pares de variáveis utilizadas neste trabalho. Na segunda etapa foi realizada a redução do número de covariáveis, para obter os modelos mais simples e com menor número de preditores (modelos parcimoniosos). Neste processo foi utilizado a ferramenta Eliminação RFE (*RecursiveFeatureElimination*) contida no pacote 'caret' (KUHNS 2008). Após essas análises foram obtidos os coeficientes de determinação (R^2) para cada modelo entre os dados previstos e observados. Foi escolhido o menor número de covariáveis que obteve 97% do maior valor de R^2 .

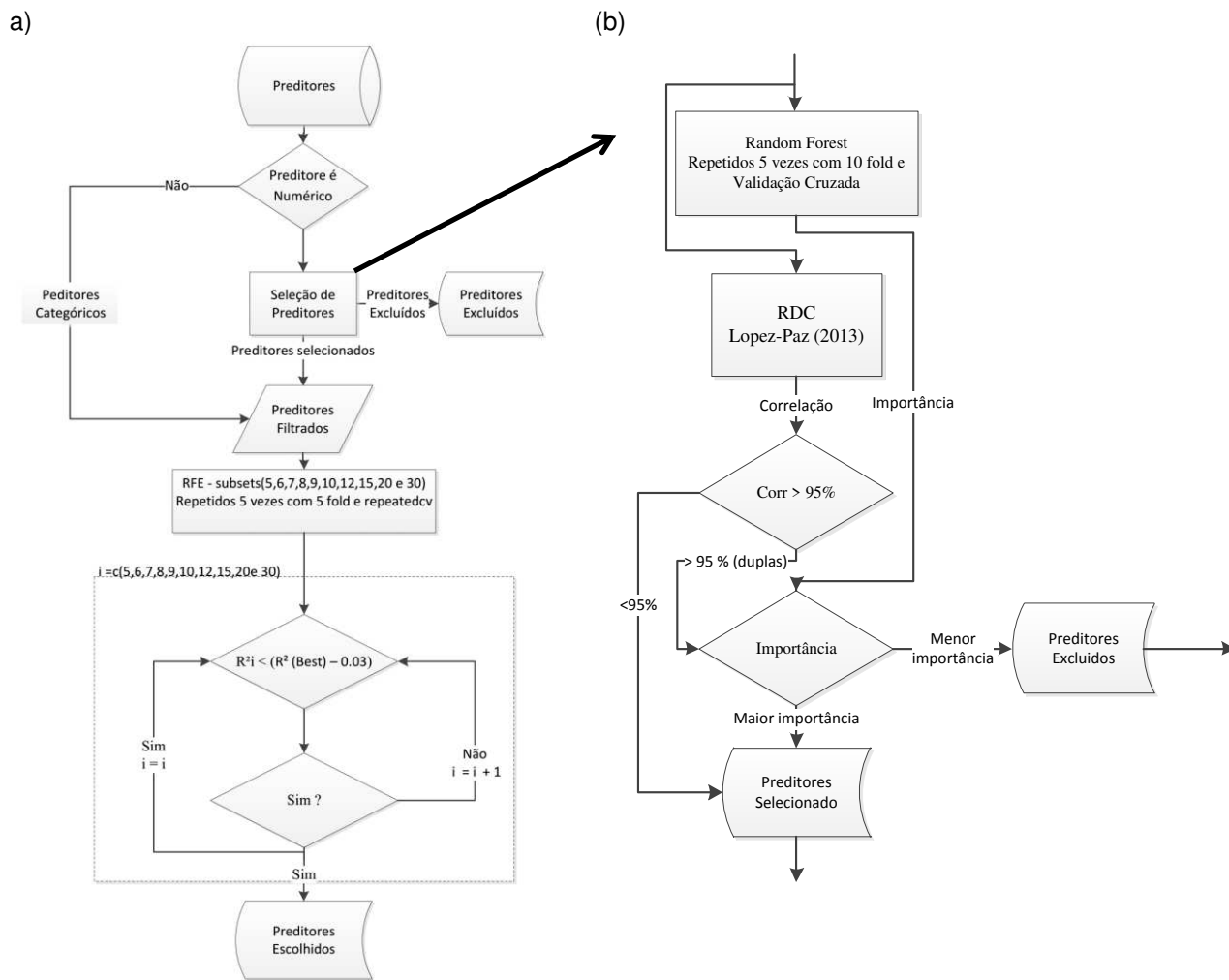


Figura 3 – Fluxograma usado para seleção de variáveis

Exemplo subprocesso de seleção de variáveis:
 Preditores A, B e C

Importância (Random Forest)
 Imp A: 3,12
 Imp B: 2,32
 Imp C: 2,31

Corr (RDC)	A	B	C
A	1	0,96	0,94
B	0,96	1	0,90
C	0,94	0,90	1

Correlação CA (AC) <0,95 e BC (CD) <0,95 =>
 C =>Preditor Escolhido

Correlação > 0,95 =>A e B (Importância)
 Importância B < A, assim:
 B =>Preditor excluído
 A =>Preditor escolhido

Treinamento e predição dos mapas

O treinamento dos algoritmos e a espacialização dos dados foram realizados a partir do número de preditores selecionados na mineração de dados (Figura 3) para cada elemento a ser predito (\hat{Y}). Além da realização do treinamento sobre os dados primários, os valores de Y foram transformados segundo as funções $\log(Y+1)$ e raiz quadrada de Y , para todos os elementos analisados (FU et al., 2009; CHIAP et al., 1999; VAYSSE e LAGACHERIE, 2015; SCHWAIGHOFER et al, 2009).

Seis algoritmos de predição foram usados neste estudo, sendo estes: Cubist, SVMRadialSigma, RandomFlorest, ExtraTree, Ranger e Rborist. A verificação do desempenho desses algoritmos foi executada pela validação cruzada com repetição. Os algoritmos utilizados estão disponíveis no pacote “*caret*” no software R (KUHN, 2008).

Alguns dos algoritmos avaliados caracterizam-se por serem do tipo árvore de decisão são incapazes de prever os valores além dos intervalos de treinamento predefinidos, tais como: ExtraTree, RandomForest, Ranger e Rborist (ALVAREZ, 2009; HANSSON et al., 2016; HILL et al., 2014). Para prever os teores dos elementos de forma mais próximas da condição real, foram adicionados valores extremos para cada elemento nos grupos utilizados no treinamento e validação dos algoritmos avaliados.

A performance dos algoritmos de predição foi avaliada a partir de uma separação aleatória entre amostras que constituem o conjunto de treinamento e de validação. O desempenho dos algoritmos foi comparado pelo coeficiente de determinação para o modelo (R^2) (EQ. 1) (YANG et al., 2016) sendo calculador por:

$$R^2 = \frac{\sum_{i=1}^n (P_i - \overline{Om})^2}{\sum_{i=1}^n (Om_i - \overline{Om})^2} \quad (\text{EQ. 1})$$

Onde n é o número de observações, Om_i são os teores dos elementos observados nas transformações utilizadas (quando essas transformações forem feitas), \overline{Om} é a média dos valores observados e P_i são teores dos elementos previstos pelos algoritmos em cada grupo. Com os resultados preditos pelos modelos foi calculada a raiz quadrada do erro quadrado médio (EQ. 2) (HALLAK e PEREIRA FILHO, 2011; MEHTATALO et al., 2007; AHMED et al., 2015; GUO et al., 2015).

$$RMSE = \left[\frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2 \right]^{1/2} \quad (\text{EQ. 2})$$

Foi calculado o valor da raiz quadrada do erro quadrático médio normalizado (nRMSE) (LOAGUE e GREEN 1991, LOPATIN et al., 2016). Este valor foi calculado com o valor do RMSE, dividido pela amplitude amostral e seu valor é expresso em percentagem (EQ. 3). O nRMSE é utilizado quando se compara diferentes amostras com diferentes variâncias amostrais ou grupo de amostras com diferente amplitudes (HAN et al., 2003; COOK e KRISHNAN, 2015). A predição é considerada “excelente” quando os valores de nRMSE são inferiores a 10%, e predição com valores de nRMSE entre 10% a 20% são consideradas “boas”, e entre 20 a 30% são “regulares” e acima de 30% de nRMSE, a predição é considerada ruim (JAMIESON et al., 1991).

$$NRMSE = \frac{RMSE}{(\max(O_i) - \min(O_i))} \quad (\text{EQ. 3})$$

Os menores resultados de nRMSE dos modelos avaliados e suas transformações foram selecionadas para a confecção dos mapas de predição de cada elemento. Para os elementos que apresentaram os melhores resultados de nRMSE (nRMSE com valores próximos zero), no qual também ocorreu com os dado transformados (log1p e raiz quadrada). Estes dados transformados houve a transformação inversa (back-transformation) para escala real dos teores dos elementos (HENGL et al., 2013). As áreas cobertas por cursos d’água, represas ou zonas urbanas acima da área mínima mapeável não tiveram seus valores preditos apresentados.

As análises e gráficos gerados neste trabalho foram desenvolvidos com o software R e suas bibliotecas Caret, corrplot, Cubist, doParallel, dplyr, e1071, extraTree, ggplot2, gtools, hydroGOF, janitor, kernlab, Metrics, raster, randomForest, ranger, rgdal, e RWeka (KUHN, 2008; WEI.e WEI, 2016;. KUHN et al 2016; ANALYTICS e WESTON, 2014; MEYER et al, 2017; SIMM e ABRIL, 2015; WICKHAM, 2010; WARNES et al., 2015; Zambrano-Bigiarini, 2017; FIRKE et al., 2017; KARATZOGLOU et al., 2005; HAMNER, 2017; HIJMANS et al, 2016; LIAW & WIENER, 2002; WRIGHT.e ZIEGLER, 2015; BIVAND et al., 2017; HORNIK, et al., 2009).

RESULTADOS

Seleção de Variáveis (Mineração de Dados)

Os valores de coeficiente do algoritmo de treinamento (R^2) utilizado na seleção de covariáveis variaram de 0,0346 a 0,4300, estes valores pertencentes aos elementos Hg e V (Tabela 4).

Foram selecionadas no máximo dez covariáveis para a predição da maioria dos elementos (Figura 4). As covariáveis categóricas litologia, geomorfologia e classificação de solo apresentaram a maior importância na predição de 17 elementos, não sendo importante somente para Sb e Sb. A litologia foi a covariável com maior importância para predição da maioria dos teores dos elementos (Al, As, B, Co, Cr, Cu, Fe, Hg, Mn, Ni, Pb, V, Zn), apresentou valor secundário de importância para predição de Ba, Cd, Mo, Sr, e não foi considerada importante para a predição de Sb (Tabela 4). O MDE foi importante para a predição de 11 elementos (Ba, Co, Cr, Fe, Hg, Mn, Pb, Sb, Se, Sr, V), e a Geomorfologia para a predição de dez elementos (Al, As, Ba, Cd, Co, Cr, Mo, Ni, Sr e Zn).

A classificação de solo foi uma covariável importante para predição de seis elementos, Al, Ba, Cu, Fe, Sr e V, e segunda variável com maior importância para Al, Cu e V. O NDVI foi importante predição para seis elementos, Cr, Hg, Mn, Pb, Sb, Zn e Sb foi a variável de maior importância na predição.

A temperatura anual (Bio 1), a sazonalidade da temperatura (Bio 4), a precipitação anual (Bio 12), a precipitação do mês mais úmido (Bio 13), a precipitação do mês mais seco (Bio 14), a variação da precipitação anual (Bio 15), a precipitação do trimestre mais quente (Bio 18) e a precipitação do trimestre mais frio (Bio 19) foram importantes na predição dos teores de 16 elementos.

As covariáveis relacionadas à gamaespectrometria foram importantes na predição dos teores de dez dos elementos analisados (Al, As, Ba, Cd, Co, Mo, Pb, Sb, Sr, Zn). Por sua vez, a magnetometria apresentou baixa importância para a predição do teor da maioria dos elementos em solos, sendo importante somente para predição do Sr.

As covariáveis gama eU/eTh, Gama K/eU, magnetometria e 11 das variáveis bioclimáticas (Bio 2, Bio 3, Bio 5, Bio 6, Bio 7, Bio 8, Bio 9, Bio 10, Bio 11, Bio 16 e Bio 17) não foram selecionadas para as etapas posteriores de processamento dos dados, devido à baixa correlação das covariáveis com o teor de elementos no solo, ou por apresentarem alta correlação com outras covariáveis.



Figura 4 - Seleção das variáveis para os elementos*
 *Nome das variáveis na tabela 2

Em geral, a transformação dos dados não promoveu um aumento na capacidade de predição dos elementos, exceto para Al, Sr e Zn. Os algoritmos *Ranger* e *SvmRadialSigma* predisseram melhor os teores de Al, As, B, Ba, Cd, Co, Cr, Cu, Fe, Hg, Mn, Ni, Pb, Sb e V. Por outro lado, Mo, Se, Sr e Zn foram melhores preditos pelo algoritmo *RandomFlorest e Extratree*.

Os algoritmos *Cubist e Rborist* não alcançando desempenho satisfatório na predição para nenhum elemento analisado. Avaliando os valores de NRMSE, somente o elemento Mo obteve uma predição regular, enquanto que para os outros elementos a predição pode ser considerada boa, de acordo com os critérios estabelecidos por Jamieson et al. (1991).

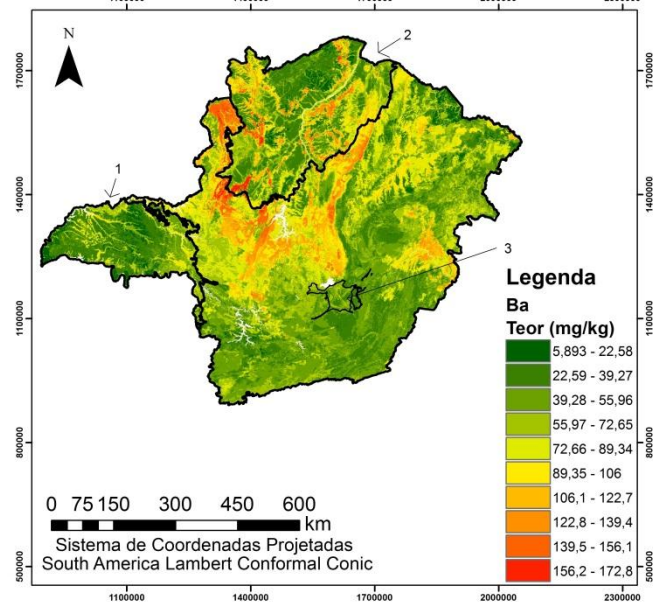
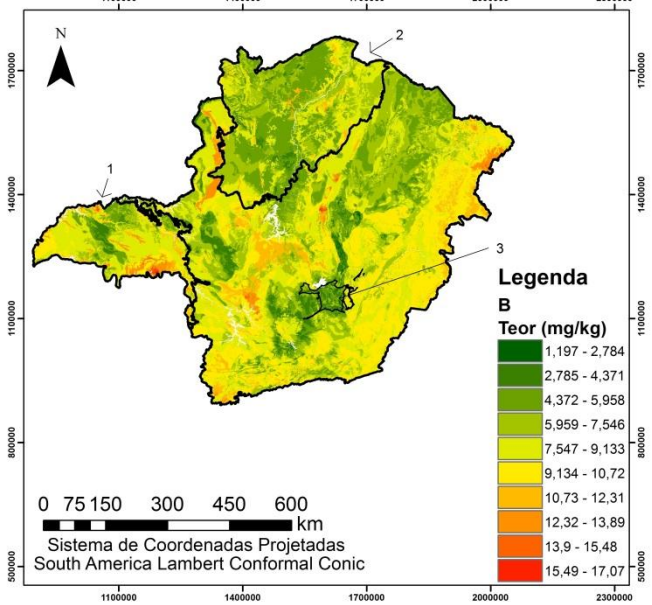
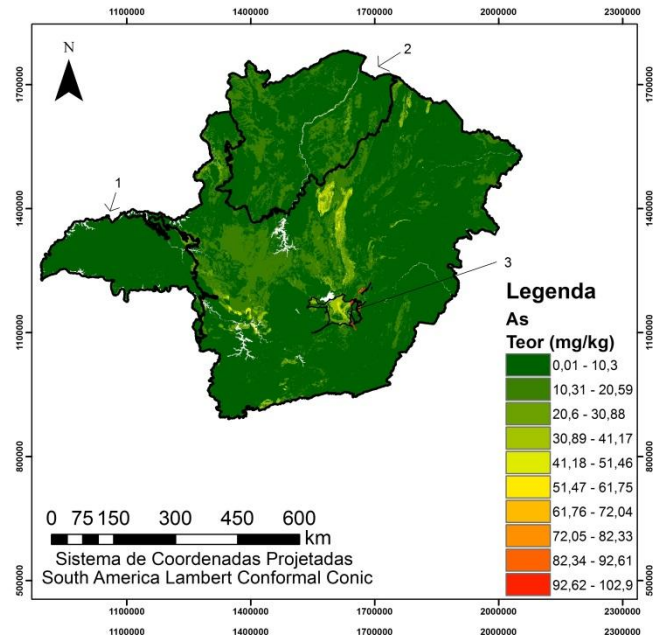
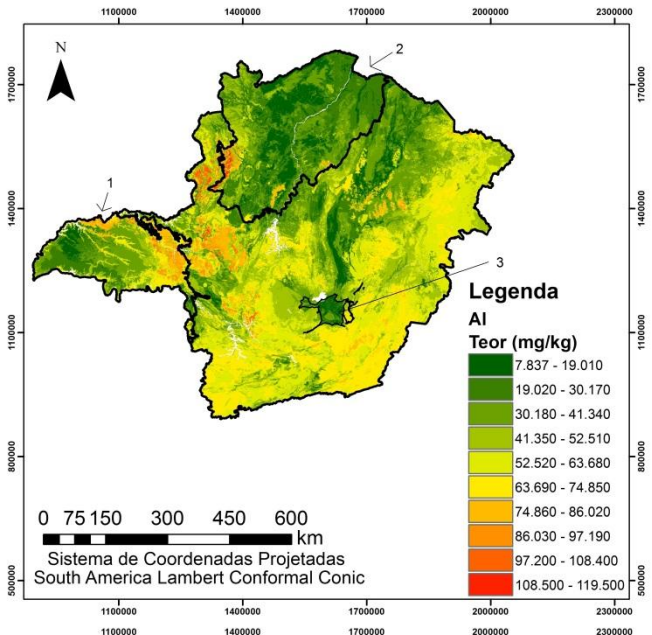
Tabela 5. Sumário dos resultados estatísticos do treinamento dos algoritmos para os elementos avaliados

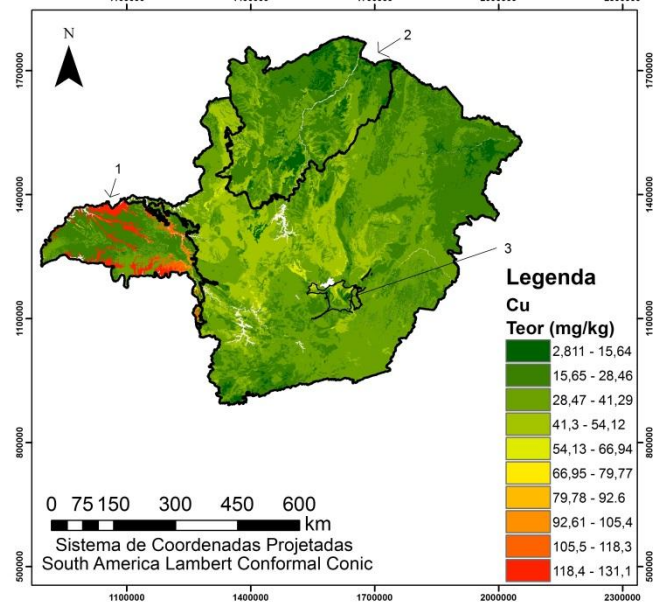
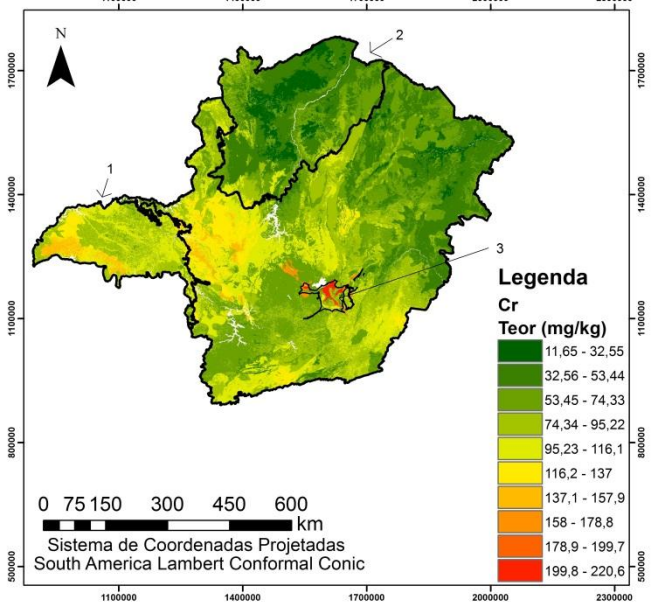
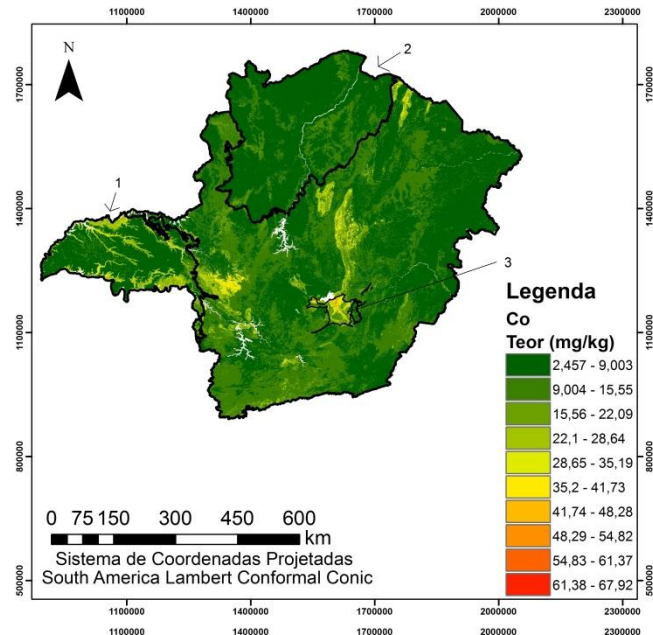
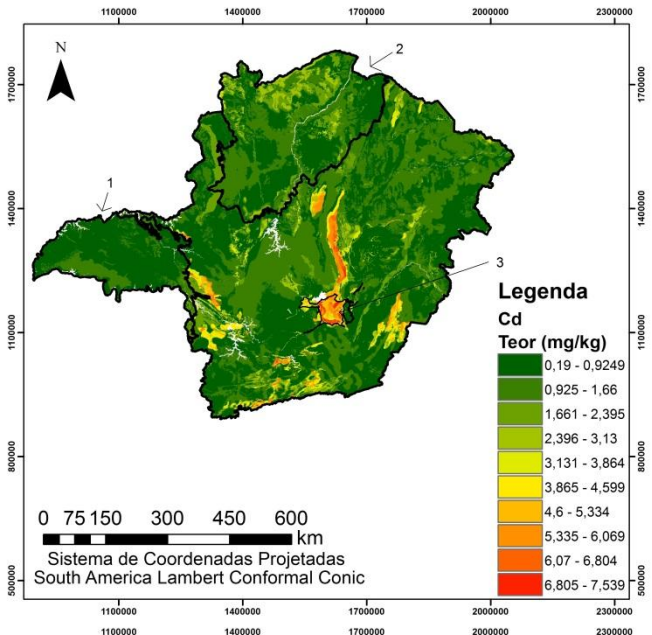
Elemento	Algoritmo	Transformação	R ²	RMSE	nRMSE
Al	Ranger	log1p	0,440	0,658	16,310
As	SvmRadialSigma	sem	0,321	33,641	13,970
B	Ranger	sem	0,223	5,821	22,920
Ba	SsvmRadialSigma	sem	0,109	166,878	16,070
Cd	SvmRadialSigma	sem	0,319	3,102	18,100
Co	Ranger	sem	0,225	15,556	16,520
Cr	SvmRadialSigma	sem	0,184	126,585	17,390
Cu	SvmRadialSigma	sem	0,229	43,445	18,220
Fe	Ranger	sem	0,319	27560,413	16,190
Hg	SvmRadialSigma	sem	0,067	0,489	16,260
Mn	SvmRadialSigma	sem	0,120	652,854	16,570
Mo	Rf	sem	0,248	1,881	22,670
Ni	SvmRadialSigma	sem	0,244	31,900	15,660
Pb	Ranger	sem	0,174	14,593	16,840
Sb	SvmRadialSigma	sem	0,120	2,762	15,920
Se	Rf	sem	0,242	0,466	17,640
Sr	ExtraTrees	sqrt	0,261	2,063	15,940
V	Ranger	sem	0,408	89,775	16,350
Zn	ExtraTrees	log1p	0,288	0,650	17,710

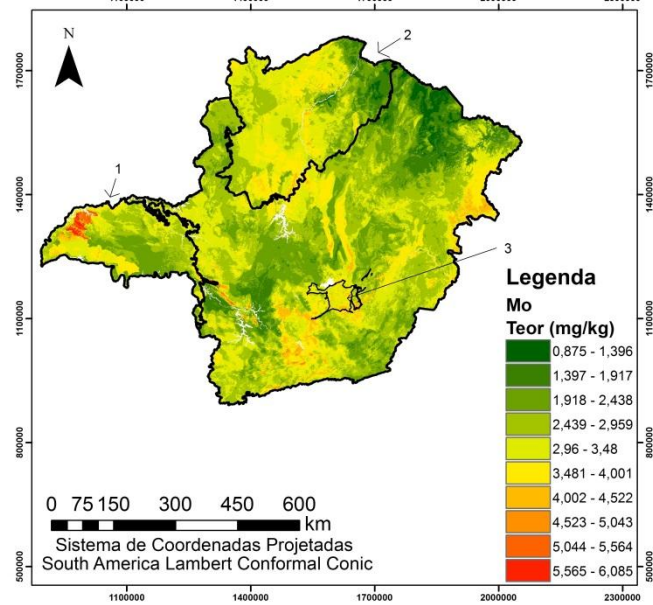
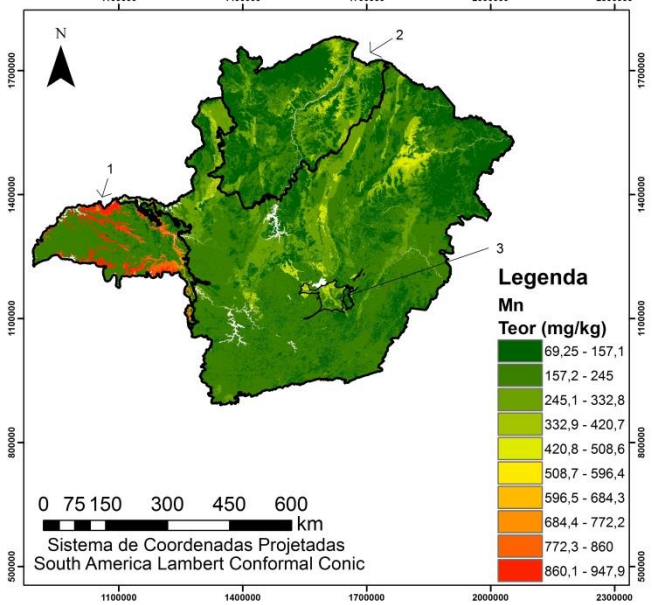
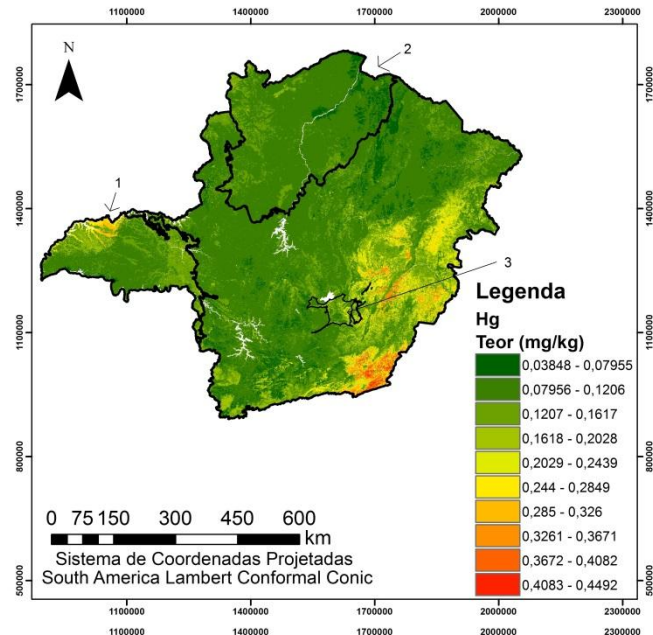
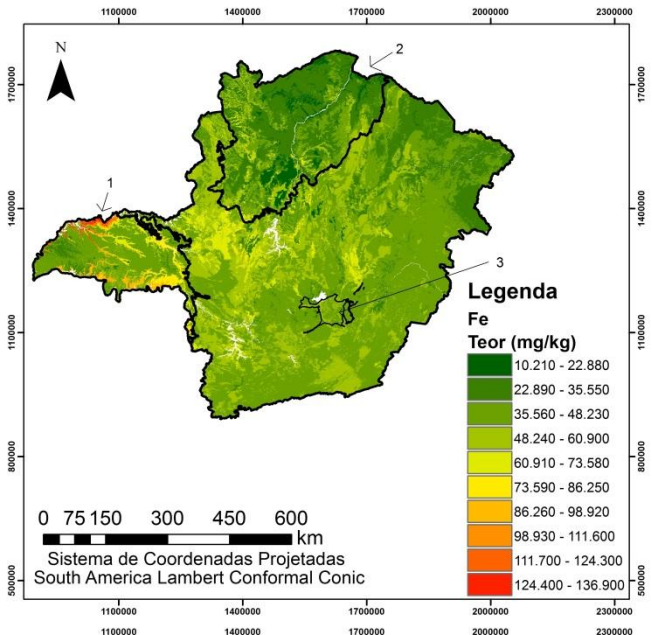
* Transformações: sem: dados sem transformações, log1p: transformações dos dados por $\log(1+Y)$ e sqrt: Transformação dos dados usando raiz quadrada.

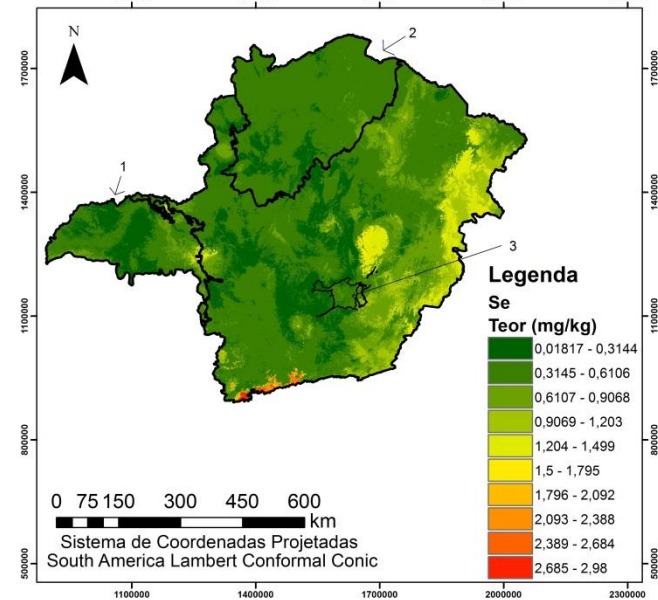
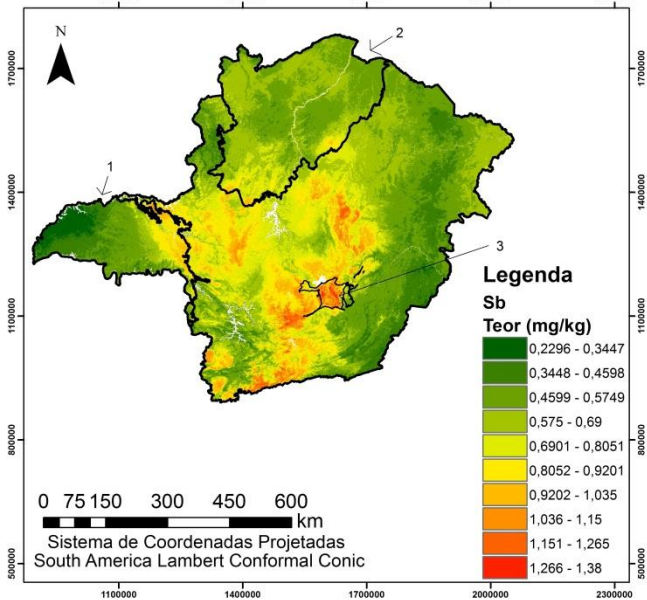
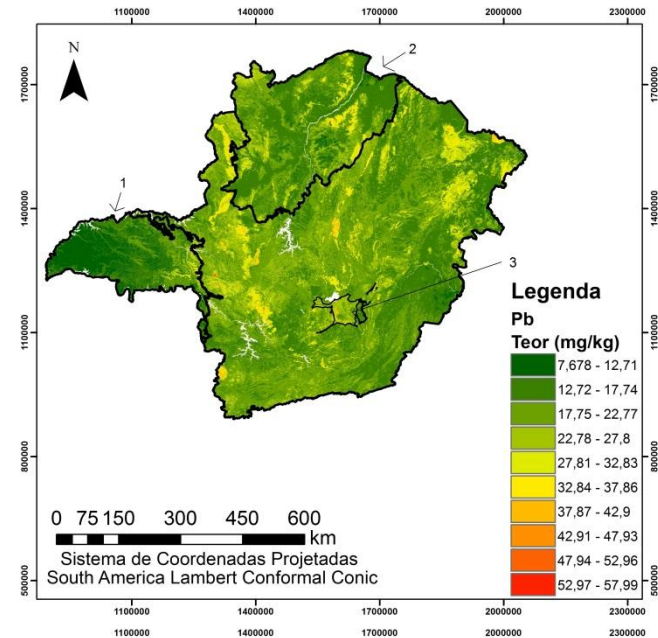
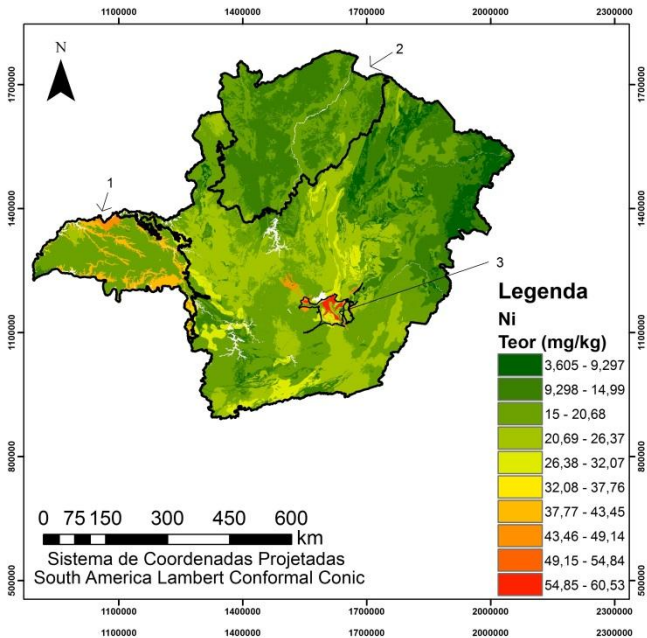
Mapas de predição dos elementos

Os mapas das predições referentes aos melhores algoritmos e elementos avaliados (Figura 5) apresentaram valores máximos e mínimos dentro dos limites superior e inferior de cada elemento (Tabela 1). Os maiores teores de As, Cd, Cr e Ni estão localizados na região do Quadrilátero Ferrífero. Os elementos Cu, Fe, Mn, V e Zn apresentaram os maiores teores na região do triângulo mineiro, enquanto o elemento Co apresentou teores altos nas duas regiões, Quadrilátero Ferrífero e Triângulo mineiro. Por outro lado, B e Mo não apresentaram um padrão definido de distribuição espacial dos teores. O Hg apresentou os maiores teores na região sudeste do Estado. A região norte do estado apresentou os mais baixos teores para a maioria dos elementos (Al, As, Co, Cr, Cu, Fe, Hg, Se, Sr e V).









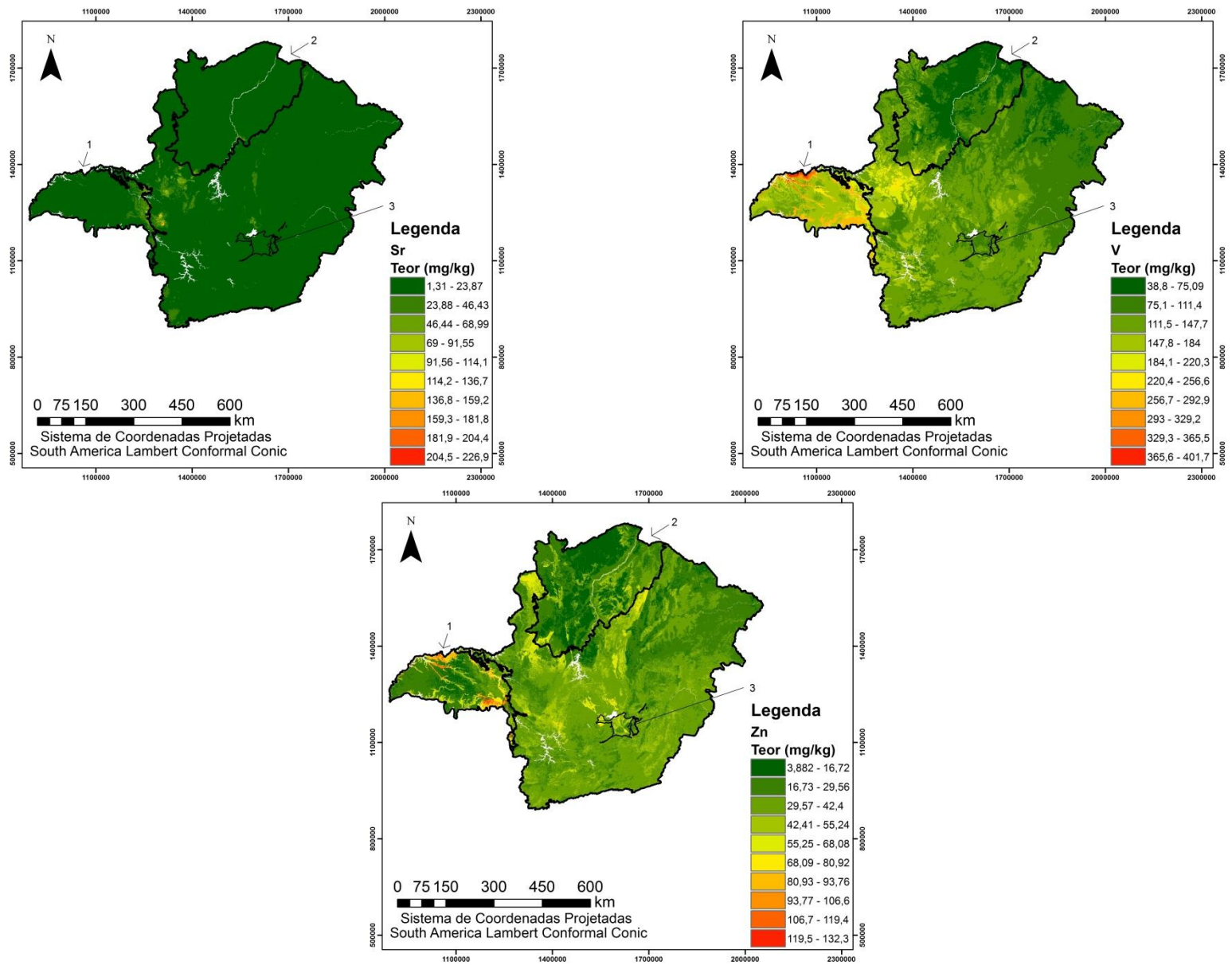


Figura 5 – Mapas predição dos melhores resultados de nRMSE para os elementos analisados Legenda: 1 - Triângulo Mineiro, 2 - Norte e 3 - Quadrilátero Ferrífero

DISCUSSÃO

A importância da covariável litologia na predição de grande parte dos elementos analisados (Al, As, B, Ba, Cd, Co, Cr, Cu, Fe, Hg, Mn, Mo, Ni, Sr, V, Zn) pode ser atribuído a influência do material de origem sobre composição química dos solos (FASSBENDER, 1980; KABATA-PENDIAS, 2011; DAVIES e WIXSON, 1987; GUILHERME et al., 2005; ALLEONI et al., 2005). Rochas ígneas básicas e ácidas e xistos apresentam teores elevados de As, Ba, Cd, Cr, Cu, Mn, Mo, Ni, Pb, V e Zn em comparação à outras rochas (ALLOWAY, 1995). O teor de metais e metaloides em rochas sedimentares é variável, mas em geral é mais baixo do que outras rochas.

Em geral, as rochas metamórficas apresentam grande variação na concentração de elementos traço, que são dependentes da concentração do protólito metamorfizado. As rochas metamórficas também podem apresentar processos de enriquecimento geogênico, que acarretam na concentração de alguns metais e metaloides, tais como As, Cd, Cr e Ni. Assim, derivados de diferentes litologias podem herdar diferentes teores de metais e metaloides.

Os elevados teores de metais e metaloides em solos do Quadrilátero Ferrífero podem ser associados à dominância de rochas metamórficas que sofreram diversos processos de mineralização ao longo dos últimos 1,5 bilhões de anos (ALKMIM et al, 2015).

Os altos teores dos elementos no material de origem podem não refletir em altos teores dos elementos nos solos, devido aos processos de intemperismo, pois este processo tende a lixiviar parte dos elementos. Sendo assim, são necessários elementos que adsorvem os íons e, conseqüentemente reduzam sua mobilidade nos solos. Elementos tri e pentavalentes em sua maioria apresentam elevada afinidade por argilas (hidr)óxidos e/ou matéria orgânica do solos (KABATA-PENDIAS, 2011). Além disso, processo como laterização e dessilicação tendem a aumentar a concentração relativa de Al e Fe sob a forma de (hidr)óxidos e, conseqüentemente, reter outros metais e metaloides no solo contra os processos de intemperismo e lixiviação.

Os elementos Cd, Hg, Mo, Se, Sb apresentam baixos teores médios nos solos mundialmente, com valores inferiores a 1 mg L^{-1} (KABATA-PENDIAS, 2011).

A covariável classes de solos tem importância na predição dos teores dos elementos analisados, pois reflete a influência dos principais processos pedogenéticos

na composição química do solo. Os minerais secundários formados durante a dessilicação e ferralitização são importantes para imobilização de elementos menores (Cd, Co, Cr, Cu etc.) no solo através da co-precipitação (ALLOWAY, 1995). Os processos de lessivagem, elutriação e a podzolização promovem o empobrecimento relativo de Al e Fe nos horizontes superiores do solo pela diminuição do teor de argila, e diferenciam quimicamente Argissolos, Espodosolos, Luvisolos e Planossolos dos demais. Por outro lado, a gleização conduz a redução e lixiviação de Fe e, conseqüentemente, de outros metais, formando Gleissolos. A plintização promove a mobilização e concentração de Fe em nódulos e outros elementos por 'efeito de arraste' em Plintossolos (KABATA-PENDIAS, 2011; DE SOUZA et al., 2015).

As covariáveis geomorfologia e MDE são representações do relevo, sendo o MDE usado para descrever a geomorfologia (TRAN et al., 2014; BADURA et al., 2004). O relevo é um dos principais fatores de formação de solo, que culminam na acumulação ou lixiviação de elementos dos solos (LESPSCH, 2016). A declividade da superfície bem como sua posição topográfica afetam os processos de infiltração de água, erosão física, a mobilidade e a concentração de elementos nos solos. Além disso, as características físico-químicas da água são outro fator a ser considerado, principalmente, os fatores pH e Eh (BERTOLO et al., 2009). No entanto, regiões de planaltos na maioria das vezes, apresentam condições favoráveis à maior infiltração da água no solo e por conseqüência a lixiviação dos elementos. Enquanto que, em regiões de relevo ondulado ou ainda mais declivoso os processos erosivos reduzem a profundidade dos solos, culminando em solos incipientes e com teor de metais e metaloides de maior similaridade com o conteúdo da rocha matriz.

A associação das covariáveis relacionadas à gamespectrometria (eTh, K, eU e suas razões) e a distribuição dos metais nos solos podem ser atribuídas ao nível de intemperismo dos solos e também pela diferença do material de origem dos solos. O K é encontrado em altas concentrações em rochas félsicas, que são abundantes em feldspatos, biotitas, muscovitas, e alguns anfibólios, tais como granitos, sienitos e riolitos em comparação às rochas máficas (FAURE, 1991). Sob condições úmidas e bem drenadas, o K^+ é lixiviado durante o processo de intemperismo. Metais alcalinos e alcalino terrosos de raio iônico hidratado ligeiramente similar ao K, tais como Ba, Co e Sr, possuem comportamento semelhante durante o intemperismo e são correlacionados. Minerais contendo Th são mais abundantes em rochas ígneas félsicas e gnaisses. Uma vez liberado destes minerais primários, o Th é adsorvido por óxidos de Fe ou Al, argilas silicatadas ou complexado por compostos orgânicos. Além disso, diversos minerais contendo Th são resistentes ao intemperismo, tais como o

zircão, a ilmenita etc. Com isso, o Th é tipicamente concentrado em solos durante o processo de intemperismo em comparação ao K e outros elementos mais solúveis. Elementos trivalentes também acumulados em solos altamente intemperizados e/ou, com horizontes cimentados por sílica, óxidos de Fe ou Al podem ser preditos pela contagem de Th (Figura 4). O incremento do teor de sílica das rochas é acompanhado pelo aumento do teor de U na rocha, sendo, portanto mais comum em rochas félsicas (MARKOVAARA-KOIVISTO et al., 2008).

A mobilidade do U é complexa e altamente dependente do pH e Eh. Uma vez liberado pelo intemperismo dos minerais, o U apresenta características semelhantes ao Th quando relacionado processo de adsorção com argilominerais, óxidos de Al ou Fe, e complexação com matéria orgânica do solo (CLAVERANNE-LAMOLÈREET et al, 2008). Mas ao contrário do Th e da maioria dos elementos analisados, o U é lixiviado em condições oxidantes com boa drenagem e precipitado em condições redutoras. Sua ocorrência é fortemente associada a anomalias supergênicas, o que pode enfraquecer seu poder preditivo sobre os elementos analisados (Figura 4). A ausência de um padrão similar entre litologia e gamaespectrometria na predição dos elementos analisados pode indicar a atenuação da variabilidade do material parental dos solos em detrimento da atuação do intemperismo e dos processos pedogenéticos.

Com relação à temperatura média anual e precipitação, nota-se tanto o valor acumulado quanto à amplitude e os valores dos períodos secos e chuvosos, mostraram-se de grande relevância para a predição dos mapas para os elementos avaliados (Figura 4), estando associada com a influência sobre os processos de lixiviação e acumulação de elemento no solo e, indiretamente, por meio da erosão do solo. A baixa importância da temperatura média no estado na predição de metais e metaloides pode ser associada à baixa variação térmica na área de estudo quando comparadas com regiões com latitudes maiores onde ocorrem temperaturas negativas. A precipitação influencia o intemperismo e lixiviação dos elementos nos solos, assim, solos originados da mesma rocha em regiões com maiores precipitações terão teores inferiores de metais e metaloides (KÄMPF e CURI; 2012).

A covariável NDVI mostrou-se relevante para demonstrar o efeito biológico no aporte e saída de elementos como Cr, Hg, Mn, Pb, Sb e Zn nos solos (Tabela 4). A vegetação altera a distribuição vertical e disponibilidade de elementos químicos de acordo com a magnitude dos processos de ciclagem e lixiviação (JOBAGY & JACKSON, 2004). Os padrões verticais são condicionados pelas plantas e consistentes com o nível de importância do elemento para a planta quando comparados com elementos de menor importância ou tóxicos. Para o elemento Hg, os

maiores teores estão localizados na parte leste, centro e sul do estado, que coincide com a região de predomínio da mata atlântica. Por outro lado, não foi possível ver um padrão de localização para os elementos Cr, Mn, Pb, Sb e Zn.

O padrão similar entre as covariáveis relacionadas à gravimetria (primeira derivada vertical, total e Sinal Analítico) e litologia sugere-se que o contraste de densidade entre diferentes materiais rochosos é sensível e permite a predição de alguns dos elementos analisados (CASTRO e COSTA, 1999).

Os elementos Ba, Cr e Sr apresentaram um número de covariáveis selecionadas superior a dez. O elevado número de preditores pode estar relacionado a um baixo desempenho de predição dos algoritmos quando não existe uma alta correlação com a variável predita. Dessa forma, este fato mostra-se que não existe uma covariável de elevada importância para predição desses elementos, e os algoritmos tentam utilizar o maior número de covariáveis tentando buscar relações entre eles para tentar aumentar o desempenho de treinamento. Um grande número de preditores acarreta em uma maior complexidade para explicação das correlações entre os elementos e as covariáveis utilizadas para a predição.

O baixo poder explicativo dos modelos preditivos para o Mn pode ser atribuído a sua elevada mobilidade, uma vez que este elemento pode ser removido sob condições redutoras, por soluções ácidas, bicarbonatos e forma complexa com a matéria orgânica (KABATA-PENDIAS, 2011).

Os elevados teores de As, Cd, Cr na região do Quadrilátero Ferrífero (Figura 5) podem ser associados às mineralizações frequentes na área (SPIER et al., 2007; DARDENNE e SCHOBENHAUS, 2001). As regiões próximas à zona de mineração tendem a apresentar altos teores dos elementos minerados e outros associados (KABATA-PENDIAS, 2011; ALLOWAY, 1999).

Os maiores teores de Co, Cu, Fe, V e Zn Triângulo Mineiro quando comparados ao resto do estado (Figura 5), reafirma o encontrado por De Souza et al. (2015), que observou elevados teores destes elementos nesta região e, além disso, Oliveira e Costa (2004) encontraram maior disponibilidade de Cu e Zn em solos de originados do basalto desta região. Rochas basálticas e gabroicas apresentam maiores teores de Cu, Mn, V e Zn quando comparados com outros tipos de rochas (ALLOWAY, 1999), o que explicaria os altos teores desses elementos nesta região, quando comparado ao restante do Estado. Os teores de Fe nesta região estão relacionados ao desenvolvimento de solos a partir de rochas máficas (basaltos) (KABATA-PENDIAS, 2011)

Os baixos teores dos elementos avaliados na região norte do Estado podem ser atribuídos à influência da litologia local (Figura 1). Assim, os estes solos são derivados de depósitos sedimentares, arenitos e outras rochas pelíticas (CODEMIG, 2014), que contêm em sua composição baixos teores dos elementos avaliados (PERES et al., 2004, DE SOUZA et al., 2015).

Ajuste dos modelos preditivos

Para os elementos que tiveram como melhores resultados sem a transformação dos dados (As, B, Ba, Cd, Cr, Cu, Fe, Hg, Mn, Mo, Ni, Pb, Sb, Se, V) os resultados R^2 ficaram abaixo de 50%. Este resultado pode ser explicado pela grande diversidade geológica-pedológica. A transformação dos dados não gerou melhora no desempenho da maioria dos elementos. Porém, três elementos apresentaram aumento de desempenho nos resultados com a transformação dos dados, sendo estes o Al, Sr e Zn. Os valores de R^2 para os elementos não podem ser comparados entre dados não transformados e transformados, visto que a transformação dos dados leva a uma redução do desvio padrão e amplitude dos dados previstos e observados que geram os R^2 superiores e, assim, não demonstrando a realidade dos processos de treinamento. Para o RMSE houve uma tendência de redução dos valores para transformações utilizadas neste trabalho pelos mesmos motivos que leva a melhoria dos valores de R^2 .

Para Hg e Pb, os baixos valores de R^2 podem estar relacionados aos fluxos de entrada e saída por deposição e volatilização atmosférica. O Hg é um elemento facilmente volatilizado em sua forma elementar e também forma compostos orgânicos gasosos do tipo metal-Hg, com os íons hidroxila e/ou cloro forma compostos facilmente solubilizáveis nos solos (KABATA-PENDIAS, 2011, MURKHERJEE, 2007). Dessa forma, os teores de Hg tendem a aumentar próximos de regiões de mineração e fundição (KABATA-PENDIAS, 2011). Enquanto o aporte de chumbo na camada superior dos solos pode ocorrer por tetraetila, utilizada até 1989 como aditivo na gasolina de veículos automotores no Brasil (LANDRIGAN, 2002).

Quatro algoritmos de predição foram selecionados, a partir dos seis avaliados, devido aos maiores valores de nRMSE para os elementos avaliados. Assim, o uso de apenas um algoritmo levaria a um desempenho inferior para a maioria dos elementos avaliados. Outro ponto relevante foi o uso de algoritmos de aprendizado de máquina e a definição da importância de covariáveis na predição dos elementos, esses métodos não são possíveis de serem executados na krigagem.

CONCLUSÃO

As covariáveis selecionadas para análise dos elementos apresentaram alta interação com a litologia, relevo e clima. Os elementos Al, As, Ba, Co, Mo, Pb, Sb, Sr, Zn apresentam correlação com covariável gamaespectrometria. A vegetação apresentou correlacionada com os teores de Cr, Hg, Mn, Mo, Pb, Sb, Zn, enquanto que, para B, Mo e Sr, as covariáveis relacionadas para a predição foram a magnetometria e gravimetria. Assim, os mapas gerados contribuem para a regionalização dos teores dos VRQ no Estado de Minas Gerais.

Os algoritmos ExtraTrees, RandomFlorest, Ranger e Suport Vector Machine Radial Sigma proporcionaram melhores resultados para os elementos obtidos, a partir da avaliação dos algoritmos.

Este trabalho situa-se entre os primeiros que utilizam algoritmos de aprendizado de máquinas na América Latina para a predição de teores de metais e metalóides. Dessa forma, resultados mais precisos podem ser alcançados com utilização de novos preditores, não utilizadas neste trabalho.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABRAHÃO, W. A. F.; MARQUES, J. J. **Manual de procedimentos analíticos para determinação de VRQ de elementos-traço em solos do Estado de Minas Gerais**. Belo Horizonte, 2013. 34.p.
- ALKMIM, A. R. et al. Geochemical fingerprint of siliceous, amphibolitic and magnetitic itabirite types of the region of Serra Azul–Quadrilátero Ferrífero, MG. **Rem: Revista Escola de Minas**, v. 68, n. 1, p. 37-42, 2015.
- AHMAD, S.; KALRA, A.; STEPHEN, Haroon. Estimating soil moisture using remote sensing data: A machine learning approach. **Advances in Water Resources**, v. 33, n. 1, p. 69-80, 2010.
- AHMED, O. S. et al. Characterizing stand-level forest canopy cover and height using landsat time series, samples of airborne LiDAR, and the random forest algorithm. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 101, p. 89-101, 2015.
- ALLEONI, L. R. F.; BORBA, R. P.; CAMARGO, OA de. Metais pesados: da cosmogênese aos solos brasileiros. **Tópicos em ciência do solo**, v. 4, p. 1-42, 2005.
- ALLOWAY, B. J. Heavy metals in soils Blackie Academic and Professional. **London, UK**, 1995.
- ALVAREZ, R. Predicting average regional yield and production of wheat in the Argentine Pampas by an artificial neural network approach. **European Journal of Agronomy**, v. 30, n. 2, p. 70-77, 2009.
- ANALYTICS, R.; WESTON, S.. doParallel: Foreach parallel adaptor for the parallel package. **R package version**, v. 1, n. 8, 2014.
- AYOTTE, J. D.; NOLAN, B. T.; GRONBERG, J. A.. Predicting Arsenic in Drinking Water Wells of the Central Valley, California. **Environmental Science & Technology**, v. 50, n. 14, p. 7555-7563, 2016.
- BADURA, J.; PRZYBYLSKI, B. . Application of digital elevation models to geological and geomorphological studies—some examples. **Przeegląd Geologiczny**, v. 53, n. 10/2, p. 977-983, 2005.
- BERTOLO, R. A.; MARCOLAN, L. N. O.; BOUROTTE, C. L. M.. Relações água-rocha e a hidrogeoquímica do cromo na água subterrânea de poços de monitoramento multiníveis de Urânia, SP, Brasil. **Geologia USP. Série Científica**, v. 9, n. 2, p. 47-62, 2009.
- BIONDI, C. M. et al. Teores de Fe, Mn, Zn, Cu, Ni e Co em solos de referência de Pernambuco. **Revista Brasileira de Ciência do Solo**, v. 35, n. 3, 2011.
- Bivand, R, et. al. **Package “rgdal.”** R Package, 2016.
- BRUNGARD, C. W. et al. Machine learning for predicting soil classes in three semi-arid landscapes. **Geoderma**, v. 239, p. 68-83, 2015.
- CASTRO, D. L.; COSTA, F. C. N.. Mapas gravimétricos do Estado do Ceará. **Revista de Geologia**, v. 12, n. 1, 1999.
- CARROLL, M. L. et al. 250m MODIS normalized difference vegetation index. **University of Maryland, College Park, Maryland**, 2004.
- CEMPEL, M.; NIKEL, G. Nickel: a review of its sources and environmental toxicology. **Polish Journal of Environmental Studies**, v. 15, n. 3, p. 375-382, 2006

CHIAP, P. et al. Validation of an automated method for the liquid chromatographic determination of atenolol in plasma: application of a new validation protocol. **Analytical chimica acta**, v. 391, n. 2, p. 227-238, 1999.

CLAVERANNE-LAMOLÉRE, Céline et al. Characterization of the interactions between uranium and colloids in soil by on-line fractionation multi-detection methods. **Uranium, Mining and Hydrogeology**, p. 807-814, 2008.

CONSELHO NACIONAL DO MEIO AMBIENTE (CONAMA). Resolução nº 420, de 28 de dezembro de 2009, 2009 [Brasil].

CODEMIG. Mapa Geológico do Estado de Minas Gerais; 2014. p. 1 [Belo Horizonte].

COOK, D. J.; KRISHNAN, N. C. **Activity learning: discovering, recognizing, and predicting human behavior from sensor data**. John Wiley & Sons, 2015, 282.p.

DARDENNE, M. A.; SCHOBENHAUS, C.. **Metalogênese do Brasil**. Editora UnB, 2001.

DAVIES, B. E.; WIXSON, B. G. Use of factor analysis to differentiate pollutants from other trace metals in surface soils of the mineralized area of Madison County, Missouri, USA. **Water, Air, & Soil Pollution**, v. 33, n. 3, p. 339-348, 1987.

DE SOUZA, J. J. L. et al. Geochemistry and spatial variability of metal (loid) concentrations in soils of the state of Minas Gerais, Brazil. **Science of the Total Environment**, v. 505, p. 338-349, 2015.

FADIGAS, F. de S. et al. Concentrações naturais de metais pesados em algumas classes de solos brasileiros. 2002.

FASSBENDER, H. W.; BORNEMISZA, E.. **Química de suelos con énfasis en suelos de América Latina**. lica, 1987.

FARR, T G. et al. The shuttle radar topography mission. **Reviews of geophysics**, v. 45, n. 2, 2007.

FAURE, Gunter. **Principles and applications of inorganic geochemistry**. McMillan Pub. Company, 1991.

FIRKE, S.; HAID, C. and KNIGHT, R, Package 'janitor'. 2017

FORKUOR, G. et al. High Resolution Mapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso: A Comparison of Machine Learning and Multiple Linear Regression Models. **PLoS one**, v. 12, n. 1, p. e0170478, 2017.

FU, Y. et al. A tree-based approach to preserve the privacy of software engineering data and predictive models. In: **Proceedings of the 5th International Conference on Predictor Models in Software Engineering**. ACM, 2009. p. 3.

FUNDAÇÃO JOÃO PINHEIRO. *Perfil demográfico do Estado de Minas Gerais - 2002*. Belo Horizonte: Centro de Estatística e Informações, 2003. 111 p.

GÖHRE, Vera; PASZKOWSKI, Uta. Contribution of the arbuscular mycorrhizal symbiosis to heavy metal phytoremediation. **Planta**, v. 223, n. 6, p. 1115-1122, 2006.

GRAY, J. M.; BISHOP, T. F. A; WILFORD, J. R. Lithology and soil relationships for soil modelling and mapping. **Catena**, v. 147, p. 429-440, 2016.

GUILHERME, L. R. G. et al. Elementos-traço em solos e sistemas aquáticos. **Tópicos em ciências do solo**, v. 4, p. 345-390, 2005.

- GUO, P et al. Digital mapping of soil organic matter for rubber plantation at regional scale: An application of random forest plus residuals kriging approach. **Geoderma**, v. 237, p. 49-59, 2015.
- HALLAK, R.; PEREIRA FILHO, A. J. Metodologia para análise de desempenho de simulações de sistemas convectivos na região metropolitana de São Paulo com o modelo ARPS: sensibilidade a variações com os esquemas de advecção e assimilação de dados. **Revista Brasileira de Meteorologia**, v. 26, n. 4, p. 591-608, 2011.
- HAMNER, B. *Package 'Metrics'*. 2017
- HAN, K.; VIAU, A. A.; ANCTIL, F.. High-resolution forest fire weather index computations using satellite remote sensing. **Canadian journal of forest research**, v. 33, n. 6, p. 1134-1143, 2003.
- HANSSON, K. et al. Machine learning algorithms in heavy process manufacturing. **American Journal of Intelligent Systems**, v. 6, n. 1, p. 1-13, 2016
- HAO, L. et al. Spatial distribution and sources of trace elements in surface soils, Changchun, China: Insights from stochastic models and geostatistical analyses. **Geoderma**, v. 273, p. 54-63, 2016.
- HENDERSON, B. L. et al. Australia-wide predictions of soil properties using decision trees. **Geoderma**, v. 124, n. 3, p. 383-398, 2005.
- HENGL, T.; NIKOLIĆ, M.; MACMILLAN, R. A. Mapping efficiency and information content. **International Journal of Applied Earth Observation and Geoinformation**, v. 22, p. 127-138, 2013.
- HENGL, T. et al. SoilGrids250m: Global gridded soil information based on machine learning. **PloS one**, v. 12, n. 2, p. e0169748, 2017.
- HILL, R. A.; HAWKINS, C. P.; J., J.. Predicting thermal vulnerability of stream and river ecosystems to climate change. **Climatic change**, v. 125, n. 3-4, p. 399-412, 2014.
- HIJMANS, R.J. et al, *Package 'raster'*. 2016.
- HIJMANS, R. J. et al. Very high resolution interpolated climate surfaces for global land areas. **International journal of climatology**, v. 25, n. 15, p. 1965-1978, 2005.
- HORNIK, K.; BUCHTA, C.; ZEILEIS, A.. Open-source machine learning: R meets Weka. **Computational Statistics**, v. 24, n. 2, p. 225-232, 2009.
- JAMIESON, P. D.; PORTER, J. R.; WILSON, D. R. A test of the computer simulation model ARCWHEAT1 on wheat crops grown in New Zealand. **Field crops research**, v. 27, n. 4, p. 337-350, 1991.
- JOBBÁGY, E. G.; JACKSON, R. B. The uplift of soil nutrients by plants: biogeochemical consequences across scales. **Ecology**, v. 85, n. 9, p. 2380-2389, 2004.
- JENNY, H. **Factors of soil formation: a system of quantitative pedology**. New York: McGraw-Hill, 1941. 281 p.
- KABATA-PENDIAS, A. et al. **Trace elements in soils and plants**. Boca Raton: CRC press, 2010.
- KARATZOGLOU, A.; MEYER, D.; HORNIK, K. Support vector machines in R. 2005.
- KIRKWOOD, C. et al. A machine learning approach to geochemical mapping. **Journal of Geochemical Exploration**, v. 167, p. 49-61, 2016.

- KOVAČEVIĆ, M.; BAJAT, B.; GAJIĆ, B. Soil type classification and estimation of soil properties using support vector machines. **Geoderma**, v. 154, n. 3, p. 340-347, 2010.
- KÄMPF, N.; CURTI, N. Formação e evolução do solo (Pedogênese). **Pedologia: Fundamentos. Viçosa, MG: Sociedade Brasileira de Ciência do Solo**, p. 207-302, 2012.
- KUHN, M. Caret package. **Journal of Statistical Software**, v. 28, n. 5, p. 1-26, 2008.
- KUHN, M. et al. Cubist package. **Version 0.0. 19**, 2013.
- LANDRIGAN, Philip J. The worldwide problem of lead in petrol. **Bulletin of the World Health Organization**, v. 80, n. 10, p. 768-768, 2002.
- LEPSCH, I. F. **Formação e conservação dos solos**. Oficina de textos, 2016.
- LIAW, A.; WIENER, M.. The randomforest package. **R News**, v. 2, n. 3, p. 18-22, 2002.
- LOAGUE, K.; GREEN, R. E. Statistical and graphical methods for evaluating solute transport models: overview and application. **Journal of contaminant hydrology**, v. 7, n. 1-2, p. 51-73, 1991.
- LOPATIN, J. et al. Comparing generalized linear models and random forest to model vascular plant species richness using LiDAR data in a natural forest in central Chile. **Remote sensing of environment**, v. 173, p. 200-210, 2016.
- LOPEZ-PAZ, David; HENNIG, Philipp; SCHÖLKOPF, Bernhard. The randomized dependence coefficient. In: **Advances in neural information processing systems**. 2013. p. 1-9.
- MACHADO, M. F.; SILVA, S. F. da. Geodiversidade do estado de Minas Gerais. 2010.
- MALAFAIA, P. et al. Equívocos arraigados no meio pecuário sobre deficiências e suplementação minerais em bovinos no Brasil. **Pesq. Vet. Bras**, v. 34, n. 3, p. 244-249, 2014.
- MARKOVAARA-KOIVISTO, Mira et al. Uranium mineralogy at the Askola ore deposit, Southern Finland. In: **Materials Research Society Symposium Series. Scientific Basis for Nuclear Waste Management**. 2008. p. 1124.
- MCBRIDE, M. B. **Environmental chemistry of soils**. New York: Oxford University, 1994. 406 p. YU. TR.
- HIJMANS, L.; MALTAMO, M.; KANGAS, A.. The use of quantile trees in the prediction of the diameter distribution of a stand. **Silva Fennica**, v. 40, n. 3, p. 501, 2007.
- MEYER, D. et al. Package 'e1071'. 2017.
- MUKHERJEE, Santanu et al. Asymmetric enamine catalysis. **Chem. Rev**, v. 107, n. 12, p. 5471-5569, 2007.
- PERES, G. G.; ALKIMIM, F. F.; JORDT-EVANGELISTA, H. The southern Araçuaí belt and the Dom Silvério Group: geologic architecture and tectonic significance. **Anais da Academia Brasileira de Ciências**, v. 76, n. 4, p. 771-790, 2004.
- PELOZATO, M. et al. Comparação entre métodos de extração de cádmio, cobre e zinco de solos catarinenses derivados de basalto e granito-migmatito. **Revista de Ciências Agroveterinárias**, v. 10, n. 1, p. 54-61, 2011.
- PENG, Y. et al. Digital Mapping of Toxic Metals in Qatari Soils Using Remote Sensing and Ancillary Data. **Remote Sensing**, v. 8, n. 12, p. 1003, 2016.

OLIVEIRA, T. S.; COSTA, L. M. Metais pesados em solos de uma topolitosseqüência do Triângulo Mineiro. **Revista Brasileira de Ciências do Solo**, v. 28, p. 785-796, 2004.

OLIVEIRA, F. J. S.; JUCÁ, J. F. T.. Acúmulo de metais pesados e capacidade de impermeabilização do solo imediatamente abaixo de uma célula de um aterro de resíduos sólidos. **Eng. sanit. ambient**, v. 9, n. 3, p. 211-217, 2004.

REIMANN, C.; FILZMOSER, P. Normal and lognormal data distribution in geochemistry: death of a myth. Consequences for the statistical treatment of geochemical and environmental data. **Environmental geology**, v. 39, n. 9, p. 1001-1014, 2000.

SILVA, T. I.; RODRIGUES, S. C. ELABORAÇÃO DE UM TUTORIAL DE CARTOGRAFIA GEOMORFOLÓGICA COMO ALTERNATIVA PARA O ENSINO DE GEOMORFOLOGIA. **Revista Geográfica Acadêmica**, v. 3, n. 2, 2009.

SIMM; J. and ABRIL; I. M. Package 'extraTrees'. 2015

SPIER, C. A. et al. Geochemistry and genesis of the banded iron formations of the Cauê Formation, Quadrilátero Ferrífero, Minas Gerais, Brazil. **Precambrian Research**, v. 152, n. 3, p. 170-206, 2007.

SCHWAIGHOFER, A. et al. How wrong can we get? A review of machine learning approaches and error bars. **Combinatorial chemistry & high throughput screening**, v. 12, n. 5, p. 453-468, 2009.

TRAN, T. A. et al. A geomorphology-based approach for digital elevation model fusion-Case study in Danang City, Vietnam. **Earth Surface Dynamics**, v. 2, n. 2, p. 403, 2014.

VAYSSE, K.; LAGACHERIE, P. Evaluating digital soil mapping approaches for mapping GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France). **Geoderma Regional**, v. 4, p. 20-30, 2015.

UNIVERSIDADE FEDERAL DE VIÇOSA; FUNDAÇÃO CENTRO TECNOLÓGICO DE MINAS GERAIS; UNIVERSIDADE FEDERAL DE LAVRAS; FUNDAÇÃO ESTADUAL DO MEIO AMBIENTE. **Mapa de solos do Estado de Minas Gerais** [Belo Horizonte], 2010.

YANG, R. et al. 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.

WARNES, G. R.; BOLKER, B. and LUMLEY, T. Package 'gtools'. 2015

WEI, T.; WEI, M. T.. Package 'corrplot'. **Statistician**, v. 56, p. 316-324, 2016.

WICKHAM, H.. ggplot2: elegant graphics for data analysis. **J Stat Softw**, v. 35, n. 1, p. 65-88, 2010.

WILLIAMS, Chris Keefer et al. Package 'caret'. 2016.

WILFORD, J.; DE CARITAT, P.; BUI, E.. Predictive geochemical mapping using environmental correlation. **Applied Geochemistry**, v. 66, p. 275-288, 2016.

WRIGHT, M. N.; ZIEGLER, A.. ranger: A fast implementation of random forests for high dimensional data in C++ and R. **arXiv preprint arXiv:1508.04409**, 2015.

ZAMBRANO-BIGIARINI, M. Package "hydroGOF". 2017

CAPÍTULO 3

COMPARAÇÃO DE SOFTWARES PARA ESPACIALIZAÇÃO DE METAIS E METALÓIDES NO ESTADO DE MINAS GERAIS PELA KRIGAGEM ORDINÁRIA

INTRODUÇÃO

A composição química do solo é influenciada diretamente pelo material de origem. Os solos tropicais ácidos e de forte drenagem são um reflexo da dissolução de minerais primários que conduziu à lixiviação de metais alcalino e alcalino terrosos e o aumento da concentração relativa de elementos trivalentes na forma de óxidos. O teor de metais de transição no solo é diretamente associado ao teor desses óxidos, uma vez que são retidos por estes, via processo de adsorção e/ou coprecipitação.

Alguns elementos químicos definidos como traços apresentam elevada toxicidade para a flora e fauna, sendo estes: antimônio, arsênio, cádmio, chumbo, cromo, cobalto, mercúrio, níquel e zinco (CEMPEL e NIKEL, 2006; GÖHRE e PASZKOWSKI, 2006). Por essa razão, existe uma grande preocupação ambiental quanto à quantificação dos teores destes elementos traço em solos e resíduos. No Brasil, o Conselho Nacional de Meio Ambiente (CONAMA), por meio da Resolução nº 420 de 2009, dispôs sobre os critérios e valores orientadores da qualidade (VRQ) dos solos quanto à presença de substâncias químicas potencialmente tóxicas (CONAMA, 2009). Nesse sentido, os Estados brasileiros ficaram responsáveis pela determinação dos VRQs para seus solos. O Estado de Minas Gerais iniciou seus esforços em 2009 para elaboração dos seus VRQs, os quais foram estabelecidos pelo Conselho Estadual de Política Ambiental (COPAM), através da resolução nº 166 de 29 de julho em 2011 (COPAM, 2011).

A resolução COPAM 166/2011 foi de grande importância para o avanço do conhecimento e proteção do solo e do meio ambiente. A determinação dos VRQ foi realizada por meio de 499 amostragens de solos superficiais em todo o Estado de Minas Gerais, armazenadas no Banco de Solos do Estado de Minas Gerais (BSMG).

Atualmente o BSMG conta com 648 amostras georreferenciadas e busca validar uma superfície de predição para teor de metais e metaloides no Estado, avaliando a necessidade de novas amostragens e o estudo da variabilidade espacial dos teores de elementos químicos. A interpolação de dados geoquímicos para a área

do Estado é importante para prever o teor de metais traços em áreas não amostradas. Tal previsão por si expressa influência dos fatores pedogenéticos sobre o material parental. Além disso, ela permite uma análise da variação espacial dos teores e, possivelmente, gerar os VRQs regionais.

O método de krigagem é um dos principais métodos utilizados para a interpolação dos dados pontuais no solo (CAMARGO et al., 2008; DA SILVA JÚNIOR et al., 2012a,b; OLIVEIRA JÚNIOR et al., 2011). Este método é aplicado em diversas áreas da ciência, tais como: climatologia (CASTRO et al., 2010; DE MELLO et al., 2003), solos (DE SOUZA et al., 2010; LIMA et al., 2014) e estudos florestais (PAIVA et al., 2010). Além disso, a krigagem foi aplicada na interpolação de teores de elementos químicos inorgânicos para solos, em grande parte do continente europeu, com a obtenção de resultados satisfatórios, de acordo com os autores (DE VOS et al. 2006; LADO et al. 2008). Trata-se de um processo geoestatístico de estimativa de valores aplicado a uma variável distribuída no espaço e, ou, tempo. O método se baseia em valores adjacentes e conhecidos, quando estes são de caráter interdependente pelas análises variográficas (YAMAMOTO e LANDIM, 2013).

De Souza et al, (2015) utilizaram a krigagem para prever os teores de elementos traço e os atributos dos solos na área do Estado de Minas Gerais, a partir das amostras contidas no BSMG. Pela natureza dinâmica do BSMG, novas amostras foram inseridas ao seu acervo e, a partir de um maior controle de qualidade das análises, é possível aumentar o nível de detalhamento e precisão da diversidade pedológica e de background de elementos traço no Estado. Neste sentido, o presente estudo comparou o software Arcgis, a ferramenta Toolbars Geostatistical Analyst do software Arcgis e o R para previsão do teor de elementos químicos (Al, As, B, Ba, Cd, Co, Cr, Cu, Fe, Hg, Mn, Mo, Ni, Pb, Sb, Se, Sr, V e Zn) no Estado de Minas Gerais pela krigagem ordinária e posterior espacialização destes teores.

MATERIAL E MÉTODOS

Banco de dados

Foram utilizadas 648 amostras georreferenciadas obtidas do banco de dados de solos do Estado de Minas Gerais (Figura 1). Estas amostras foram coletadas a profundidade de 0-20 cm. O processo de amostragem e quantificação dos elementos químicos baseou-se no protocolo estabelecido por De Souza et al (2015) e Abrahão e

Marques (2013). Na Tabela 1, encontra-se a estatística descritiva dos elementos avaliados:

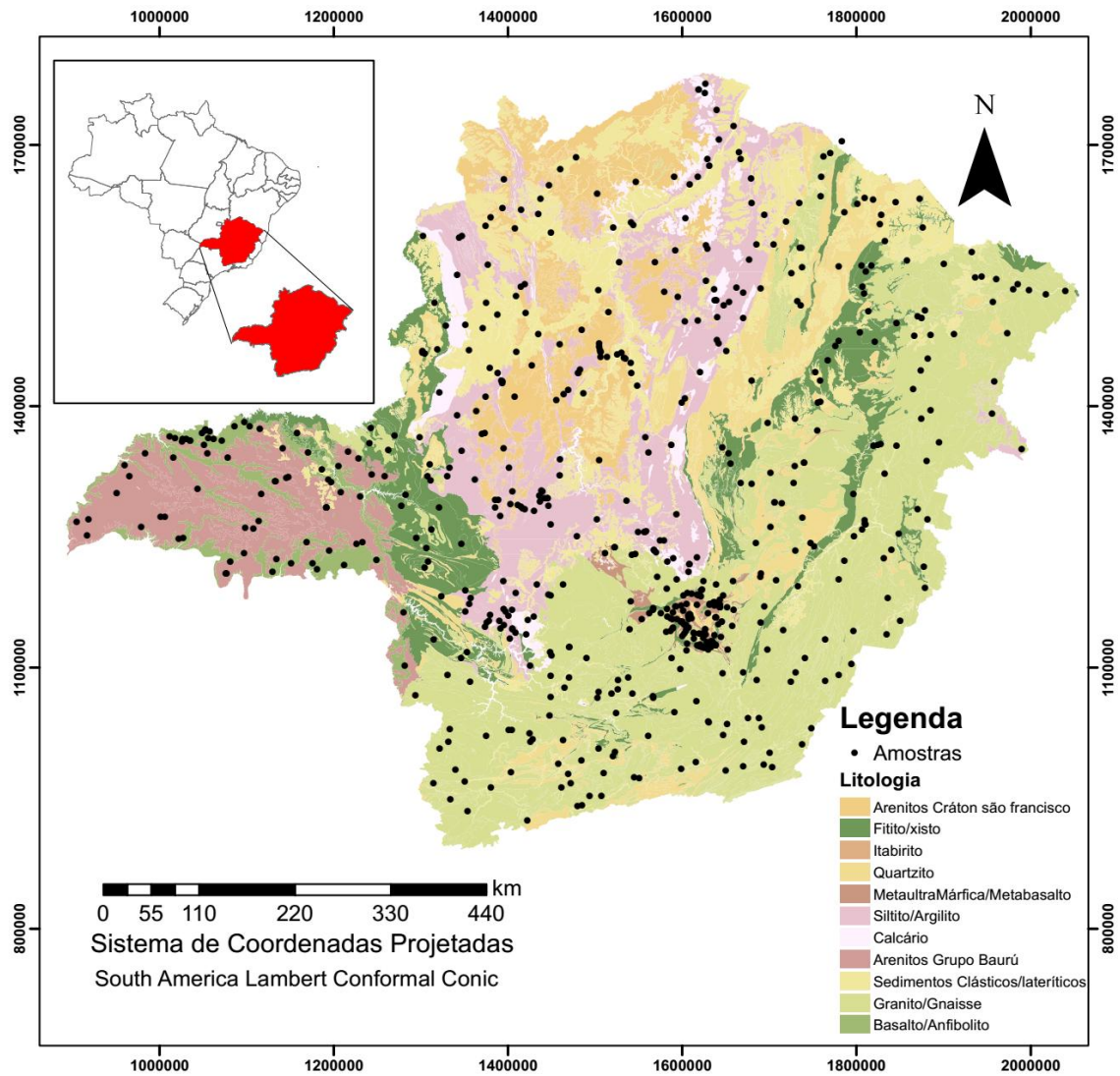


Figura 1. Distribuição de pontos coletados e mapa litológico simplificado da área estudo.

Tabela 1 – Estatísticas descritivas das amostras de solos na cama de 0-20 cm, para os teores de metais e metaloides;

	Número de amostras	Média	Desvio Padrão	CV**	Mediana	Máximo	Mínimo
Elemento	*	mg/kg	mg/kg	%	mg/kg	mg/kg	mg/kg
Al	647	5,55x10 ⁴	4.11x10 ⁴	74,02	4.70x10 ⁴	2,51x10 ⁵	850,99
As	576	16,17	58,13	359,49	4,26	854,51	0,13
B	647	7,71	6,74	87,42	6,67	43,64	0,28
Ba	647	99,82	272,09	272,58	46,62	4472,50	0,72
Cd	576	2,64	4,35	164,50	1,06	56,82	0,18
Co	647	11,52	19,97	173,36	5,27	227,12	0,221
Cr	647	115,38	192,01	166,41	71,13	2940,16	33,59
Cu	647	47,34	51,02	107,77	32,31	388,05	0,20
Fe	647	4,62x10 ⁴	3,40x10 ⁴	73,64	39377,91	2,11x10 ⁵	506,23
Hg	576	0,22	0,66	299,20	0,05	8,59	0,01
Mn	647	418,33	974,46	232,94	172,60	1,56x10 ⁴	0,77
Mo	647	2,47	2,23	90,33	2,36	16,91	0,46
Ni	647	29,04	42,51	146,40	18,12	591,13	0,50
Pb	576	20,19	17,68	87,58	16,70	183,33	1,30
Sb	559	1,22	4,85	399,43	0,31	81,86	0,00
Se	559	0,45	0,56	124,97	0,28	4,62	0,02
Sr	629	17,62	55,85	317,02	6,54	815,88	0,02
V	647	123,79	118,25	95,53	81,73	705,00	2,45
Zn	647	43,96	38,07	86,60	33,90	392,70	1,00

* Adimensional

** Coeficiente de Variação

Análises geoestatísticas

O protocolo de espacialização foi aplicado sob os dados inalterados e os transformados, segundo as funções $\log(Y+1)$ e raiz quadrada de Y . As transformações tem o objetivo de aumentar o desempenho na espacialização dos elementos nos softwares avaliados, (MUKHOPADHAYA, 2016; MASSON et al., 2014; KIM et al, 2016; SOVANN e POLYA, 2010).

Os parâmetros de dependência espacial e anisotropia foram testadas por meio do software PASSaGE® (ROSENBERG e ANDERSON, 2011). Foi realizada uma avaliação do número de pares mínimos de pontos dentro do espaçamento entre pontos do semivariograma (lag) sendo definido por 30 pares (JOURNEL e HUIJBREGTS, 1978).

Com a confirmação da dependência espacial dos dados ocorreu na realização dos seguintes testes de modelos teóricos de semivariograma: esférico, exponencial, estavel e gaussiano, no software ArcGis 10.2® – Toolbars, Geostatistical Analyst. No mínimo de cinco pontos mais próximos foram utilizados na espacialização dos teores da área de estudo. Paralelamente, a krigagem ordinária no R foi executada de forma automática usando a biblioteca gstat (PEBESMA, 2004), a automap (HIEMSTRA e HIEMSTRA, 2013) e as bibliotecas de apoio, sendo estas: raster, rgdal, hydroGOF e doParallel (HIJMANS et al, 2016, BIVAND et al., 2017; ZAMBRANO-BIGIARINI e BIGIARINI, 2013). A anisotropia não foi aplicada na confecção do semivariograma teórico para a krigagem feita no R devido à ausência deste método nas bibliotecas do software. A qualidade dos resultados da predição foi avaliada por validação cruzada do tipo leave-one-out e, em seguida, foram exportados e analisados externamente.

Avaliação das krigagens

O desempenho da krigagem foi comparado pelo coeficiente de determinação para o modelo (R^2) (EQ. 1) (YANG et al., 2016) sendo calculador por:

$$R^2 = \frac{\sum_{i=1}^n (P_i - \overline{Om})^2}{\sum_{i=1}^n (Om_i - \overline{Om})^2} \quad (\text{EQ. 1})$$

Onde n é o número de observações, Om_i são os teores dos elementos observados nas transformações utilizadas (quando essas transformações forem feitas), \overline{Om} é a média dos valores observados e P_i são teores dos elementos previstos

pelos métodos em cada grupo. A partir dos valores observados e preditos foi também calculada a raiz quadrada do erro quadrado médio (EQ. 2) (HALLAK e PEREIRA FILHO, 2011; MEHTATALO et al., 2007; AHMED et al., 2015; GUO et al., 2015).

$$RMSE = \left[\frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2 \right]^{1/2} \quad (\text{EQ. 2})$$

Foi também calculado a raiz quadrada do erro quadrado médio normalizado (nRMSE) (LOPATIN, et al, 2016, LOAGUE e GREEN 1991), seu valor é calculado com o valor do RMSE, dividido pela amplitude amostral, e seu valor é expresso em percentagem (EQ. 3). O nRMSE é utilizado para comparar as diferentes amostras, quando existem diferentes variâncias amostrais ou grupo de amostras com diferente amplitudes (HAN et al., 2003; COOK e KRISHNAN, 2015). A predição é considerada “excelente” quando os valores de nRMSE são inferiores a 10%, “boa” para valores entre 10% a 20%, “regular” entre 20 a 30% e acima de 30% de nRMSE é considerada “ruim” (JAMIESON et al., 1991).

$$nRMSE = \frac{RMSE}{(\max(O_i) - \min(O_i))} \quad (\text{EQ. 3})$$

Os modelos teóricos do semivariograma com menores valores de nRMSE foram obtinham para cada elemento. A partir, dos melhores modelos teóricos do semivariograma selecionados foram confeccionados os mapas de predição destes elementos d . Para os elementos que apresentaram os melhores resultados de nRMSE (nRMSE com valores mais próximos 0), com transformações (log1p e raiz quadrada), foram feitas a transformação inversa (back-transformation) (HENGL et al., 2013).

RESULTADOS

A dependência espacial e a anisotropia foram observadas para todos os elementos e métodos avaliados. Não foi feita a krigagem do elemento Mn pelo software R em função da observação de efeito pepita puro, nos semivariogramas dos três tipos de transformações de dados (Tabela 3).

Os valores de nRMSE observados na krigagem do software R apresentaram melhor performance para a predição de 10 elementos (Al, B, Ba, Cd, Cu, Fe, Hg, Mo, Pb, Se e Zn) quando comparado aos resultados obtidos pelo ArcGis.

A transformação dos dados de entrada em log1p e raiz quadrada foram eficientes para aumentar a capacidade preditiva dos teores, em ambos os softwares para elementos (Tabela 3). Os melhores desempenhos de krigagem, com base no estimador nRMSE, ocorreram para Sr e Cd, com valores de nRMSE de 14,64 e 14,96 para os software ArcGis e R, respectivamente. Em contrapartida, os piores desempenhos foram de Hg no R e Ni no Arcgis, com valores de nRMSE de 42,39 e 26,66, respectivamente.

Os valores de nRMSE estão dentro de uma predição boa, com valores entre 10 a 20 %, para a maioria dos elementos, e no software R houve um número maior de elementos neste intervalo (Jamieson et al. 1991).

Os valores de alcance nos semivariogramas variaram entre os softwares para o mesmo elemento, com resultados de alcance maiores para 13 elementos (Al, As, Cd, Co, Cr, Fe, Hg, Pb, Sb, Sr, V e Zn), no R (Tabela 3). Estes elementos apresentaram valores de alcance variando de 10,78 a 5929,35% superiores aos alcances obtidos no Arcgis.

Tabela 3 – Resultados da validação cruzada dos dois elementos para krigagem nos softwares R e ArcGis**

Resultados ArcGis									Dados R								
Elemento	Trans**	Modelo	Patamar	Alcance (Km)	Efeito pepita	R2	RMSE	nRMSE	Elemento	Trans**	Modelo	Patamar	Alcance (Km)	Efeito pepita	R2	RMSE	nRMSE
Al	log1p	Gaussiano	0,822	9,738	0,618	0,334	0,740	22,49	Al	sem	Gaussiano	1,06E+09	85,24	7,62E+08	0,26	35321,83	18,21
As	log1p	Exponencial	1,632	10,631	1,174	0,387	0,920	20,49	As	log1p	Estável	1,57	195,19	0,00	0,07	1,13	23,21
B	sem	Gaussiano	48,96	121,275	23,775	0,185	6,257	26,64	B	sem	Estável	33,82	98,06	16,11	0,22	5,93	22,24
Ba	sqrt	Gaussiano	39,98	317,909	22,77	0,093	5,929	19,94	Ba	srqt	Estável	30,30	244,13	15,30	0,12	5,34	19,30
Cd	sem	Gaussiano	17,84	7,268	5,997	0,176	4,123	23,23	Cd	sem	Estável	16,79	8,77	1,92	0,26	3,31	14,96
Co	sqrt	Esférico	5,30	8,146	1,572	0,112	1,877	18,58	Co	log1p	Esférico	0,53	19,29	0,51	0,15	0,91	21,61
Cr	sem	Exponencial	6,60E+04	1,284	2,23E+04	0,023	205,929	16,18	Cr	sem	Esférico	23495,71	14,30	36905,05	0,02	157,31	18,09
Cu	sem	Exponencial	2204,95	11,551	1020,000	0,120	49,749	22,87	Cu	sem	Estável	2093,21	6,98	136,55	0,09	47,54	18,07
Fe	sem	Exponencial	9,32E+08	78,866	4,29E+08	0,187	3,18E+04	19,17	Fe	sem	Estável	6,97E+08	183,93	3,92E+08	0,23	29498,47	16,94
Hg	log1p	Exponencial	0,0585	10,893	0,022	0,035	0,257	19,75	Hg	log1p	Esférico	0,01	30,97	0,04	0,01	0,23	16,69
Mn	log1p	Gaussiano	3,941	4,366	2,547	0,0465	1,491	25,65	Mn*	---	---	---	---	---	---	---	---
Mo	sem	Exponencial	5,06	139,606	1,590	0,213	2,024	26,66	Mo	sem	Gaussiano	2,66	65,41	2,34	0,19	1,99	20,88
Ni	sem	Exponencial	2610,556	8,902	725,809	0,047	45,257	18,72	Ni	sem	Estável	2670,78	5,835	0,00	0,00	118,50	42,39
Pb	sem	Exponencial	480,894	6,132	66,304	0,048	18,305	24,97	Pb	sem	Estável	82,04	369,72	310,61	0,06	16,51	17,64
Sb	sem	Exponencial	81,249	2,261	0,000	0,013	5,291	15,25	Sb	log1p	Esférico	0,03	15,05	0,57	0,14	0,54	18,21
Se	sqrt	Exponencial	0,137	98,010	0,057	0,274	0,310	23,88	Se	srqt	Estável	0,13	49,20	0,01	0,30	0,30	19,16
Sr	sem	Exponencial	1848,729	8,215	8,895	0,084	55,816	14,68	Sr	srqt	Gaussiano	5,41	68,38	2,51	0,13	2,50	15,72
V	sem	Esférico	8755,030	13,362	2958,289	0,360	96,829	15,04	V	sem	Estável	3542,14	15,26	7107,80	0,32	98,13	19,17
Zn	log1p	Gaussiano	0,645	3,746	0,302	0,147	0,749	23,14	Zn	sem	Estável	1628,30	4,15	0,00	0,03	37,04	20,21

*Efeito pepita puro

** Melhores resultados para nRMSE trans em cinza

*** Transformações: sem: sem transformação dos dados, log1p: transformações dos dados por log1p e sqrt : Transformação dos dados usando raiz quadrada.

Em geral, os teores máximos e mínimos dos elementos preditos pelos softwares variaram acima de 20 %, exceto para B, Cu e Se (Tabela 4).

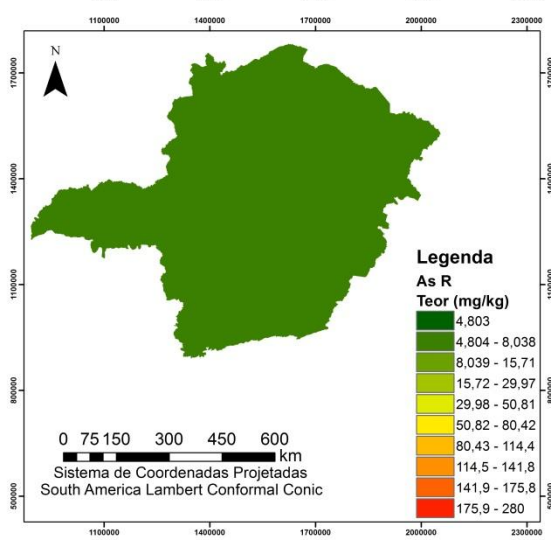
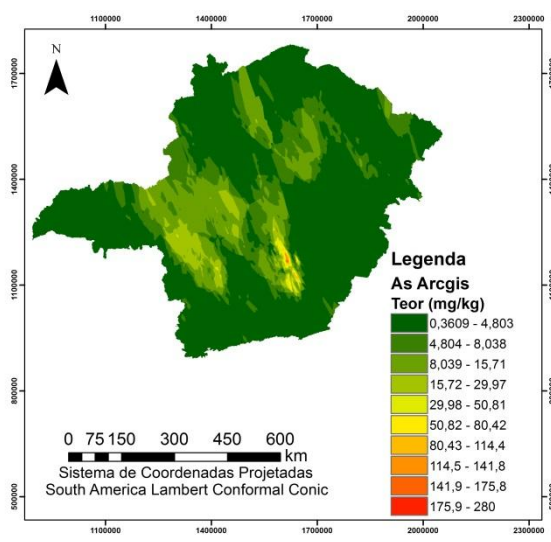
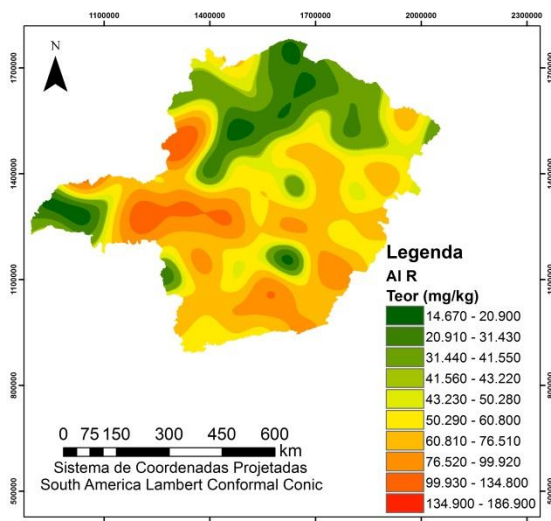
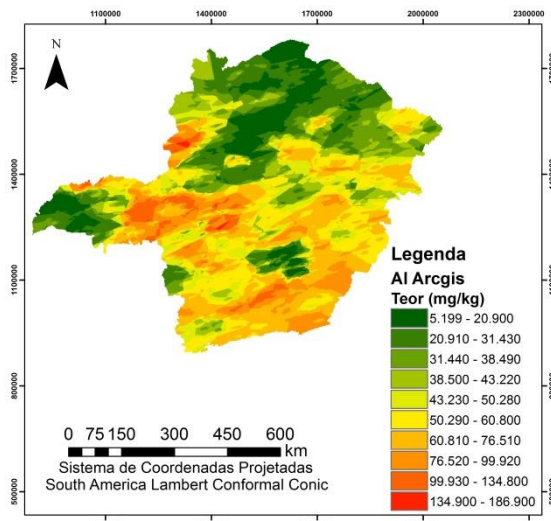
Tabela 4 – Diferença (%) entre os teores máximos e mínimos dos elementos preditos entre os dois softwares avaliados (ArcGis e R)**

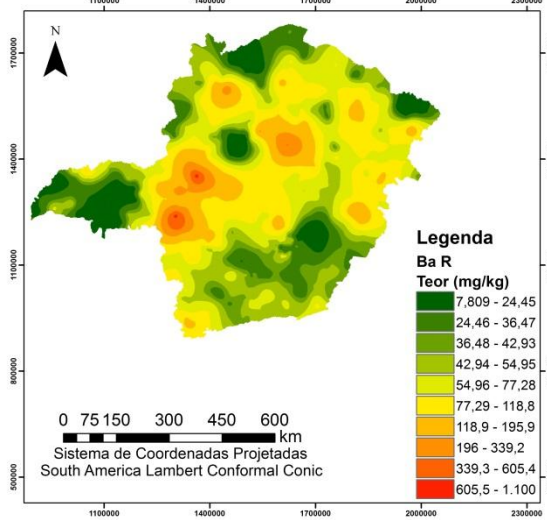
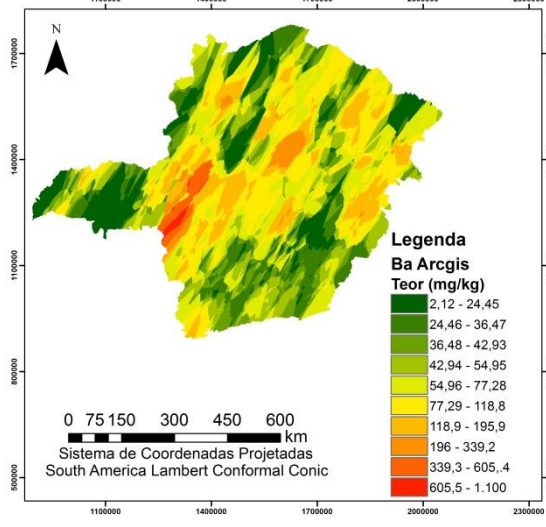
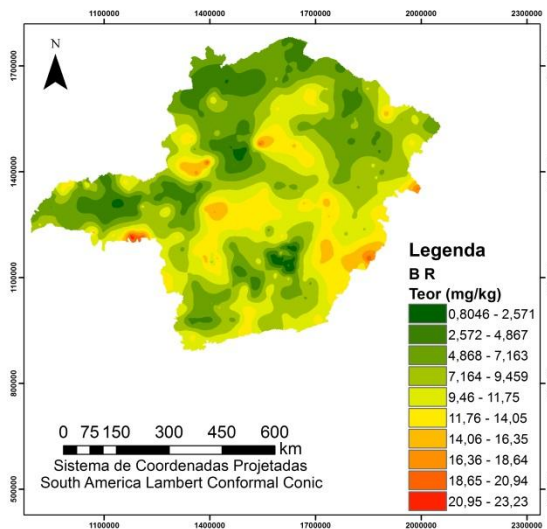
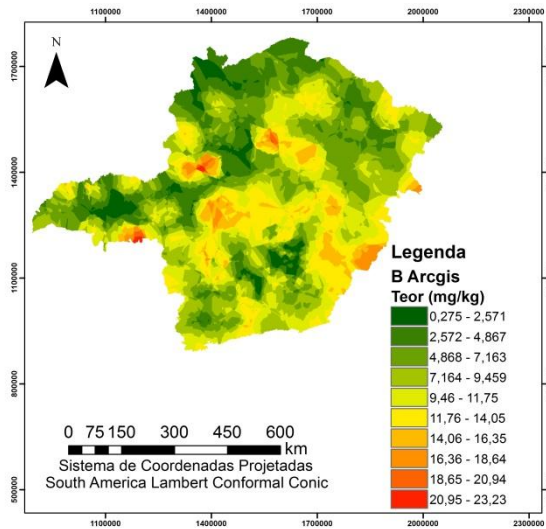
	Mínimo	Máximo
Al	182,17	43,77
As	1230,84	3807,89
B	192,58	5,98
Ba	268,35	40,66
Cd	205,25	0,38
Co	142,03	141,37
Cr	63,35	74,38
Cu	34,75	-33,65
Fe	255,16	12,15
Hg	343,30	336,29
Mn	-*	-*
Mo	99,91	56,50
Ni	-277,88	-2168,32
Pb	513,62	394,66
Sb	2809,41	6,755,00
Se	2,43	-61,81
Sr	-468,36	282,79
V	231,72	38,51
Zn	141,40	-124,06

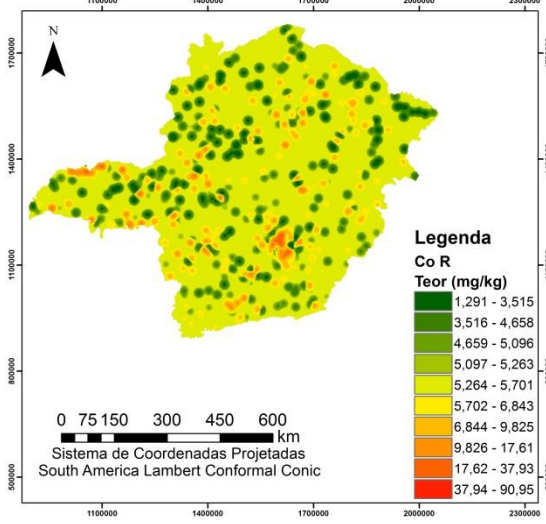
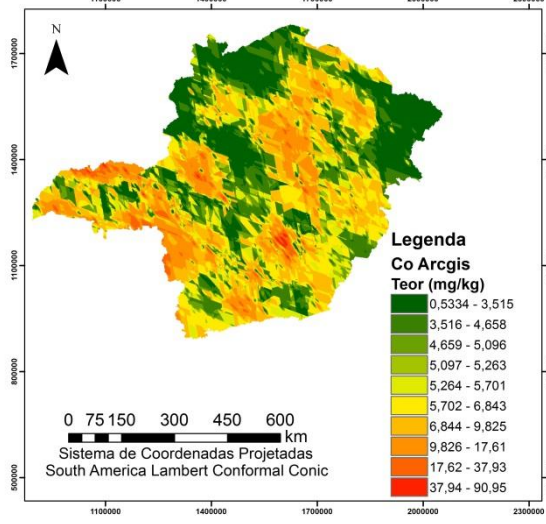
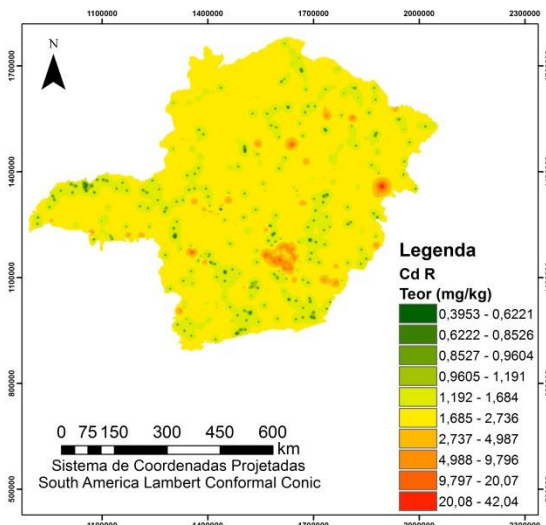
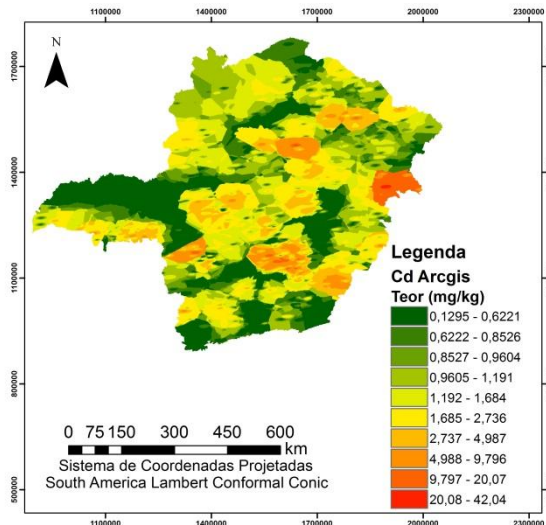
*O Mn não foi espacializado.

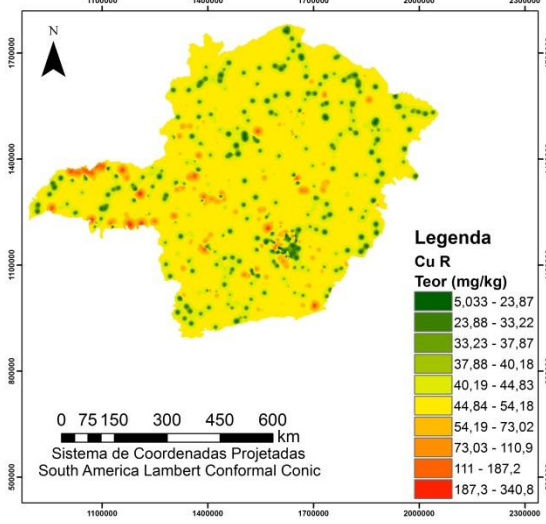
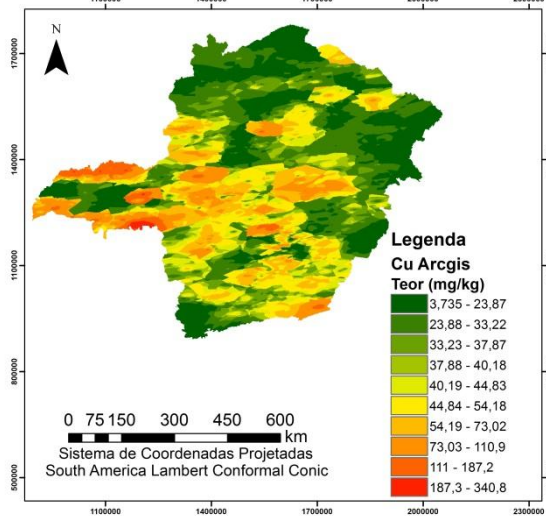
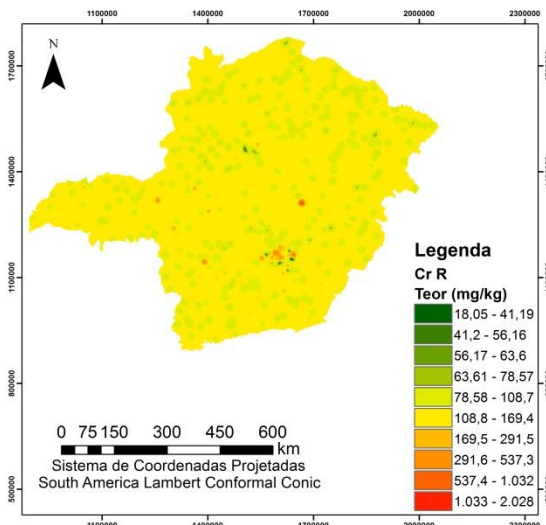
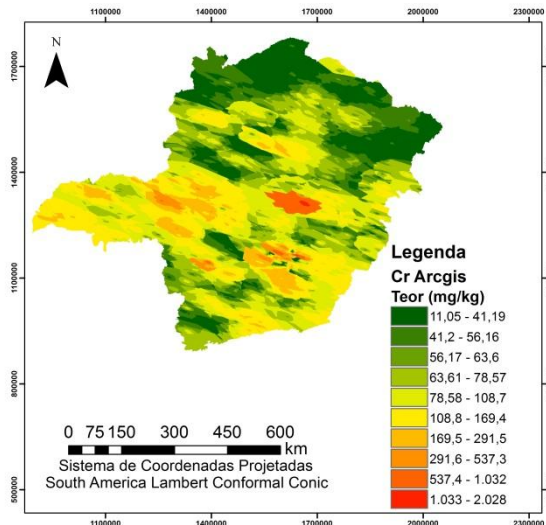
** Valores positivos: Teores do software ArcGis superiores as encontrado pelo software R; Valores negativos: Teores do software R superiores as encontrados pelo software ArcGis.

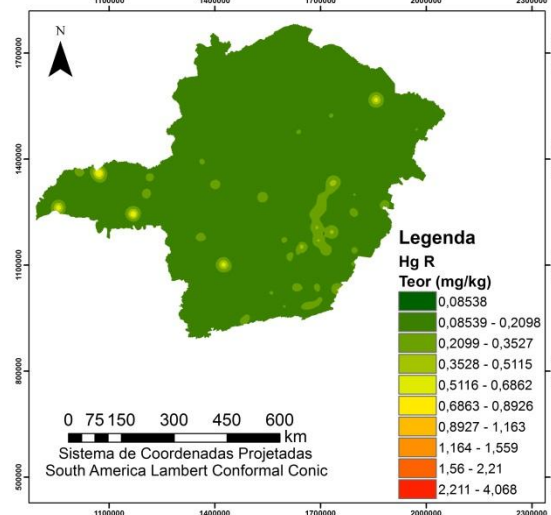
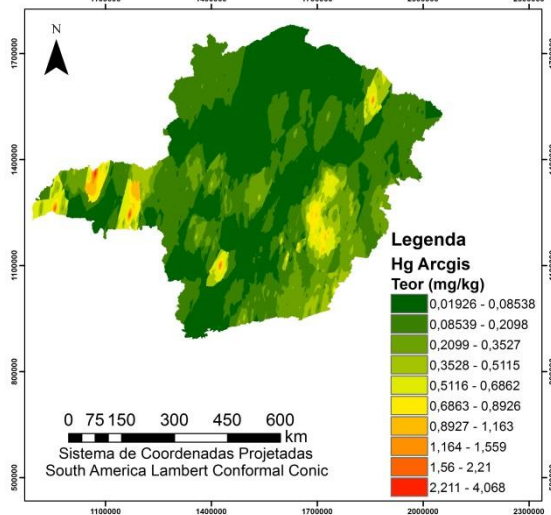
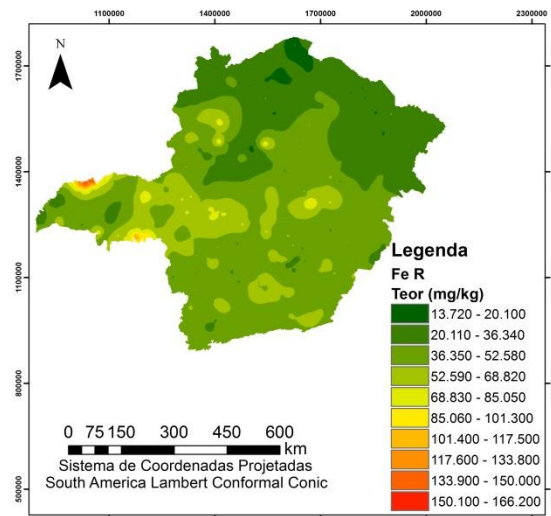
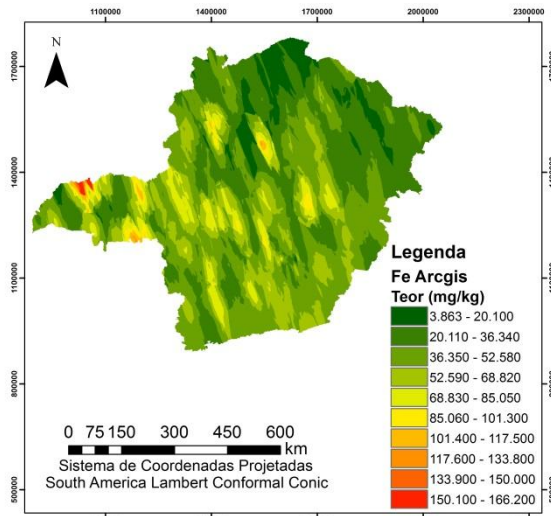
Os mapas de predição gerados pelo software R para Cd, Co, Cr, Cu, Hg, Ni, Se, V e Zn apresentaram distribuição espacial concêntrica dos teores em volta das amostras, enquanto na espacialização no Arcgis, a transformação dos dados de Shapefile Feature Class para formato raster promoveu suavização dos mapas.

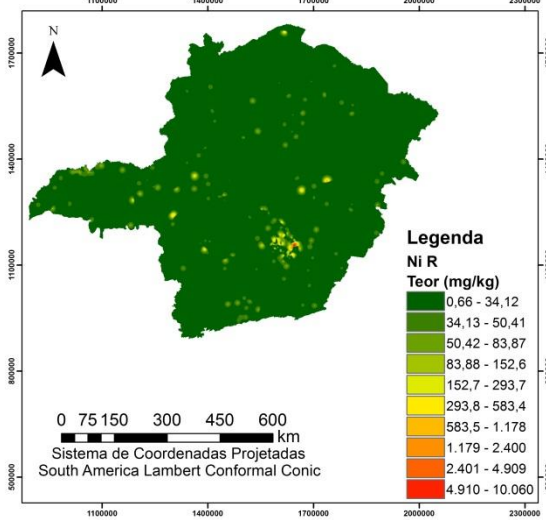
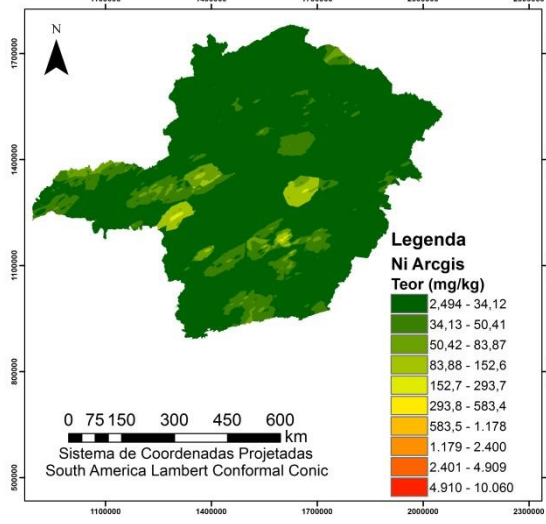
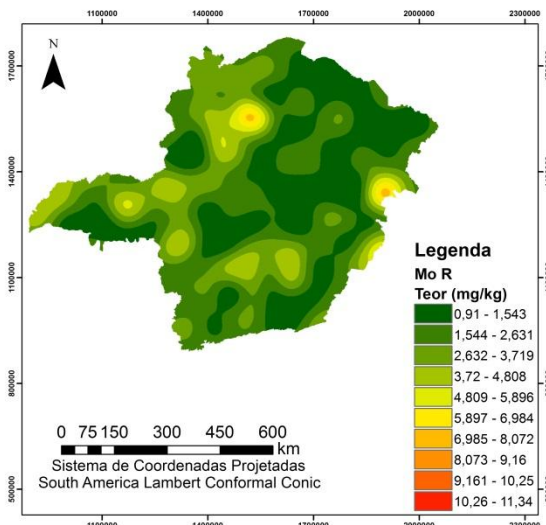
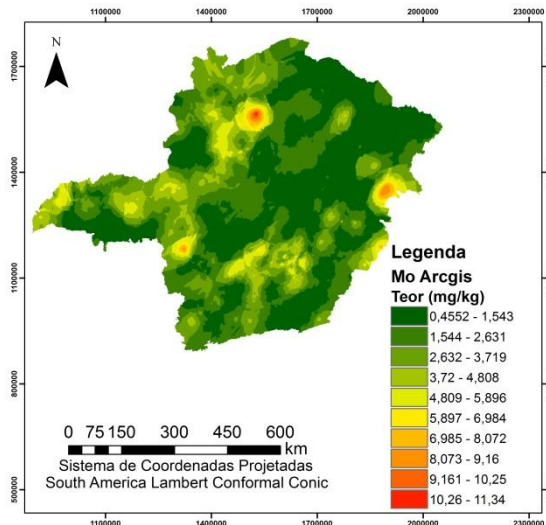


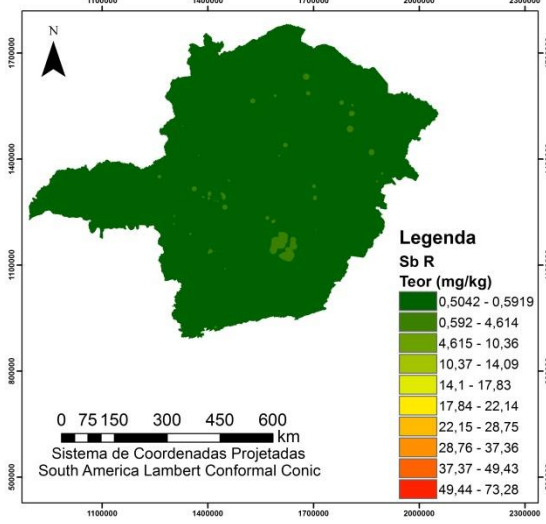
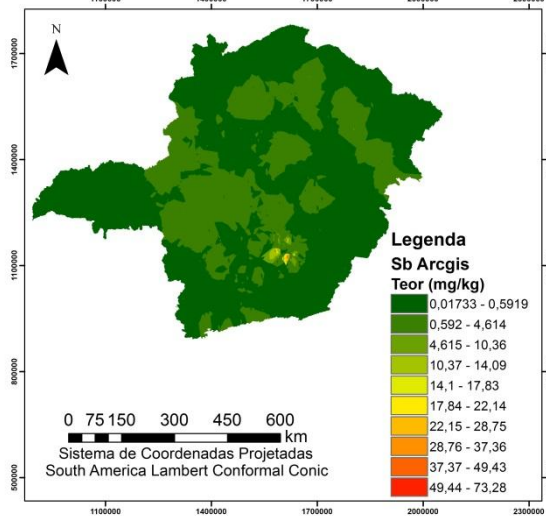
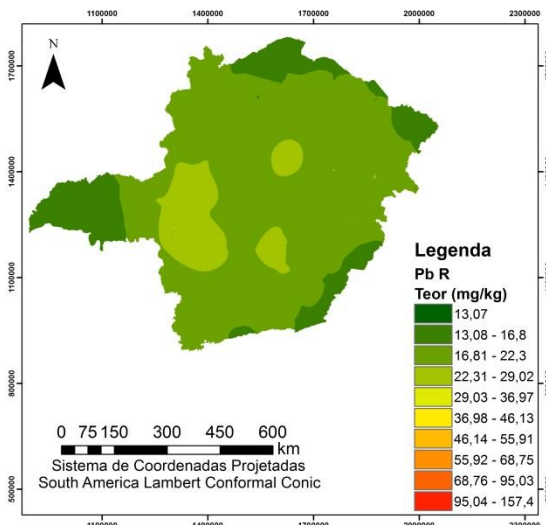
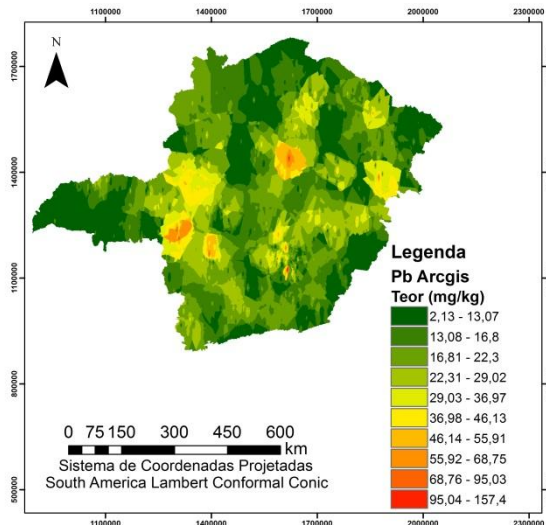


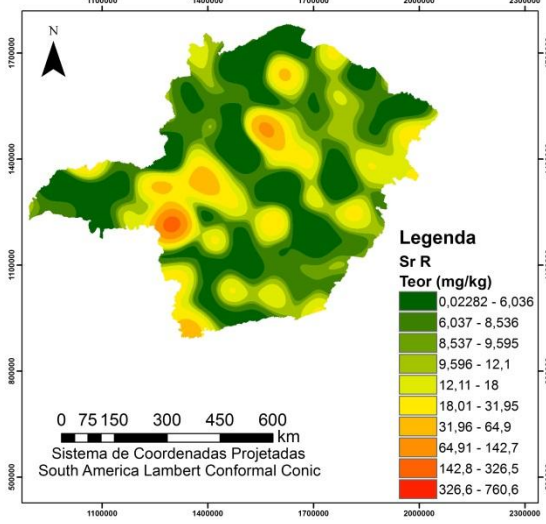
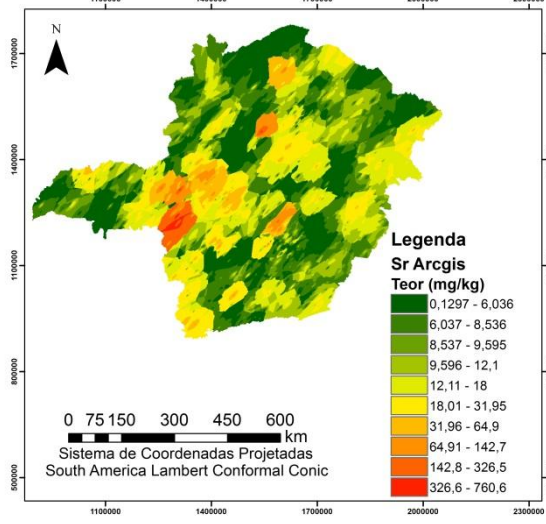
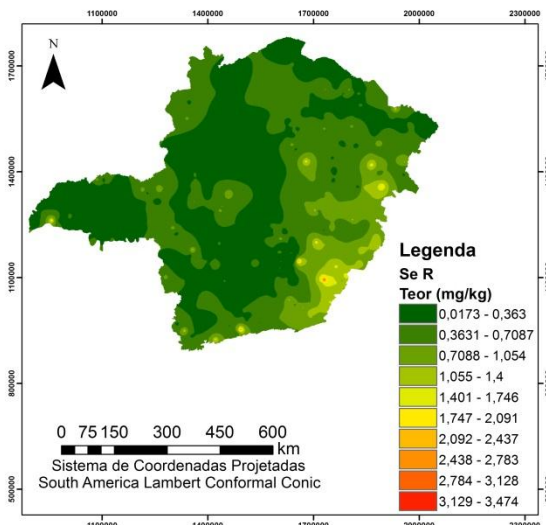
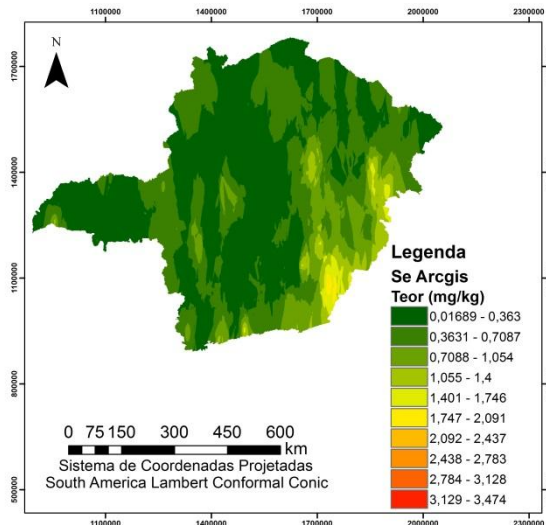












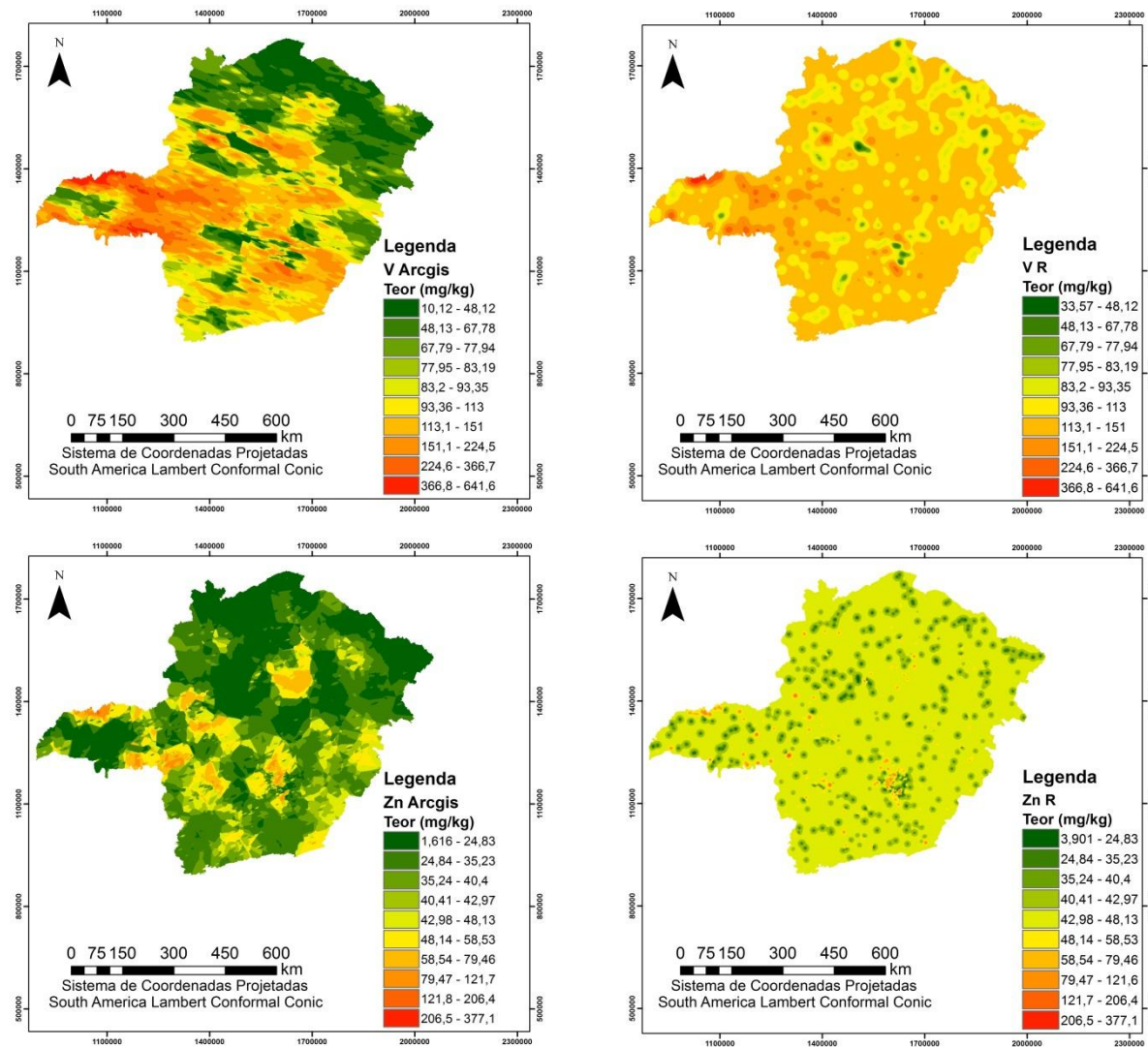


Figura 2 – Espacialização dos elementos Al, As, B, Ba, Cd, Co, Cr, Cu, Fe, Hg, Mo, Ni, Pb, Sb, Sr, Se, V e Zn para os Arcgis e R, respectivamente, e a diferenças entre os dois softwares.

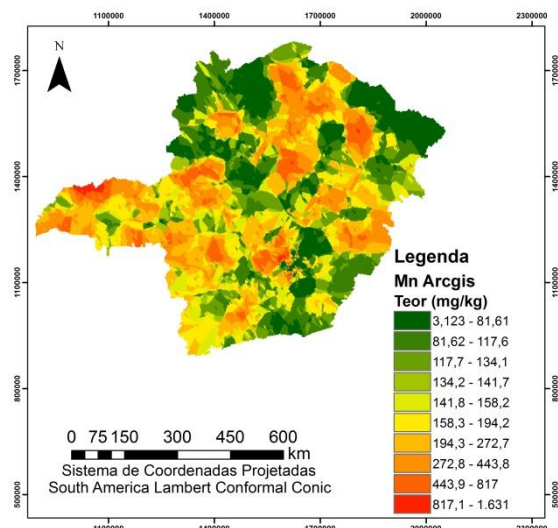


Figura 3 – Espacialização do elemento Mn feita pelo Arcgis.

DISCUSSÃO

Os resultados de nRMSE pela krigagem apresentaram menor desempenho (nRMSE mais próximo a 0) foram para os elementos B, Cd, Co, Cr, Co, Mo, Sb, Se, e V, quando comparados aos nRMSE obtidos pelos métodos aprendizado de máquinas (1ª Capítulo). Para os elementos B e V a baixas dependência espacial e densidade amostral podem explicar estes resultados. A baixa dependência espacial pode ser atribuída também à diversidade de materiais parentais dos solos no Estado que exercem esse efeito a curtas distâncias. Diversos trabalhos mostram a correlação entre propriedades físico-químicas dos solos e teores de metais pesados com a litologia (GRAY et al., 2016; LADO et al., 2008; HENGL, 2017; HENDERSON et al., 2005).

Os elementos B e V apresentaram características que diferem dos outros elementos avaliados. Por exemplo, Kabata-Pendias (2011) alta mobilidade do Mo, sendo a maior entre os micronutrientes e apresenta lixiviação com a lixiviação da água nos solos sendo acumulado na superfície do solos em regiões áridas e semi-áridas, esse processo pode levar teores de maior homogeneidade em regiões próximas que indica uma maior dependência espacial e levando a resultados na melhor desempenho da krigagem do Mo. O V tende a ser mais móvel sob a forma do cátion VO^{2+} em solos ácidos, que são comuns na região de estudo (KABATA-PENDIAS, 2011, De SOUZA et. al, 2013). Assim, as características ácidas dos solos amostrados podem acarretar uma diminuição dos teores de V em áreas com elevado teor natural e, por consequência, que ocasiona uma distribuição mais espacial mais adequada

para krigagem com heterogeneidade em maiores distâncias e homogêneas entre amostras próximas.

Para os elementos Cd, Co Sb e Sr, os resultados de predição foram melhores pela krigagem, mostrando o melhor desempenho desse método comparado com aprendizado de máquinas. Contudo pela avaliação do mapa da predição pode-se observar o aparecimento de olhos em voltas das amostras para os elementos citados anteriormente, esse fato está relacionado ao baixo alcance do semivariograma, que produz um mapa que não demonstra a realidade da espacialização destes elementos (Figura 2).

O desempenho superior da krigagem para os elementos Mo e Sb, pode estar relacionado a menor amplitude dos dados amostrais desses elementos (Tabela 1). A maior homogeneidade local das amostras também seria uma explicação dos melhores resultados da krigagem dos elementos Cr e V.

Uma explicação para o baixo desempenho da krigagem para os elementos que tiveram melhores resultados pelos métodos de aprendizado de máquinas (Al, As, B, Co, Fe, Hg, Mn, Ni, Pb, Se e Zn), possivelmente devido ao tipo de grid utilizado na amostragem, e que interferi no desempenho da predição. Nesse sentido, o grid de amostragem neste estudo caracterizou-se por ser aleatório estratificado. Por sua vez, este grip apresenta problemas de distribuição espacial e, conseqüentemente, resultados inferiores aos obtidos em um estudo com grid regular (YAMAMOTO e LANDIM, 2013).

Para os elementos que tiveram espacialização (menor NRMSE) na krigagem do software Arcgis, pode ser atribuídos aos ajustes do semivariograma teórico, enquanto aa krigagem no R o ajuste é feito automaticamente (HIEMSTRA e HIEMSTRA 2013), para reduzir tempo de processamento. O ajuste do semivariograma do Arcgis é realizado para possibilitar estes ajustes manuais.

A possibilidade de aplicação da anisotropia no Arcgis é um fator que explicaria os melhores desempenhos da krigagem neste software. A ansiotropica ocorre quando há a distribuição desigual do elemento em todas as direções (GUEDES et al, 2008).

Os alcances do semivariograma estimados foram maiores que os obtidos no estudo de De SOUZA et. al, (2013), para 14 elementos (Al, B, Ba, Cd, Cu, Fe, Hg, Mn, Mo, Pb, Sr, V, Zn) dos 17 elementos avaliados. Não foi possível a comparação dos dados estatísticos dos estudos da validação cruzada, pois os parâmetros são exclusivos do software Arcgis (Mean Standardized Error e Mean Standardized Error).

A explicação dos maiores alcances estaria relacionada à releitura de amostras contidas no centro do Estado na região denominada de Quadrilátero Ferrífero, onde estão contidos os maiores teores para As, Cd, Co, Cr, Mn, Ni e Sb (Figura 2).

A ferramenta interna Geostatistical Analyst possibilitou o cálculo da validação cruzada dos valores a partir dos valores de RMSE, mas não apresentaram valores de R^2 e MAE, nRMSE. Estes parâmetros podem ser calculados externamente com extração dos valores preditos pela validação cruzada (ARSLAN, 2012; NAS et al, 2009; NAS e BERKTAY, 2010; MERWADE et al. 2006; TRIPATHI et al 2015; XIE et al. 2011). Alguns autores apresentam os dados do R^2 modelo, mas um elevado valor alto de R^2 modelo do semivariograma pode apresentar baixa performance para a predição da krigagem (ADHIKARY et al 2010, BEELEN et al., 2009; DELBARI et al., 2013; IHL et al, 2015; WERE et al, 2016). Explore isso melhor.

Os maiores valores máximos de predição obtidos pelo ArcGis em comparação ao R pode ser atribuído ao menor número de vizinhos considerados na krigagem. O primeiro software fez uso de 5 vizinhos, enquanto o software R utilizou todo o banco de amostras, dentro do raio de alcance esse fato leva a uma predição mais exata com a redução de possíveis efeitos espúrios de alguma amostra vizinha.

No ArcGis, a seleção do número pontos próximos para predição de qualquer ponto de predição reduziu o efeito de amortecimento da predição dos teores, que ocorre quando se prediz os teores dos elementos com o uso de todas as amostras dentro do alcance. Isso levou a espacialização de predições com teores em áreas não amostradas, que estarão associados aos valores amostrais próximos e aos teores mais próximos em regiões com amostras mais altas. Isso explicaria os teores mais elevados das interpolações do software do Arcgis. Outro ponto importante na krigagem do Arcgis foi o tipo de dado gerado, no formato Shapefile Feature Class, que não pode ser modificado posteriormente. A modificação nos resultados da krigagem requer a transformação dos dados para raster. Porém, esta transformação reduz a qualidade do resultado deixando o mapa com qualidade inferior ao resultado gerado pelo R (Figura 2).

A transformação dos dados indicou uma maior eficiência do desempenho da predição pela krigagem, gerando menores valores de nRMSE. Entretanto, o uso das transformações reduziu a possibilidade de comparação estatístico sendo somente possível comparar as krigagem testadas com dados normalizados como o nRMSE e nMAE.

Os baixos valores de alcances, menores que 47 km, obtidos no R afetaram a espacialização dos teores. Isto gerou uma distribuição espacial concêntrica às amostras, sem grande representatividade ao longo da área do Estado. Esses nódulos foram observados em menor escala nos mapas de krigagem no Arcgis, principalmente, em amostras com teores altos, o que demonstrou a dependência espacial dos elementos para estes elementos baixa.

Um elevado alcance é ponto fundamental para predição em grandes áreas como foi o caso deste estudo, que houve uma associação entre o alcance à dependência espacial das amostras. Dessa forma, para conseguir um aumento do alcance é necessária uma maior dependência espacial entre as amostras. O aumento no número de amostras pode aumentar a dependência espacial na área de estudo ou reduzir a área de estudo a fim de conseguir uma área com maior homogeneidade das amostras dentro dessa área. A redução da área também reduz a necessidade de um elevado alcance, mas a dependência espacial também depende da distribuição amostral, dessa forma um elemento com distribuição aleatória não leva ao aumento do alcance do semivariograma.

Por fim o devido discutido anteriormente, não foi possível definir um software entre os dois estudados. A escolha dependerá do banco de dados, da transformação que se deseja realizar com os dados e a dependência espacial das amostras. O uso do software R apresentou características convidativas, por ser um software “open source” e pela possibilidade de automação do processo. Enquanto isso, a Arcgis apresentou pontos positivos em relação à capacidade de escolha do número de vizinhos e por consequência uma amplitude de dados mais próxima da amplitude da variável na qual se deseja proceder à espacialização, mas a extração formato Shapefile Feature Class para formato raster reduziu a qualidade da predição sendo necessário levar esse fato em consideração.

CONCLUSÃO

Os elementos As, Co, Cr, Mn, Ni, Sb, Sr e V obtiveram melhores performance da krigagem para o software ArcGis, enquanto os outros elementos (Al, B, Ba, Cd, Cu, Fe, Hg, Mo, Pb, Se, Zn) tiveram melhores performance da krigagem via R.

O software R apresentou baixos valores de alcance, gerando uma distribuição espacial concêntrica dos teores em volta amostras com altos teores para krigagem dos elementos Cd, Co, Cr, Hg, Ni, Se, V e Zn.

A geração do dado na krigagem pelo ArcGis no formato Shapefile Feature Class torna um impeditivo ao seu uso, visto que, não pode ser modificado posteriormente. A modificação nos resultados da krigagem requer a transformação dos dados para raster. Porém, esta transformação reduz a qualidade do resultado deixando o mapa com qualidade inferior ao resultado gerado pelo R.

REFERENCIAS BIBLIOGRÁFICAS

ABRAHÃO, W. A. F; MARQUES, J. J. **Manual de procedimentos analíticos para determinação de VRQ de elementos-traço em solos do Estado de Minas Gerais**. Belo Horizonte, 2013. 28 p

ADHIKARY, Partha Pratim et al. Assessment of groundwater pollution in West Delhi, India using geostatistical approach. **Environmental Monitoring and Assessment**, v. 167, n. 1-4, p. 599-615, 2010.

AHMED, O. S. et al. Characterizing stand-level forest canopy cover and height using landsat time series, samples of airborne LiDAR, and the random forest algorithm. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 101, p. 89-101, 2015.

ARSLAN, H.. Spatial and temporal mapping of groundwater salinity using ordinary kriging and indicator kriging: the case of Bafra Plain, Turkey. **Agricultural Water Management**, v. 113, p. 57-63, 2012.

BEELLEN, Rob et al. Mapping of background air pollution at a fine spatial scale across the European Union. **Science of the Total Environment**, v. 407, n. 6, p. 1852-1867, 2009.

BIVAND, R. et al. **Package “rgdal.”** R Package. (2016)

CASTRO, Fábio da S. et al. Avaliação do desempenho dos diferentes métodos de interpoladores para parâmetros do balanço hídrico climatológico. **Revista Brasileira de Engenharia Agrícola e Ambiental-Agriambi**, v. 14, n. 8, 2010.

CAMARGO, L. A. et al. Variabilidade espacial de atributos mineralógicos de um latossolo sob diferentes formas do relevo: II-correlação espacial entre mineralogia e agregados. **Revista Brasileira de Ciência do Solo**, p. 2279-2288, 2008.

CEMPEL, M.; NIKEL, G. Nickel: a review of its sources and environmental toxicology. **Polish Journal of Environmental Studies**, v. 15, n. 3, p. 375-382, 2006.

CONSELHO NACIONAL DO MEIO AMBIENTE (CONAMA). Resolução nº 420, de 28 de dezembro de 2009; 2009 [Brasil].

COPAM-CONSELHO ESTADUAL DE POLÍTICA AMBIENTAL,. Deliberação Normativa nº 166, de 29 de junho de 2011. **Altera o Anexo I da Deliberação Normativa Conjunta COPAM/CERH**, n. 2.

COOK, D. J.; KRISHNAN, N. C. **Activity learning: discovering, recognizing, and predicting human behavior from sensor data**. John Wiley & Sons, 2015.

DA SILVA JÚNIOR, João F. et al. Classificação numérica e modelo digital de elevação na caracterização espacial de atributos dos solos. **Revista Brasileira de Engenharia Agrícola e Ambiental**, p. 415-424, 2012 (a).

- DA SILVA JUNIOR, João F. et al. Simulação geoestatística na caracterização espacial de óxidos de ferro em diferentes pedofomas. **Revista Brasileira de Ciência do Solo**, v. 36, n. 6, 2012 (b).
- DE SOUZA, G. S. et al. Krigagem ordinária e inverso do quadrado da distância aplicados na espacialização de atributos químicos de um argissolo. **Scientia Agraria**, v. 11, n. 1, 2010.
- DE VOS, W. et al. Geochemical atlas of Europe. Part 2—interpretation of geochemical maps, additional tables, figures, maps. **and Related Publications.—Geological Survey of Finland. Espoo, Finland. Geol**, v. 98, n. 2, p. 69-82, 2006.
- DELBARI, Masoomeh et al. Investigating spatio-temporal variability of groundwater quality parameters using geostatistics and GIS. **Intl. Res. J. Appl. Basic. Sci**, v. 4, n. 10, p. 3623-3632, 2013.
- DE SOUZA, J. J. L. L. et al. Geochemistry and spatial variability of metal (loid) concentrations in soils of the state of Minas Gerais, Brazil. **Science of the Total Environment**, v. 505, p. 338-349, 2015.
- DE MELLO, Carlos Rogério et al. Erosividade mensal e anual da chuva no Estado de Minas Gerais. **Pesquisa Agropecuária Brasileira**, v. 42, n. 4, p. 537-545, 2007.
- GÖHRE, Vera; PASZKOWSKI, Uta. Contribution of the arbuscular mycorrhizal symbiosis to heavy metal phytoremediation. **Planta**, v. 223, n. 6, p. 1115-1122, 2006.
- GRAY, Jonathan M.; BISHOP, Thomas FA; WILFORD, John R. Lithology and soil relationships for soil modelling and mapping. **Catena**, v. 147, p. 429-440, 2016.
- GUEDES, Luciana Pagliosa Carvalho et al. Anisotropia no estudo da variabilidade espacial de algumas variáveis químicas do solo. **Revista Brasileira de Ciência do Solo**, v. 32, n. 6, p. 2217-2226, 2008.
- GUO, P et al. Digital mapping of soil organic matter for rubber plantation at regional scale: An application of random forest plus residuals kriging approach. **Geoderma**, v. 237, p. 49-59, 2015.
- HALLAK, R.; PEREIRA FILHO, A. J.. Metodology for performance analysis of simulations of convective systems in the metropolitan area of São Paulo with the ARPS model: sensitivity to variations with the advection and the data assimilation schemes. **Revista Brasileira de Meteorologia**, v. 26, n. 4, p. 591-608, 2011.
- HAN, K.; VIAU, A. A.; ANCTIL, F.. High-resolution forest fire weather index computations using satellite remote sensing. **Canadian journal of forest research**, v. 33, n. 6, p. 1134-1143, 2003.
- HENGL, T. et al. SoilGrids250m: Global gridded soil information based on machine learning. **PloS one**, v. 12, n. 2, p. e0169748, 2017.
- HENGL, T.; NIKOLIĆ, M.; MACMILLAN, R. A. Mapping efficiency and information content. **International Journal of Applied Earth Observation and Geoinformation**, v. 22, p. 127-138, 2013.
- HENDERSON, B. L. et al. Australia-wide predictions of soil properties using decision trees. **Geoderma**, v. 124, n. 3, p. 383-398, 2005.
- HIEMSTRA, P.; HIEMSTRA, M. P. Package 'automap'. **compare**, v. 105, p. 10, 2013.
- HIJMANS, R.J. et al, Package 'raster'. 2016.
- IHL, T. et al. Concentration of toxic elements in topsoils of the metropolitan area of Mexico city: a spatial analysis using ordinary kriging and indicator kriging. **Rev Int Contam Ambient**, v. 31, p. 47-62, 2015.

JAMIESON, P. D.; PORTER, J. R.; WILSON, D. R. A test of the computer simulation model ARCWHEAT1 on wheat crops grown in New Zealand. **Field crops research**, v. 27, n. 4, p. 337-350, 1991.

JOURNEL, A. G.; HUIJBREGTS, Ch J. **Mining geostatistics**. Academic press, 1978.

KABATA-PENDIAS, A. et al. **Trace elements in soils and plants**. Boca Raton: CRC press, 2010.

KIM, T. J.; BULLOCK, B. P.; WIJAYA, A.. Spatial Interpolation of Above-Ground Biomass in Labanan Concession Forest in East Kalimantan, Indonesia. **Mathematical and Computational Forestry & Natural Resource Sciences**, v. 8, n. 2, p. 26, 2016.

LADO, L. R.; HENGL, T.; REUTER, H. I. Heavy metals in European soils: a geostatistical analysis of the FOREGS Geochemical database. **Geoderma**, v. 148, n. 2, p. 189-199, 2008.

LIMA, G. C. et al. Variabilidade de atributos do solo sob pastagens e mata atlântica na escala de microbacia hidrográfica. **Revista Brasileira de Engenharia Agrícola e Ambiental-Agriambi**, v. 18, n. 5, 2014.

LOAGUE, K.; GREEN, R. E. Statistical and graphical methods for evaluating solute transport models: overview and application. **Journal of contaminant hydrology**, v. 7, n. 1-2, p. 51-73, 1991.

LOPATIN, J. et al. Comparing generalized linear models and random forest to model vascular plant species richness using LiDAR data in a natural forest in central Chile. **Remote sensing of environment**, v. 173, p. 200-210, 2016.

MASSON, D. ; FREI, C.. Spatial analysis of precipitation in a high-mountain region: exploring methods with multi-scale topographic predictors and circulation types. **Hydrology and Earth System Sciences**, v. 18, n. 11, p. 4543, 2014.

MERWADE, V. M.; MAIDMENT, David R.; GOFF, John A. Anisotropic considerations while interpolating river channel bathymetry. **Journal of Hydrology**, v. 331, n. 3, p. 731-741, 2006.

MEHTATALO, L.; MALTAMO, M.; KANGAS, A.. The use of quantile trees in the prediction of the diameter distribution of a stand. **Silva Fennica**, v. 40, n. 3, p. 501, 2007.

MUKHOPADHAYA, S. Rainfall Mapping using Ordinary Kriging Technique: Case Study: Tunisia. **Journal of Basic and Applied Engineering Research**. V.3, n. 1, 2016

NAS, B. et al. Geostatistical approach to assessment of spatial distribution of groundwater quality. **Polish Journal of Environmental Studies**, v. 18, n. 6, p. 1073-1082, 2009.

NAS, B.; BERKTAY, A. Groundwater quality mapping in urban groundwater using GIS. **Environmental monitoring and assessment**, v. 160, n. 1, p. 215-227, 2010.

OLIVEIRA JUNIOR, J. C. DE et al. Variabilidade espacial de atributos mineralógicos de solos da formação Guabirotuba, Curitiba (PR). **Revista Brasileira de Ciência do Solo**, v. 35, n. 5, 2011.

PAIVA, Y. G. et al. Delimitação de sítios florestais e análise dos fragmentos pertencentes na bacia do rio Itapemirim. **Idesia (Arica)**, v. 28, n. 1, p. 17-22, 2010.

PEBESMA, Edzer J. Multivariable geostatistics in S: the gstat package. **Computers & Geosciences**, v. 30, n. 7, p. 683-691, 2004.

ROSENBERG, Michael S.; ANDERSON, Corey Devin. PASSaGE: pattern analysis, spatial statistics and geographic exegesis. Version 2. **Methods in Ecology and Evolution**, v. 2, n. 3, p. 229-232, 2011.

SOVANN, C.; POLYA, D.. **Using GIS and Geostatistics to Develop Hazards and Risk Maps of Arsenic in Shallow Groundwaters of Cambodia**. University of Manchester, 2010.

TRIPATHI, R. et al. Characterizing spatial variability of soil properties in salt affected coastal India using geostatistics and kriging. **Arabian Journal of Geosciences**, v. 8, n. 12, p. 10693-10703, 2015.

WERE, Kennedy; SINGH, Bal Ram; DICK, Øystein Bjarne. Spatially distributed modelling and mapping of soil organic carbon and total nitrogen stocks in the Eastern Mau Forest Reserve, Kenya. **Journal of Geographical Sciences**, v. 26, n. 1, p. 102-124, 2016.

YAMAMOTO, J. K.; LANDIM, P. M. B. **Geoestatística: Conceitos e aplicações**. Oficina de Textos: So Paulo, SP, Brasil, 2013.

YANG, R. et al. 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.

XIE, Y. et al. Spatial distribution of soil heavy metal pollution estimated by different interpolation methods: accuracy and uncertainty analysis. **Chemosphere**, v. 82, n. 3, p. 468-476, 2011.

ZAMBRANO-BIGIARINI, M.; BIGIARINI, M. M. Z.. Package 'hydroGOF'. 2013.

CAPÍTULO 4

COMPARAÇÃO ENTRE MÉTODOS DE ESPACIALIZAÇÃO DE DADOS DE METAIS E METALÓIDES PARA ESTADO DE MINAS GERAIS

INTRODUÇÃO

Cada vez mais necessário o entendimento da distribuição dos elementos químicos nos solos, principalmente elementos com efeitos ecotoxicológicos, tais como: As, Ag, Cd, Cr, Co, Pb, Hg, Ni, Se, Sb (CEMPEL e NIKEL, 2006; GÖHRE e PASZKOWSKI, 2006). Os solos são o produto de fatores de formação como material de origem, relevo, clima, organismos e tempo (JENNY, 1941), assim, os teores dos elementos são proeminentes destes fatores e teor oriundo da rocha matriz.

Em 1993, a União Internacional de Ciências Geológicas e a Associação de Geoquímica e Cosmoquímica iniciaram um projeto de definição do *geochemical baseline* para determinar o teor de elementos químicos nos solos em todos os continentes (SALMINEM, 2004). No Brasil, os primeiros estudos começaram no estado de São Paulo, para determinar os teores em áreas não antropizadas e criar do background geoquímico (CETESB, 2001). O Estado de Minas Gerais estabeleceu os valores de referência em 2011 por meio do Conselho de Política Ambiental do estado de Minas Gerais liberou uma deliberação normativa nº 116, de 29 de Julho de 2011, que altera o Anexo I da deliberação Normativa conjunta COPAM CERH nº 2 de 6 de setembro de 2010. Assim, foram estabelecidos os Valores de Referência de Qualidade (VRQ) dos Solos para o Estado de Minas Gerais e o trabalho foi coordenado pela Fundação Estadual de Meio Ambiente (FEAM).

A deliberação 166/2011 correspondeu ao primeiro esforço para proteção da população, fauna, flora, qualidade do solo, qualidade das águas etc, tendo como enfoque o controle dos teores de elementos químicos em Minas Gerais. Sua atualização é necessária devido à grande diversidade existente no estado, tanto em litologia quanto em solos. Das treze ordens de solos existentes o Estado de Minas Gerais apresenta as trezes dessas ordens, sendo estas: Latossolos, Cambissolos, Neossolos, Argissolos, Plintossolos, Gleissolos, Nitossolos, Neossolos, Planossolos, Luvisolos, Espodossolos, Organossolos, Vertissolos.

A grande diversidade de ordens de solos dentro do Estado se reflete em uma desuniformidade da distribuição espacial dos teores em Minas Gerais (DE SOUZA et al., 2015). Ao analisar o mapa geológico de Minas Gerais observa-se que o Estado de Minas Gerais é inserido dentro de quatro grandes províncias geológicas, sendo elas a Bacia do Paraná, do Mantiqueira, do São Francisco e do Tocantins (COMIG, 2003). Estas províncias geológicas apresentam composições químicas diferentes entre si e também as mesmas não são homogêneas. Dessa forma, a utilização de um valor de VRQ para todo o Estado apresentaria uma grande discrepância da realidade na maioria dos casos.

A tomada de decisão é uma ferramenta utilizada para executar a espacialização dos teores e a definição de um VRQ para áreas mais homogêneas dentro da área de estudo. De Souza (2013) gerou a espacialização dos teores de elementos químicos para o Estado de Minas Gerais por meio de krigagem ordinária. São escassos os estudos de espacialização de teores de elementos químicos nos solos que não estejam relacionados à fertilidade do solo. Ademais, grande parte dos trabalhos de espacialização é executada por processos de krigagem e regressão krigagem (LIU et al, 2006; ZHENG et al, 2008). Outra metodologia de espacialização de variáveis pedológicas são os métodos de aprendizado de máquinas (AHMAD et al., 2010; BRUNGARD et al., 2015; FORKUOR et al, 2017; HENDERSON et al., 2005; HENGL et al., 2017; KOVAČEVIĆ et al., 2010). Este trabalho objetivou comparar algoritmos de aprendizagem de máquinas e krigagem ordinária nas previsões de metais e metaloides em solos do Estado de Minas Gerais.

MATERIAL E MÉTODOS

Banco de dados

Foram utilizadas 648 amostras georreferenciadas obtidas do banco de dados de solos do Estado de Minas Gerais (Figura 1). Estas amostras foram coletadas a uma profundidade 0-20 cm. O processo de amostragem e quantificação dos elementos está descrito em De Souza et al (2015) e Abrahão e Marques (2013). Na Tabela 1, encontra-se a estatística descritiva dos elementos avaliados:.

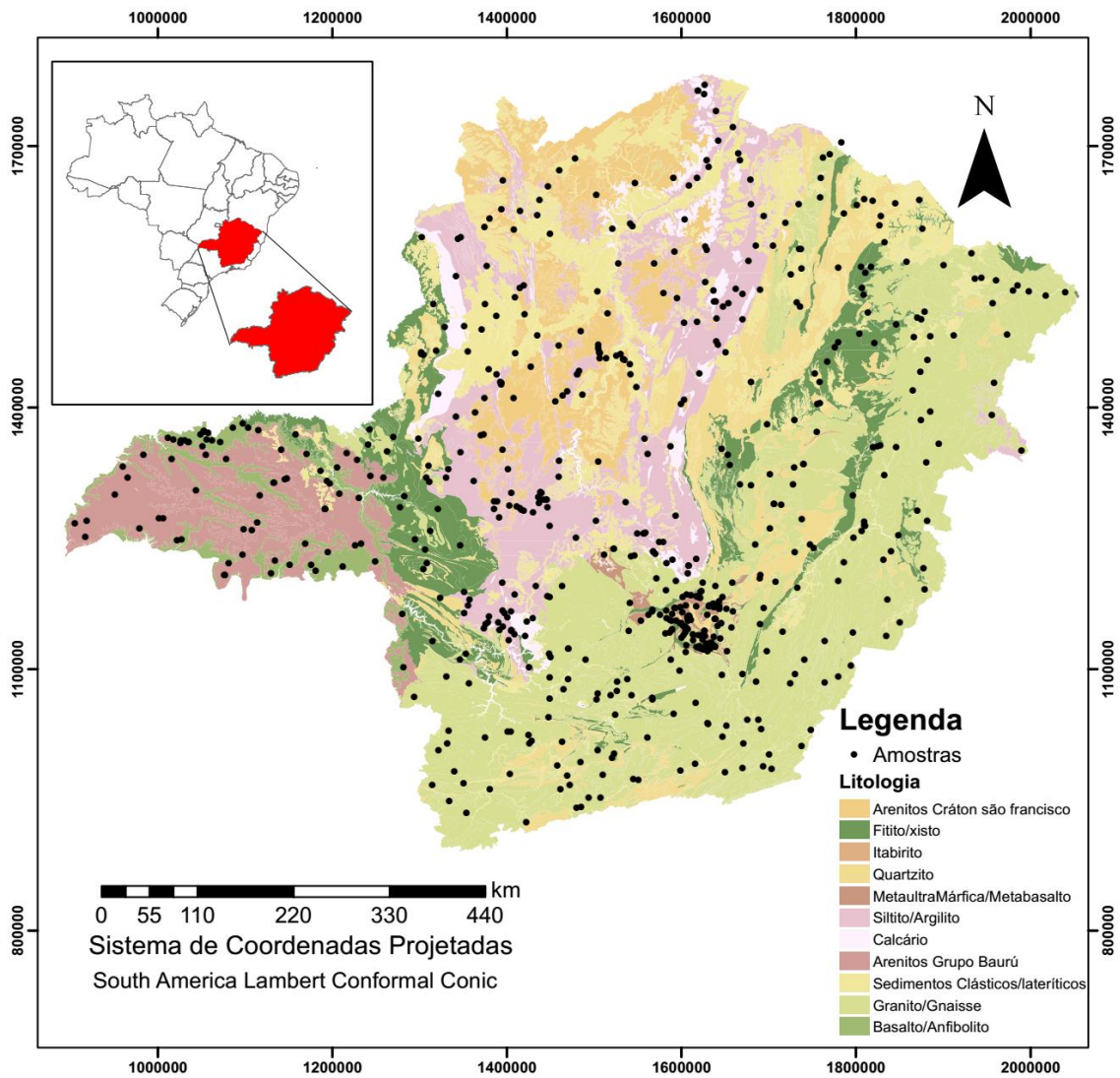


Figura 1. Distribuição de pontos coletados e mapa litológico simplificado da área estudo.

Tabela 1 – Estatísticas descritivas das amostras de solos na cama de 0-20 cm , para os teores de metais e metaloides;

	Número de amostras	Média	Desvio Padrão	CV**	Mediana	Máximo	Mínimo
Elemento	*	mg/kg	mg/kg	%	mg/kg	mg/kg	mg/kg
Al	647	5,55x10 ⁴	4.11x10 ⁴	74,02	4.70x10 ⁴	2,51x10 ⁵	850,99
As	576	16,17	58,13	359,49	4,26	854,51	0,13
B	647	7,71	6,74	87,42	6,67	43,64	0,28
Ba	647	99,82	272,09	272,58	46,62	4472,50	0,72
Cd	576	2,64	4,35	164,50	1,06	56,82	0,18
Co	647	11,52	19,97	173,36	5,27	227,12	0,221
Cr	647	115,38	192,01	166,41	71,13	2940,16	33,59
Cu	647	47,34	51,02	107,77	32,31	388,05	0,20
Fe	647	4,62x10 ⁴	3,40x10 ⁴	73,64	39377,91	2,11x10 ⁵	506,23
Hg	576	0,22	0,66	299,20	0,05	8,59	0,01
Mn	647	418,33	974,46	232,94	172,60	1,56x10 ⁴	0,77
Mo	647	2,47	2,23	90,33	2,36	16,91	0,46
Ni	647	29,04	42,51	146,40	18,12	591,13	0,50
Pb	576	20,19	17,68	87,58	16,70	183,33	1,30
Sb	559	1,22	4,85	399,43	0,31	81,86	0,00
Se	559	0,45	0,56	124,97	0,28	4,62	0,02
Sr	629	17,62	55,85	317,02	6,54	815,88	0,02
V	647	123,79	118,25	95,53	81,73	705,00	2,45
Zn	647	43,96	38,07	86,60	33,90	392,70	1,00

*Adimensional

** Coeficiente de Variação

Espacialização dos teores de metais e metaloides

Os teores dos elementos serão espacializados para área do Estado de Minas Gerais utilizando-se dos métodos de aprendizado de máquinas e da krigagem ordinária. O protocolo de espacialização foi aplicado sob os dados inalterados e os transformados, segundo as funções $\log(Y+1)$ e raiz quadrada de Y . As transformações objetivaram aumentar o desempenho na espacialização dos elementos nos softwares avaliados, (KIM, et al, 2016; MASSON et al., 2014; MUKHOPADHAYA S., 2016; SOVANN e POLYA, 2010).

Aprendizado de máquinas

Os métodos de aprendizado de máquinas utilizam de outros dados especializados (covariáveis), que encontram relações entre as covariáveis e os teores dos elementos nas amostras para prever teores em locais onde não existem dados quantificados e, conseqüentemente, estabelecer sua espacialização. Antes do processo de predição dos teores é necessário fazer a seleção das variáveis, com o objetivo de reduzir o número de covariáveis e simplificar o modelo utilizado.

Neste estudo foram feitas a seleção de covariáveis e, posteriormente, a predição dos elementos. Os métodos utilizados foram estabelecidos previamente no capítulo 1 desta tese.

Krigagem ordinária

Foi feita a espacialização dos teores por meio do método de krigagem ordinária em dois softwares: R e ArcGis 10.2[®]. Os protocolos executados nos dois softwares estão contemplados no capítulo 2 desta tese.

Avaliação dos métodos de espacialização dos teores dos elementos

O desempenho dos algoritmos de aprendizado de máquinas e dos modelos teóricos de semivariograma da krigagem foi comparado pelo coeficiente de determinação para o modelo (R^2) (EQ. 1) (YANG, et al., 2016) sendo calculado por:

$$R^2 = \frac{\sum_{i=1}^n (P_i - \overline{Om})^2}{\sum_{i=1}^n (Om_i - \overline{Om})^2} \quad (\text{EQ. 1})$$

Onde n é o número de observações, Om_i são os teores dos elementos observados nas transformações utilizadas (quando essas transformações forem feitas), \overline{Om} é a média dos valores observados e P_i são teores dos elementos previstos pelos métodos em cada grupo. Com os resultados dos valores foi calculada a raiz quadrada do erro quadrado médio transformado (EQ. 2) (AHMED et al., 2015, GUO et al., 2015; HALLAK e PEREIRA FILHO, 2011, MEHTATALO et al., 2007).

$$RMSE = \left[\frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2 \right]^{1/2} \quad (\text{EQ. 2})$$

Foi calculado o valor da raiz quadrada do erro quadrado médio normalizado (nRMSE) (LOPATIN et al, 2016, LOAGUE e GREEN 1991), seu valor é calculado com o valor do RMSE, dividido pela amplitude amostral, sendo seu valor dado em percentagem (EQ. 3). O nRMSE é utilizado quando se compara diferentes amostras com diferentes variâncias amostrais ou grupo de amostras com diferente amplitudes (COOK e KRISHNAN, 2015; HAN et al., 2003). A predição é considerada “excelente” quando os valores de nRMSE são inferiores a 10%, “boa” entre 10% a 20%, “regular” entre 20 a 30% e acima de 30% de nRMSE é considerada “ruim” (JAMIESON et al., 1991).

$$RMSE = \frac{RMSE}{(\max(O_i) - \min(O_i))} \quad (\text{EQ. 3})$$

Foram selecionados os menores resultados de nRMSE dos aprendizado de máquinas e dos modelos teóricos de semivariograma da krigagem avaliados é confeccionados dos mapas de predição de cada elemento analisado desde melhores resultados para os dois métodos de espacialização para cada elemento. Para os elementos que apresentaram os melhores resultados de nRMSe (nRMSE com valores próximos zero), que apresentava transfirmções (log1p e raiz quadrada), foram feitas a transformação inversa (back-transformation) para escala real dos teores dos elementos (HENGL et al., 2013). As áreas cobertas por cursos d'água, represas ou

zonas urbanas acima da área mínima mapeável não tiveram seus valores preditos apresentados.

Os parâmetros estatísticos foram desenvolvidos com o software R, utilizando-se das bibliotecas: dplyr, hydroGOF, Metrics.

RESULTADOS

Os resultados mostraram que 11 elementos tiveram resultados de nRMSE inferiores (melhores) pelo algoritmos de aprendizado de máquinas em comparação aos objetivos pelo ArcGis, sendo estes: Al, As, B, Co, Fe, Hg, Mn, Ni, Pb, Se e Zn (Tabela 2). Em geral, os melhores resultados de nRMSE variam entre 10% e 20%, obtendo resultados considerados bons, de acordo com Jamieson et al. (1991), exceto por B e Mo.

Os melhores resultados de nRMSE foram obtidos nos dados brutos, sem transformações. A transformação dos dados em log1p gerou melhores resultados de treinamento apenas para Al e Zn. Não foi possível comparar resultados de RMSE e R², visto que a transformações (log1p e sqrt) dos dados reduzem a variância amostral, o que gerou resultados não comparáveis aos dados não transformados.

Os valores de R² para estes elementos variaram de 0,32 a 0,013, e os melhores resultados de nRMSE foram referentes aos dados não transformados para os métodos krigagem e aprendizado de máquinas.

Para os elementos com os melhores resultados no método de aprendizado de máquinas, o algoritmo suporte vetor máquina radial sigma, obteve resultados inferiores de RMSE para cinco dos onze elementos (As, Ba, Hg, Mn e Ni). O algoritmo Ranger também se destacou obtendo menor nRMSE para com 3 elementos (Al, Co e Pb).

Dentre os algoritmos de treinamento selecionados, os algoritmos suporte vetor máquina radial sigma e Ranger obtiveram melhor desempenho para um número maior de elementos, que se mostrou mais adaptado para predição dos elementos quando comparados aos outros algoritmos utilizados (Tabela 3).

Tabela 2 – Resultados da validação cruzada dos elementos para krigagem e aprendizado de máquinas**

Resultados Krigagem										Resultados Aprendizado de máquinas					
Elemento	Software	Transformação	Modelo	sill	range	pepita	R ²	RMSE	nRMSE	Elemento	Algoritmo	Transformação	R ²	RMSE	nRMSE
Al	R	sem	Gau	1,06E+09	85,24	7,62E+08	0,26	35321,83	18,21	Al	Ranger	log1p	0,44	0,658	16,31
As	ArcMap	log1p	Exponential	1,632	10,631	1,174	0,387	0,92	20,49	As	SvmRadialSigma	sem	0,32	33,641	13,97
B	R	sem	Ste	33,82	98,06	16,11	0,22	5,93	22,24	B	Ranger	sem	0,22	5,821	22,92
Ba	R	sqrt	Ste	30,3	244,13	15,3	0,12	5,34	19,3	Ba	SsvmRadialSigma	sem	0,11	166,878	16,07
Cd	R	sem	Ste	16,79	8,77	1,92	0,26	3,31	14,96	Cd	SvmRadialSigma	sem	0,32	3,102	18,1
Co	ArcMap	sqrt	Spherical	5,3	8,146	1,572	0,112	1,877	18,58	Co	Ranger	sem	0,22	15,556	16,52
Cr	ArcMap	sem	Exponential	6,60E+04	1,284	2,23E+04	0,023	205,929	16,18	Cr	SvmRadialSigma	sem	0,18	126,585	17,39
Cu	R	sem	Ste	2093,21	6,98	136,55	0,09	47,54	18,07	Cu	SvmRadialSigma	sem	0,23	43,445	18,22
Fe	R	sem	Ste	6,97E+08	183,93	3,92E+08	0,23	29498,47	16,94	Fe	Ranger	sem	0,32	27560,41	16,19
Hg	R	log1p	Sph	0,01	30,97	0,04	0,01	0,23	16,69	Hg	SvmRadialSigma	sem	0,07	0,489	16,26
Mn	ArcMap	log1p	Gaussian	3,941	4,366	2,547	0,0465	1,491	25,65	Mn	SvmRadialSigma	sem	0,12	652,854	16,57
Mo	R	sem	Gau	2,66	65,41	2,34	0,19	1,99	20,88	Mo	Random florest	sem	0,25	1,881	22,67
Ni	ArcMap	sem	Exponential	2610,556	8,902	725,809	0,047	45,257	18,72	Ni	SvmRadialSigma	sem	0,25	31,9	15,66
Pb	R	sem	Ste	82,04	369,72	310,61	0,06	16,51	17,64	Pb	Ranger	sem	0,17	14,593	16,84
Sb	ArcMap	sem	Exponential	81,249	2,261	0	0,013	5,291	15,25	Sb	SvmRadialSigma	sem	0,12	2,762	15,92
Se	R	sqrt	Ste	0,13	49,2	0,01	0,3	0,3	19,16	Se	Rf	sem	0,24	0,466	17,64
Sr	ArcMap	sem	Exponential	1848,729	8,215	8,895	0,084	55,816	14,68	Sr	ExtraTrees	sqrt	0,26	2,063	15,94
V	ArcMap	sem	Spherical	8755,03	13,362	2958,289	0,36	96,829	15,04	V	Ranger	sem	0,41	89,775	16,35
Zn	R	sem	Ste	1628,3	4,15	0	0,03	37,04	20,21	Zn	ExtraTrees	log1p	0,29	0,65	17,71

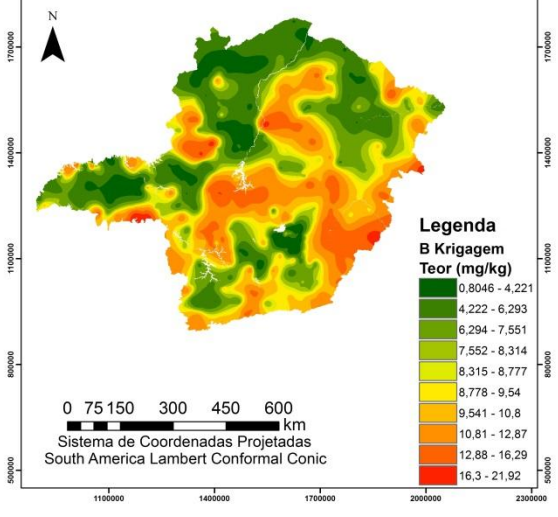
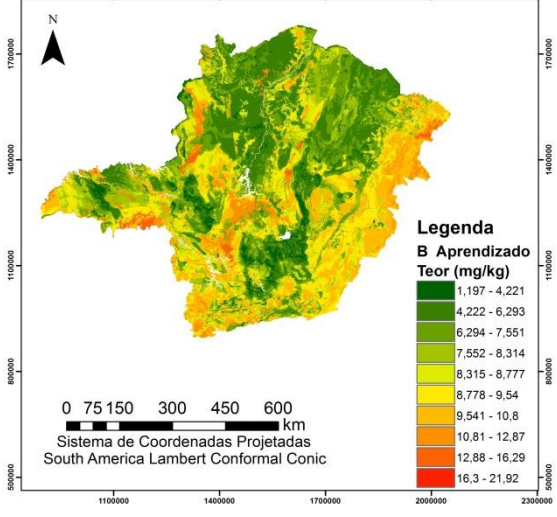
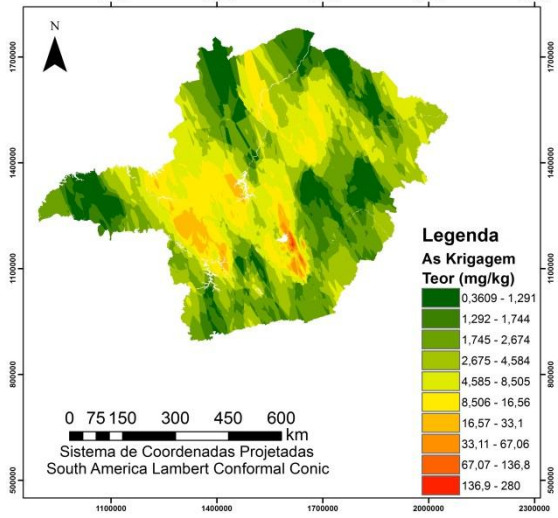
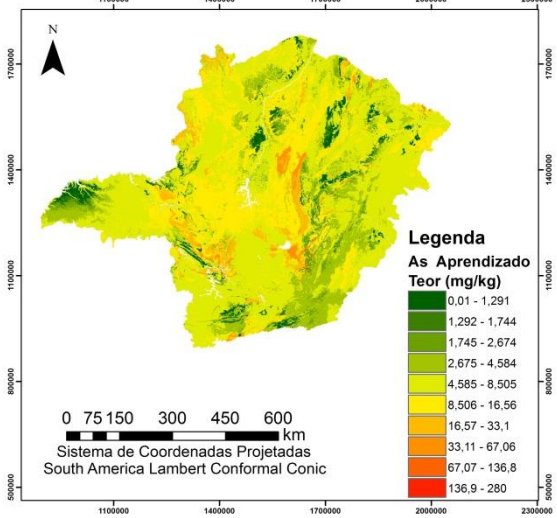
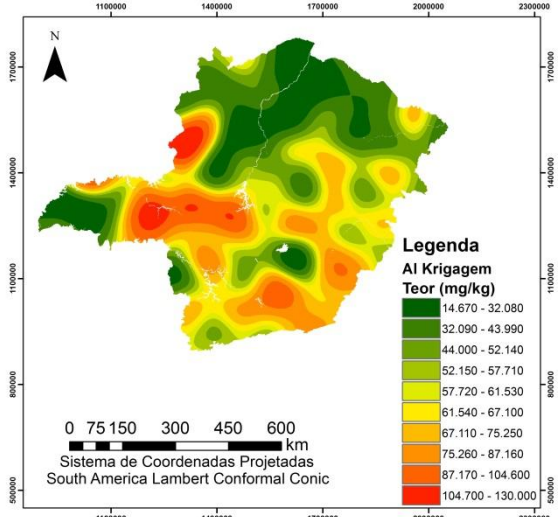
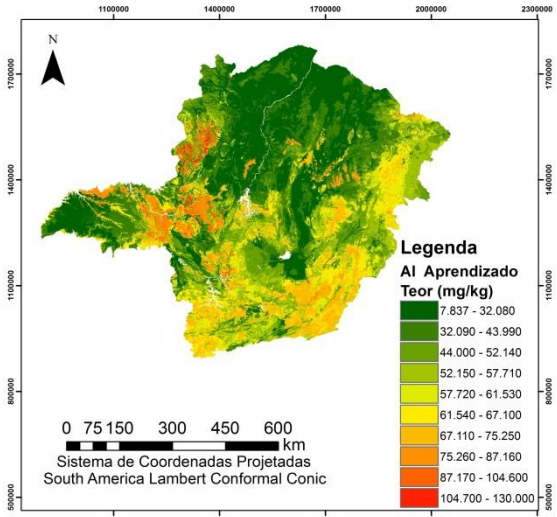
** Melhores resultados para nRMSE em cinza

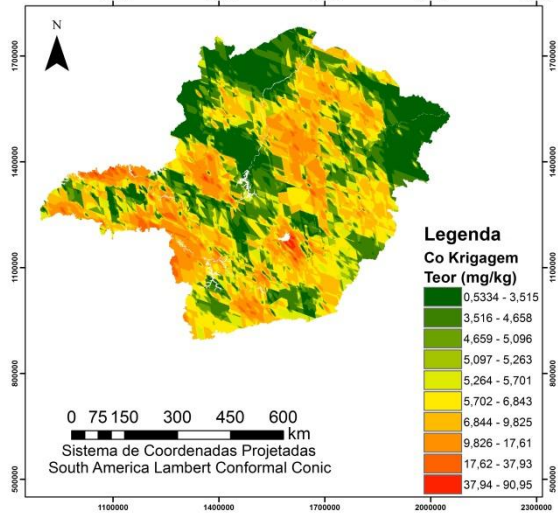
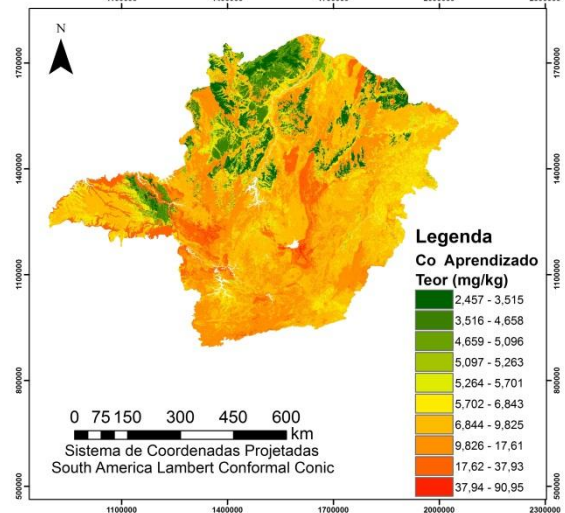
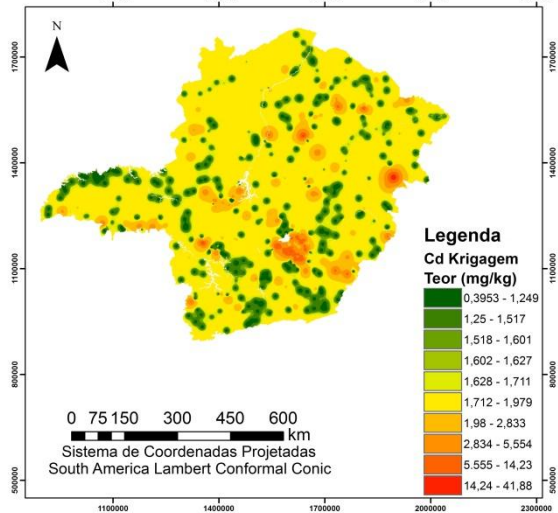
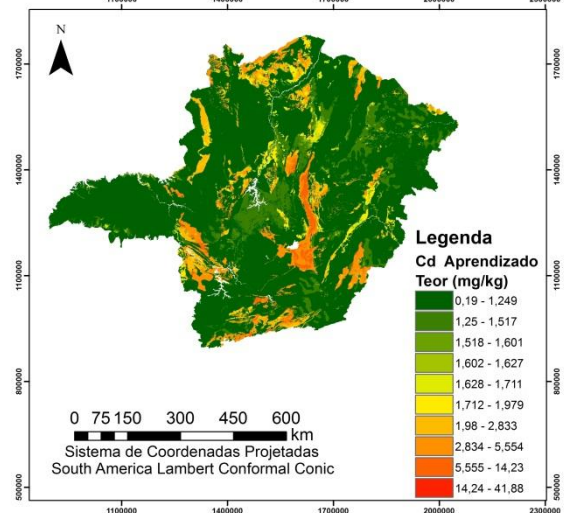
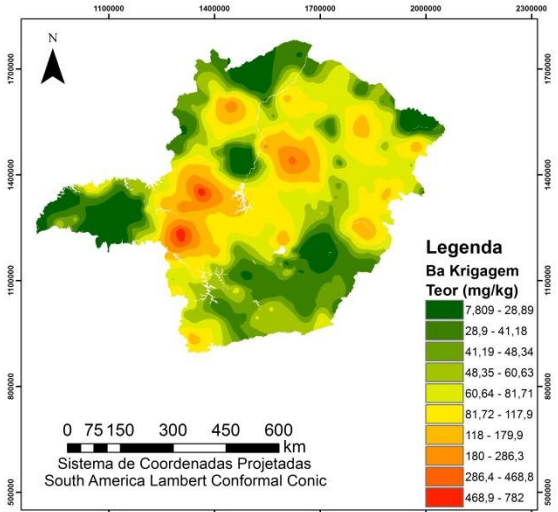
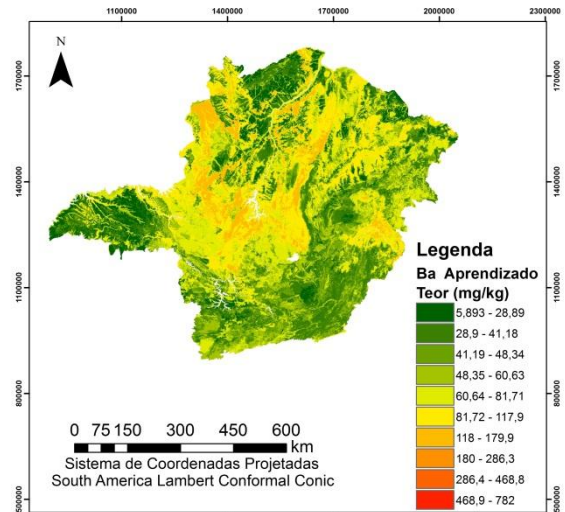
Espacialização dos elementos

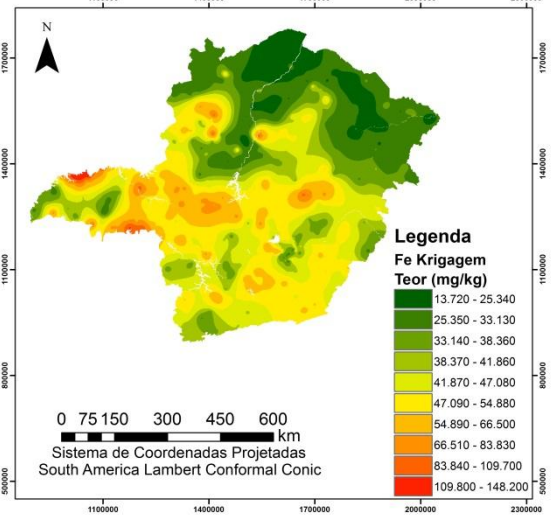
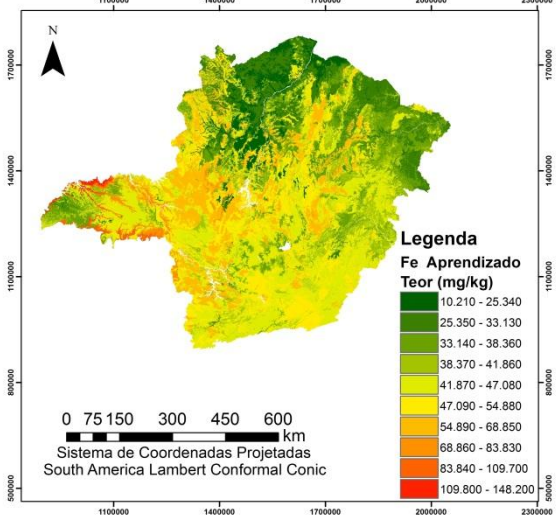
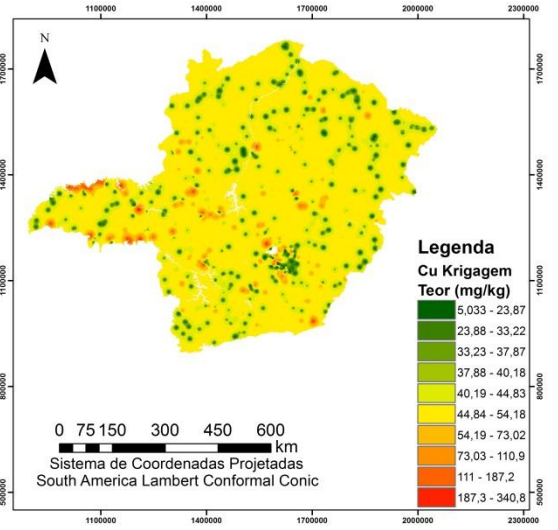
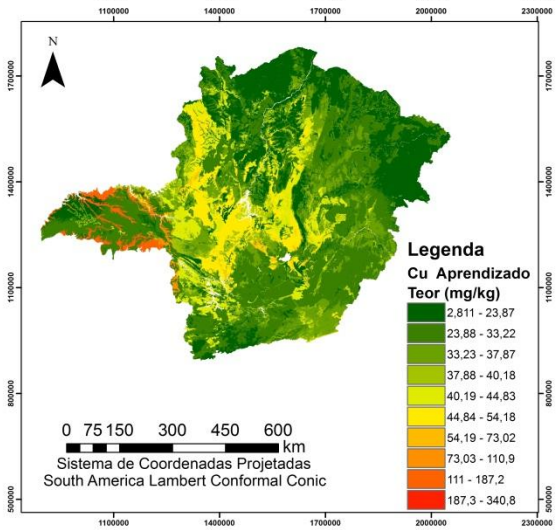
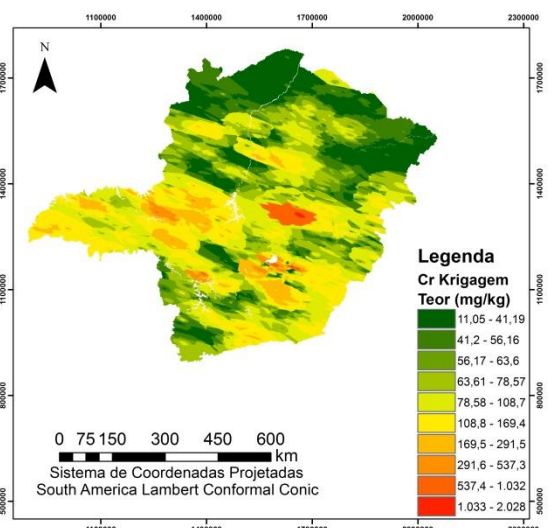
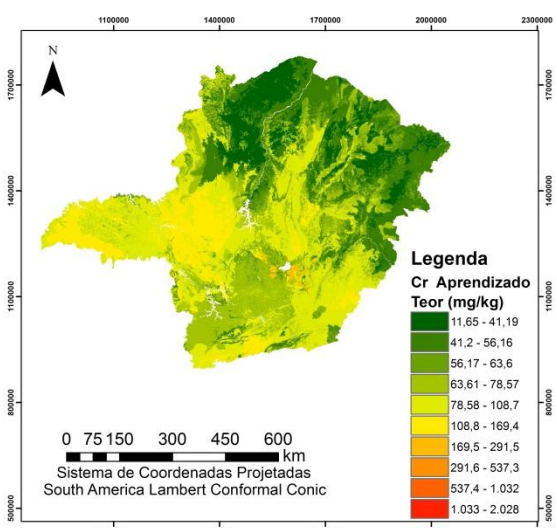
Os mapas espacializados pelos métodos de krigagem e aprendizado de máquinas apresentam similaridades entre os elementos avaliados, com estimativas dos teores próximos entre si (Figura 2). Em geral, a predição por krigagem e por algoritmos de aprendizado de máquinas apresentaram valores máximos e mínimos até 3509,00 e 5210,14 % diferentes, respectivamente. Para os elementos Fe, Mo e Se, os teores máximos e mínimos entre os métodos de espacialização ficaram próximos, e apresentaram diferenças inferiores a 40%.

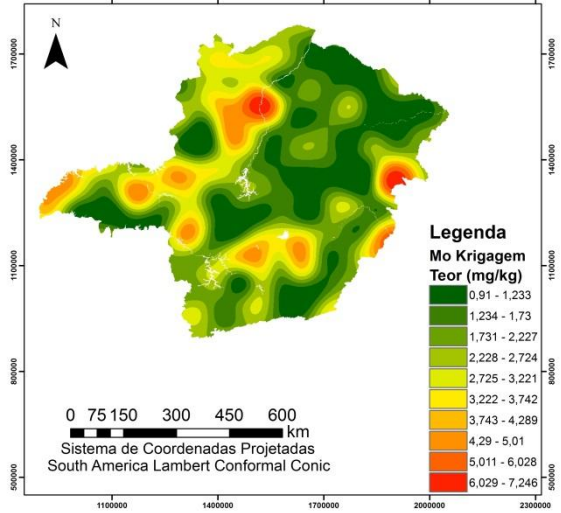
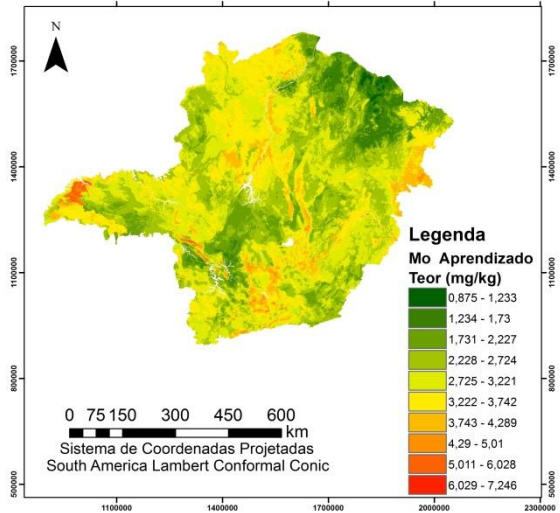
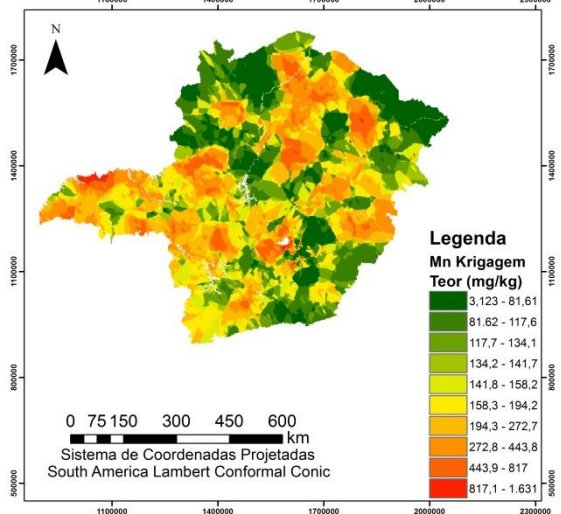
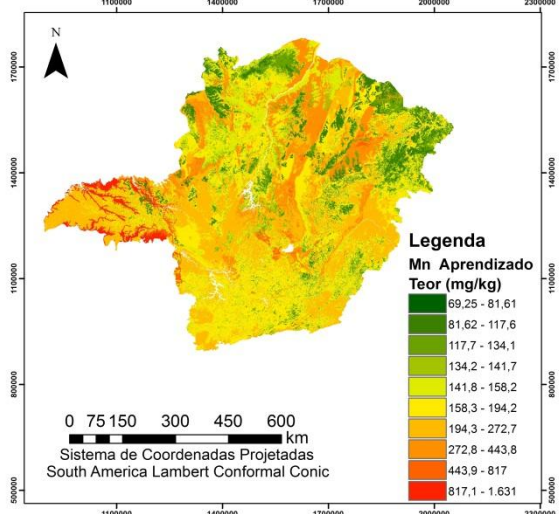
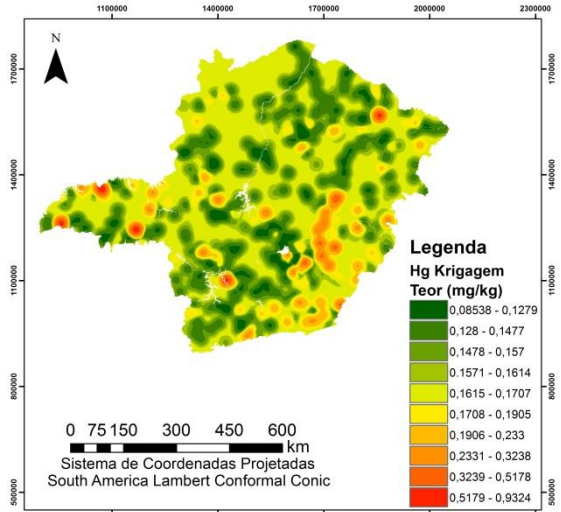
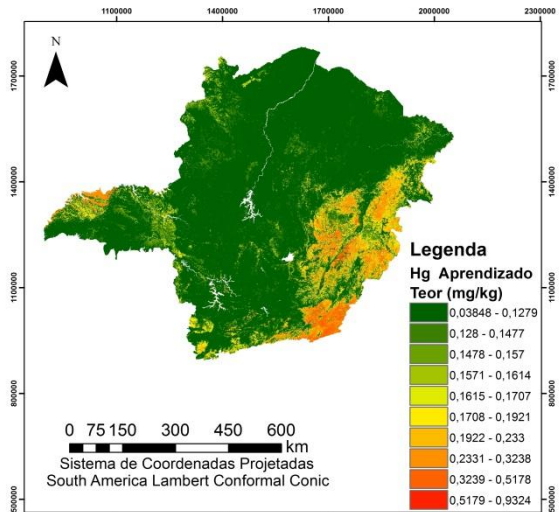
Tabela 3. Teores máximos e mínimos dos elementos para krigagem e aprendizado de máquinas**

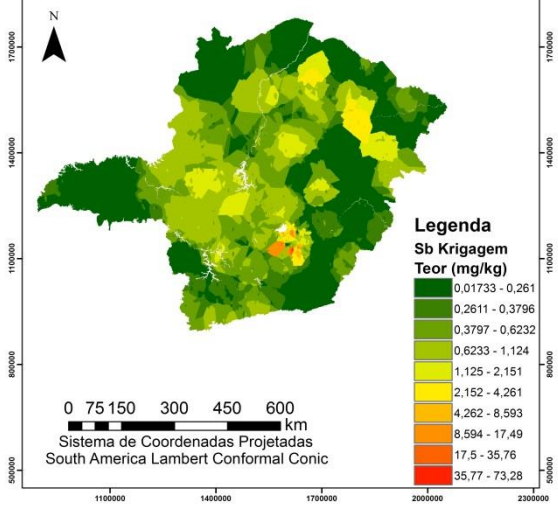
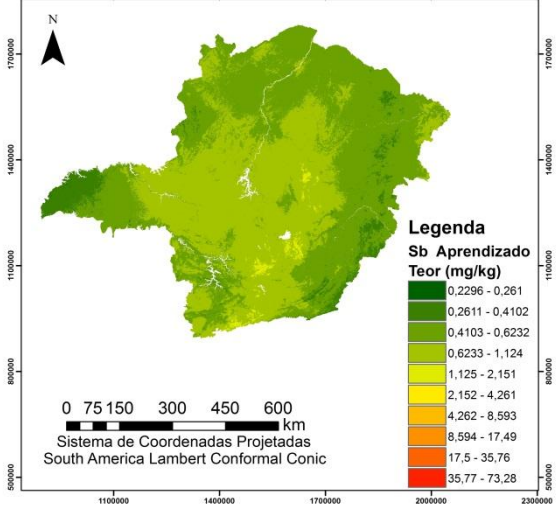
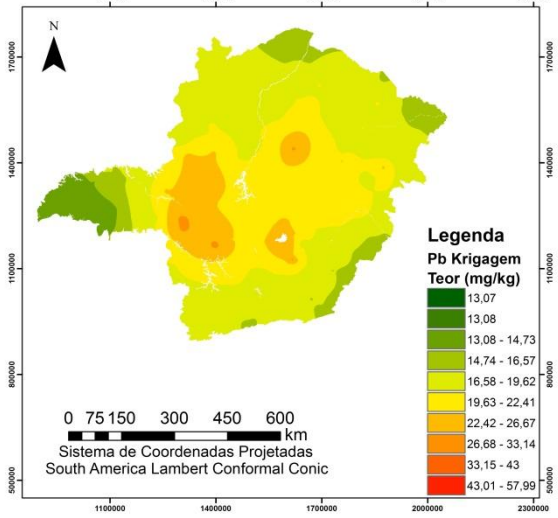
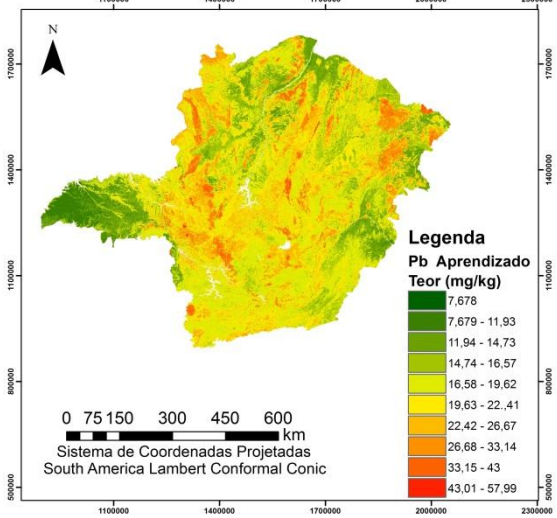
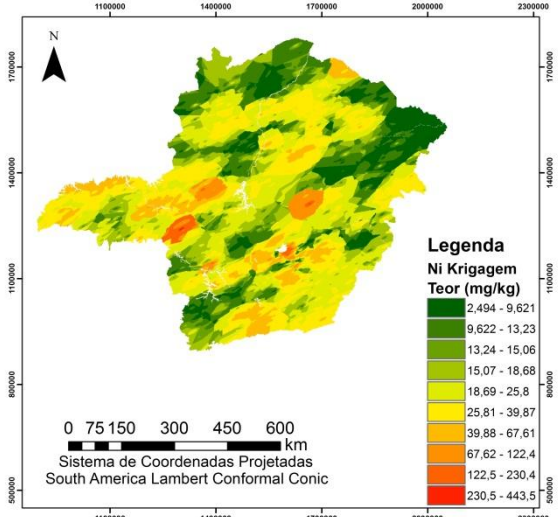
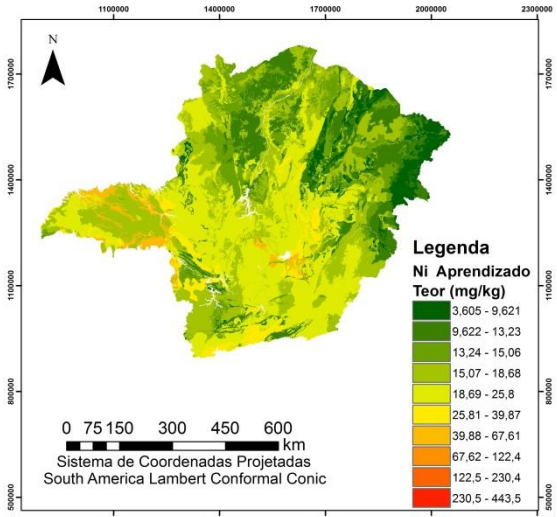
	Krigagem		Aprendizado		Diferença	
	Mínimo (mg/kg)	Máximo (mg/kg)	Mínimo (mg/kg)	Máximo (mg/kg)	Mínimo (%)	Máximo (%)
Al	1,47E+04	1,30E+05	7,84E+03	1,20E+05	87,19	8,79
As	0,36	280,00	0,01	102,90	3509,00	172,11
B	0,80	21,92	1,20	17,07	48,77	28,41
Ba	7,81	782,00	5,89	172,80	32,51	352,55
Cd	0,40	41,88	0,19	7,54	108,05	455,51
Co	0,53	90,95	2,46	67,92	360,63	33,91
Cr	11,05	2028,00	11,65	220,60	5,43	819,31
Cu	5,03	340,80	2,81	131,10	79,05	159,95
Fe	1,37E+04	1,48E+05	1,02E+04	1,37E+05	34,38	8,25
Hg	0,09	0,93	0,13	0,45	49,80	107,57
Mn	3,12	1631,00	69,52	947,90	2126,06	72,06
Mo	0,91	7,25	0,88	6,09	4,00	19,08
Ni	2,49	443,50	3,61	60,53	44,55	632,69
Pb	13,07	31,28	7,68	57,99	70,23	85,39
Sb	0,02	73,28	0,23	1,38	1224,87	5210,14
Se	0,02	3,47	0,02	2,98	5,03	16,58
Sr	0,13	760,60	1,31	226,99	910,02	235,08
V	10,12	641,60	38,80	401,17	283,40	59,93
Zn	3,90	377,10	3,88	132,30	0,49	185,03

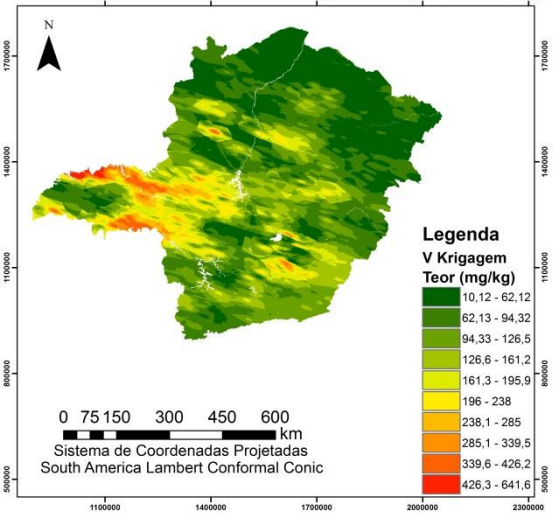
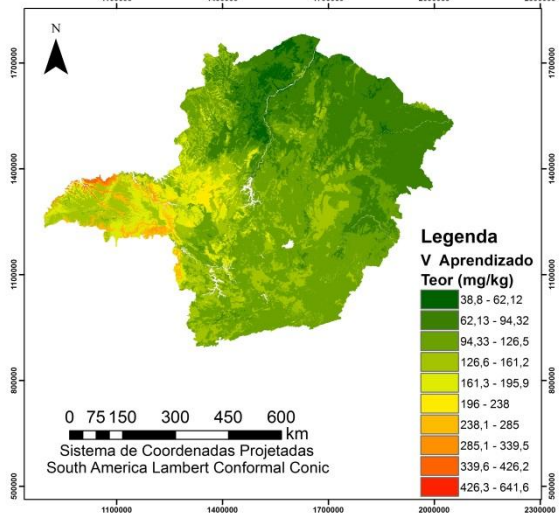
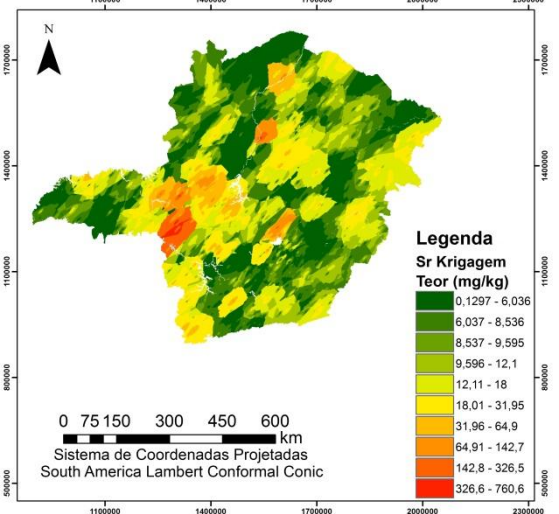
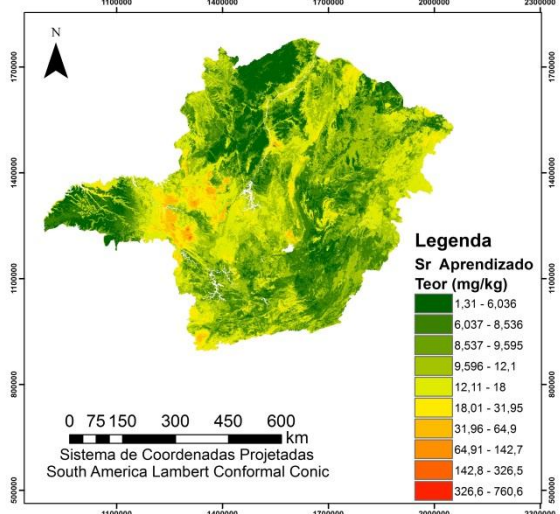
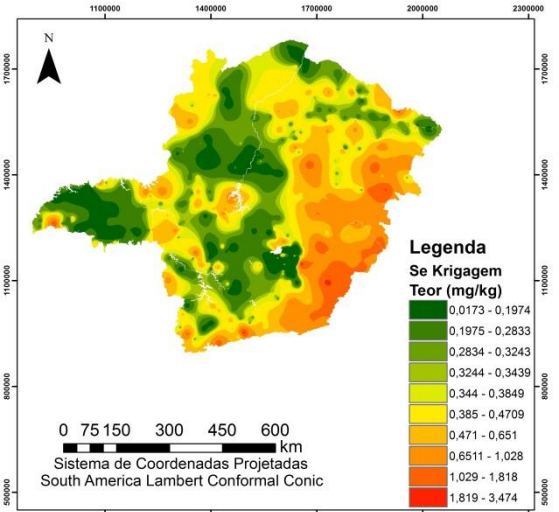
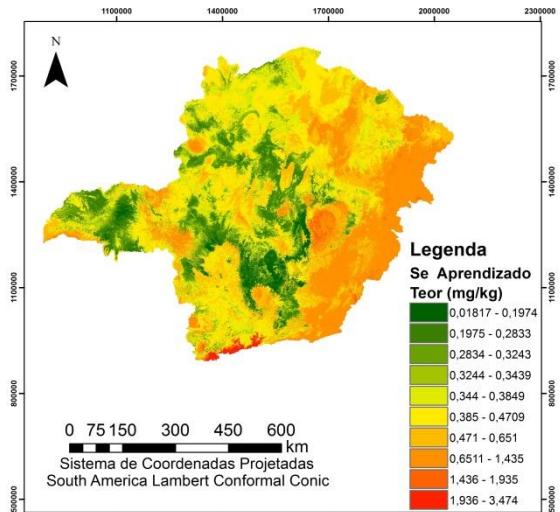












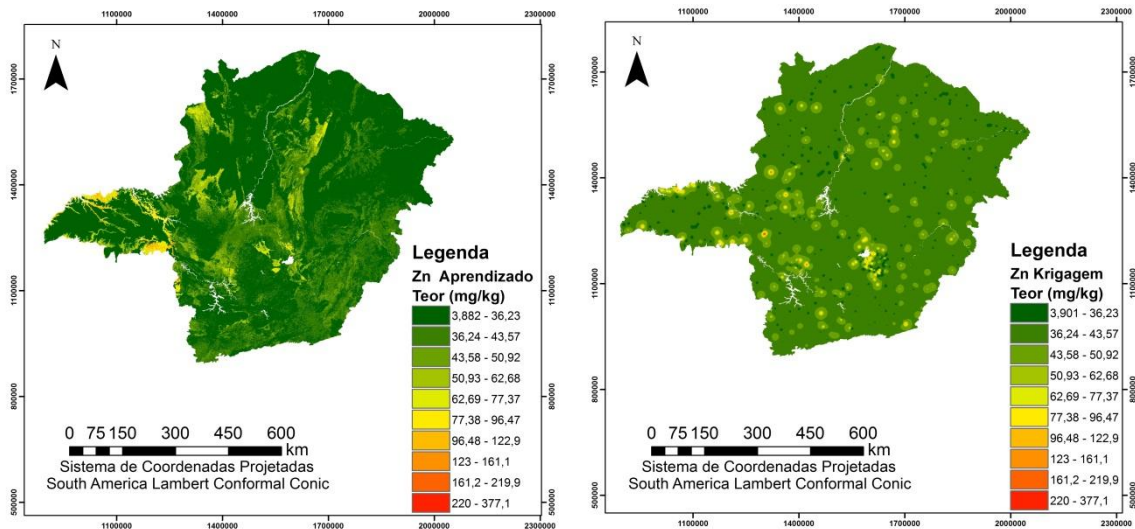


Figura 2– Espacialização dos elementos Al, As, B, Ba, Cd, Co, Cr, Cu, Fe, Hg, Mn, Mo, Ni, Pb, Sb, Sr, Se, V e Zn para os melhores resultados dos algoritmos de aprendizado de máquinas e a krigagem.

DISCUSSÃO

O baixo número de elementos químicos com bons desempenhos (menores nRMSE) pelos métodos de espacialização demonstrou a não existência um método ideal para esse tipo de espacialização. Assim, ainda há a necessidade de avaliação particular para seleção do método de acordo com a variável a ser predita. Para os elementos que tiveram desempenhos superiores no aprendizado de máquinas, estes elementos sofreram influência da litologia na predição como pode ser visto no capítulo um, com a exceção do Sb e Se. No Estado, a litologia apresenta-se de forma heterogênea e com locais onde há o encontro de rochas com altos teores de metais e metaloides e em outras locais há rochas com baixos teores, como exemplo do encontro de itabiritos com teores altos de metais e metaloides e arenitos com inferiores destes elementos (ALKMIM, et al, 2015). Essa descontinuidade litológica afetou a dependência espacial das amostras gerando resultados inferiores da krigagem em relação ao aprendizado de máquinas.

Para os elementos que obtiveram os melhores desempenhos (menores nRMSE) pela krigagem, estes resultados estariam relacionados a maior dependência espacial entre as amostras próximas. Para alguns elementos, o melhor desempenho da krigagem para os elementos Cr, Cu, Sb e Sr. porém o melhor desempenho krigagem pode não ser real, visto que o elemento obteve menores valores nRMSE pela krigagem, e com valores de R^2 inferiores aos obtidos pelo algoritmo de aprendizado de máquinas. Para estes elementos (Cr, Cu, Sb e Sr), o melhor desempenho para os

métodos de espacialização foi obtido sem transformação, portanto, quando foi possível comparar os valores de R^2 e RMSE e as variáveis estatísticas. A partir desses parâmetros, foi observado um melhor resultado para algoritmo de aprendizado de máquinas. Importante salientar que a comparação de métodos de espacialização que usam os parâmetros R^2 e RMSE, somente ser efetuado para o mesmo grupo de amostras, sendo o caso dos elementos Cr, Cu, Sb e Sr.

Quando comparamos a espacialização dos dados pelo método dois métodos de espacialização, krigagem e aprendizado de máquinas, para a maioria os mapas gerados, a krigagem gerou teores máximos superiores quando comparados com o aprendizado de máquinas, com a exceção do Pb. Isso estaria relacionado às características da krigagem, que utiliza todos os pontos amostrais dentro da amplitude amostral. Enquanto isso, os métodos de aprendizado de máquinas dão maior importância aos locais com teores onde há maior número de amostras.

Uma medida para aumentar o desempenho dos métodos de espacialização seria a divisão do Estado em regiões de maior homogeneidade. Essa divisão promoveria a redução da heterogeneidade dos elementos. A coleta de novas amostras é outra alternativa paralela a regionalização da espacialização. Todavia, o alto custo de quantificação reduz a aplicação desta alternativa e deve-se considerar a possibilidades que outras universidades, empresas ou órgãos ambientais tenham quantificado amostras com as mesmas condições utilizadas neste trabalho. É necessário uma análise exploratória desde dados para reduzir o número de novas coletas.

Os métodos de aprendizado de máquinas necessitam de covariáveis especializadas para predição dos elementos. Assim, há uma tendência de melhora na espacialização dos teores, quando menor a resolução espacial das covariáveis utilizadas. Este estudo utilizou resolução espacial de 1 km x 1 km, devido algumas covariáveis não apresentarem menor resolução com exemplo covariáveis bioclimáticas ou alto custo de aquisição de covariáveis com menor resolução espacial com covariáveis gamaespectrometria. Assim, há o convite para pesquisadores de outras áreas confeccionarem mapas de variáveis bioclimáticas com maior detalhamento para o Estado e a liberação pelos órgãos competentes de covariáveis que já tenham valores especializados com maior detalhamento sem custo para os pesquisadores, tendo em vista da importância desde tipo de pesquisa para a sociedade.

CONCLUSÃO

Os resultados mostraram que 11 elementos demonstraram melhores desempenhos pelos resultados de nRMSE na espacialização, obtidos por algoritmos de aprendizado de máquinas (Al, As, B, Co, Fe, Hg, Mn, Ni, Pb, Se e Zn). Os mapas gerados pela krigagem apresentam mapas com teores superiores aos apresentados por todos os elementos, com a exceção do Pb. Os métodos de aprendizado de máquinas geram mapas de predição com maior nível de detalhamento quando comparados com mapas gerados pela krigagem.

Este trabalho situa-se entre os primeiros estudos de comparação de algoritmos de aprendizado de máquinas e a krigagem para a predição de teores de metais e metalóides, na América Latina. Assim, os resultados mais precisos de predição podem ser alcançados com regionalização da espacialização dos teores em locais com maior homogeneidade, além da utilização de covariáveis com maior resolução espacial. Sendo convidados pesquisadores de outras áreas a confeccionar mapas de variáveis utilizadas nesse estudo e liberação pelos órgãos competentes de covariáveis espacializados para aumentar o nível de detalhamento dos mapas gerados.

REFERÊNCIAS BIBLIOGRÁFICAS

ABRAHÃO, W. A. F; MARQUES, J. J. **Manual de procedimentos analíticos para determinação de VRQ de elementos-traço em solos do Estado de Minas Gerais** . Belo Horizonte, 2013. 28 p

AHMAD, S.; KALRA, A.; STEPHEN, Haroon. Estimating soil moisture using remote sensing data: A machine learning approach. **Advances in Water Resources**, v. 33, n. 1, p. 69-80, 2010.

ALKMIM, A. R. et al. Geochemical fingerprint of siliceous, amphibolitic and magnetitic itabirite types of the region of Serra Azul–Quadrilátero Ferrífero, MG. **Rem: Revista Escola de Minas**, v. 68, n. 1, p. 37-42, 2015.

AHMED, O. S. et al. Characterizing stand-level forest canopy cover and height using landsat time series, samples of airborne LiDAR, and the random forest algorithm. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 101, p. 89-101, 2015.

BRUNGARD, C. W. et al. Machine learning for predicting soil classes in three semi-arid landscapes. **Geoderma**, v. 239, p. 68-83, 2015.

CEMPEL, M.; NIKEL, G. Nickel: a review of its sources and environmental toxicology. **Polish Journal of Environmental Studies**, v. 15, n. 3, p. 375-382, 2006

COMIG. Mapa Geológico do Estado de Minas Gerais; 2003. p. 1 [Belo Horizonte].

COMPANHIA DE TECNOLOGIA DE SANEAMENTO AMBIENTAL - CETESB. Aplicação de biossólidos de sistemas de tratamento biológico em áreas agrícolas: critérios para projeto e operação: norma P 4230. São Paulo, 1999.

COOK, D. J.; KRISHNAN, N. C. **Activity learning: discovering, recognizing, and predicting human behavior from sensor data**. John Wiley & Sons, 2015.

DE SOUZA, J. J. L. L. et al. Geochemistry and spatial variability of metal (loid) concentrations in soils of the state of Minas Gerais, Brazil. **Science of the Total Environment**, v. 505, p. 338-349, 2015.

FORKUOR, G. et al. High Resolution Mapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso: A Comparison of Machine Learning and Multiple Linear Regression Models. **PloS one**, v. 12, n. 1, p. e0170478, 2017.

GÖHRE, Vera; PASZKOWSKI, Uta. Contribution of the arbuscular mycorrhizal symbiosis to heavy metal phytoremediation. **Planta**, v. 223, n. 6, p. 1115-1122, 2006.

GUO, P et al. Digital mapping of soil organic matter for rubber plantation at regional scale: An application of random forest plus residuals kriging approach. **Geoderma**, v. 237, p. 49-59, 2015.

HALLAK, R.; PEREIRA FILHO, A. J.. Methodology for performance analysis of simulations of convective systems in the metropolitan area of São Paulo with the ARPS model: sensitivity to variations with the advection and the data assimilation schemes. **Revista Brasileira de Meteorologia**, v. 26, n. 4, p. 591-608, 2011.

HAN, K.; VIAU, A. A.; ANCTIL, F.. High-resolution forest fire weather index computations using satellite remote sensing. **Canadian journal of forest research**, v. 33, n. 6, p. 1134-1143, 2003.

HENDERSON, B. L. et al. Australia-wide predictions of soil properties using decision trees. **Geoderma**, v. 124, n. 3, p. 383-398, 2005.

HENGL, T. et al. SoilGrids250m: Global gridded soil information based on machine learning. **PloS one**, v. 12, n. 2, p. e0169748, 2017.

HENGL, T.; NIKOLIĆ, M.; MACMILLAN, R. A. Mapping efficiency and information content. **International Journal of Applied Earth Observation and Geoinformation**, v. 22, p. 127-138, 2013.

JAMIESON, P. D.; PORTER, J. R.; WILSON, D. R. A test of the computer simulation model ARCWHEAT1 on wheat crops grown in New Zealand. **Field crops research**, v. 27, n. 4, p. 337-350, 1991.

JENNY, H. Factors of soil formation: A system of quantitative pedology, 281 pp. 1941.

KIM, T. J.; BULLOCK, B. P.; WIJAYA, A.. Spatial Interpolation of Above-Ground Biomass in Labanan Concession Forest in East Kalimantan, Indonesia. **Mathematical**

KOVAČEVIĆ, M.; BAJAT, B.; GAJIĆ, B. Soil type classification and estimation of soil properties using support vector machines. **Geoderma**, v. 154, n. 3, p. 340-347, 2010.

LIU, X.; WU, J.; XU, Ji. Characterizing the risk assessment of heavy metals and sampling uncertainty analysis in paddy field by geostatistics and GIS. **Environmental Pollution**, v. 141, n. 2, p. 257-264, 2006.

LOAGUE, K.; GREEN, R. E. Statistical and graphical methods for evaluating solute transport models: overview and application. **Journal of contaminant hydrology**, v. 7, n. 1-2, p. 51-73, 1991.

LOPATIN, J. et al. Comparing generalized linear models and random forest to model vascular plant species richness using LiDAR data in a natural forest in central Chile. **Remote sensing of environment**, v. 173, p. 200-210, 2016.

MASSON, D.; FREI, C. Spatial analysis of precipitation in a high-mountain region: exploring methods with multi-scale topographic predictors and circulation types. **Hydrology and Earth System Sciences**, v. 18, n. 11, p. 4543, 2014.

MEHTATALO, L.; MALTAMO, M.; KANGAS, A.. The use of quantile trees in the prediction of the diameter distribution of a stand. **Silva Fennica**, v. 40, n. 3, p. 501, 2007.

MUKHOPADHAYA, S. Rainfall Mapping using Ordinary Kriging Technique: Case Study: Tunisia. **Journal of Basic and Applied Engineering Research**. V.3, n. 1, 2016

SALMINEN, Reijo et al. Geochemical baselines of nickel and chromium in various surficial materials in the Barents region, NW Russia and Finland. **Geostandards and Geoanalytical Research**, v. 28, n. 2, p. 333-341, 2004.

SOVANN, C.; POLYA, D. **Using GIS and Geostatistics to Develop Hazards and Risk Maps of Arsenic in Shallow Groundwaters of Cambodia**. University of Manchester, 2010.

ZHENG, Y.; CHEN, T.; HE, J. Multivariate geostatistical analysis of heavy metals in topsoils from Beijing, China. **Journal of Soils and Sediments**, v. 8, n. 1, p. 51-58, 2008.

YANG, R. et al. 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.

CAPÍTULO 5

CONCLUSÃO GERAL

Foi possível verificar a existência de elevada correlação entre os teores dos metais e metaloides com a litologia do Estado, para 17 elementos (Al, As, B, Ba, Cd, Co, Cr, Cu, Fe, Hg, Mn, Mo, Pb, Sr, V e Zn). O relevo, clima, MDE, geomorfologia e as covariáveis relacionadas à gamaespectrometria e magnetometria, mostram correlação com teores de alguns metais e metaloides específicos.

Os algoritmos SVMRadialSigma, RandomFlorest, ExtraTreee e Ranger utilizados para predição dos elementos mostraram resultados satisfatórios para alguns dos elementos analisados. Todavia, para alguns elementos como, não foi possível obter melhores predições, devido as características das características biogeoquímicas destes elementos no perfil dos solos.

Na espacialização dos elementos pelos softwares de krigagem avaliados, não foi possível selecionar um software com melhor desempenho para a krigagem de metais e metaloides em solos do Estado de Minas Gerais. Os resultados demonstraram uma dependência das características das amostras afetando a espacialização dos dados. O software R apresentou baixos valores de alcance, que geraram uma distribuição espacial concêntrica em volta das amostras com altos teores. Isso ocorreu para krigagem dos elementos Cd, Co, Cr, Hg, Ni, Se, V e Zn. A geração do dado na krigagem pelo ArcGis no formado Shapefile Feature Class foi um fator impeditivo ao seu uso, visto que, não pode ser modificado posteriormente. A modificação dos resultados desta krigagem requer a transformação dos dados para raster. Todavia, está transformação reduz a qualidade do resultado deixando o mapa com qualidade inferior ao resultado gerado pelo R

A espacialização dos teores dos metais e metaloides pelos algoritmos de aprendizado de máquinas (SVMRadialSigma, RandomFlorest, ExtraTreee e Ranger), gerou 11 elementos com desempenhos superiores à krigagem, sendo estes: Al, As, B, Co, Fe, Hg, Mn, Ni, Pb, Se e Zn. Os mapas gerados pelo método de espacialização por krigagem apresentam teores máximos superiores aos apresentados todos os elementos tirando o Pb. Os métodos de aprendizado de máquinas geram mapas com maior nível de detalhamento quando comparados com mapas gerados pela krigagem.