UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE AGRONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DO SOLO

MODELOS EMPÍRICOS E MECANÍSTICOS APLICADOS AO MAPEAMENTO DIGITAL DE ATRIBUTOS DE SOLOS

Benito Roberto Bonfatti

(Tese)

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE AGRONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DO SOLO

MODELOS EMPÍRICOS E MECANÍSTICOS APLICADOS AO MAPEAMENTO DIGITAL DE ATRIBUTOS DE SOLOS

BENITO ROBERTO BONFATTI

Bacharel em Geografia (UFSC)
Mestre em Manejo do Solo (UDESC)

Tese apresentada como
um dos requisitos à obtenção do
Grau de Doutor em Ciência do Solo

Porto Alegre (RS) Brasil
Fevereiro de 2017

CIP - Catalogação na Publicação

Bonfatti, Benito Roberto
Modelos empíricos e mecanísticos aplicados ao mapeamento digital de atributos de solos / Benito Roberto Bonfatti. -- 2017.
120 f.

Orientador: Elvio Giasson.

Tese (Doutorado) -- Universidade Federal do Rio Grande do Sul, Faculdade de Agronomia, Programa de Pós-Graduação em Ciência do Solo, Porto Alegre, BR-RS, 2017.

1. Modelos empíricos. 2. Modelos mecanísticos. 3. Mapeamento digital de solos. 4. Pedometria. I. Giasson, Elvio, orient. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

BENITO ROBERTO BONFATTI

MODELOS EMPÍRICOS E MECANÍSTICOS APLICADOS AO MAPEAMENTO DIGITAL DE ATRIBUTOS DE SOLOS

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência do Solo da Faculdade de Agronomia da Universidade Federal do Rio Grande do Sul, como requisito parcial para a obtenção do título de Doutor em Ciência do Solo.

Aprovada em 15 de fevereiro de 2017
Homologada em 07 de maio de 2017

BANCA EXAMINADORA:

Prof. Carlos Gustavo Tornquist
UFRGS

Prof. Alexandre Ten Caten
UFSC

Dr. Ivan Luiz Zilli Bacic
EPAGRI/SC

Orientador - Prof. Elvio Giasson
UFRGS

AGRADECIMENTOS

A toda minha família, pelo apoio e incentivo.

Ao meu orientador Elvio Giasson, pela orientação, suporte e confiança.

Ao professor Alfred Hartemink, pelo acompanhamento no período de doutorado sanduíche nos EUA, e à Universidade de Wisconsin-Madison por toda instrução, material e estrutura disponibilizada.

Aos professores do Departamento de Solos da UFRGS por todo ensino e orientações.

A CAPES, pela concessão da bolsa de pós-graduação.

Aos amigos do PPGCS/UFRGS, em especial ao grupo de Gênese e Classificação de Solos.

MODELOS EMPÍRICOS E MECANÍSTICOS APLICADOS AO MAPEAMENTO DIGITAL DE ATRIBUTOS DE SOLOS [1/]

Autor: Benito Roberto Bonfatti
Orientador: Elvio Giasson

RESUMO

O mapeamento digital tem se tornado uma das mais importantes ferramentas na predição e mapeamento de solos. Apesar de sua importância, é ainda pouco difundido no Brasil, principalmente na predição e mapeamento de atributos de solo. O objetivo desta tese foi apresentar e avaliar diferentes modelos que podem ser utilizados no mapeamento digital de atributos de solo. Primeiramente foram discutidos e analisados diferentes modelos empíricos e, em sequência, também foram avaliados modelos mecanísticos. Dois estudos foram apresentados, um envolvendo um modelo empírico para predição e mapeamento de concentração e estoque de carbono no solo e outro utilizando modelos mecanísticos para predição de profundidade do solo e sua alteração com o tempo, em diferentes posições da paisagem. Os estudos foram aplicados no Vale dos Vinhedos, RS. Ambos modelos apresentaram validação satisfatória e capacidade de mapear atributos de solos. O modelo empírico apresentou maior dependência em relação aos dados de campo e seus resultados variaram de acordo com o método escolhido e o número e representatividade amostral. O modelo mecanístico se mostrou complexo e importante para identificar tendências de distribuição do atributo mapeado (profundidade do solo), apesar da impossibilidade de modelar todos os fenômenos envolvidos durante a pedogênese. Também apresentou menor dependência das condições amostrais e condições para melhor compreensão do comportamento dos elementos envolvidos durante os fenômenos naturais de pedogênese. Ambos modelos podem ser utilizados no mapeamento digital de solos, considerando as suas vantagens e respeitando as limitações de cada técnica utilizada.

[1/] Tese de Doutorado em Ciência do Solo. Programa de Pós-Graduação em Ciência do Solo, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul. Porto Alegre. (110 p.) Fevereiro, 2017. Trabalho realizado com apoio financeiro da CAPES.

EMPIRICAL AND MECHANISTIC MODELS APPLIED TO DIGITAL MAPPING OF SOIL ATTRIBUTES[1/]

Author: Benito Roberto Bonfatti
Adviser: Prof. Elvio Giasson

ABSTRACT

The digital mapping has become one of the most important tools on soil predicting and mapping. Although the importance, it is still a poorly disseminated methodology in Brazil, mainly when applied in soil attributes prediction and mapping. This thesis aimed to present and evaluate different models that can be used in digital mapping of soil attributes. Firstly, different techniques from empirical models to predict and map were discussed. In sequence, techniques from mechanistic models were also evaluated. Two studies were presented. The first study involved an empirical model to predict and map soil organic carbon content and stocks. The second used a mechanistic model to predict soil thickness and its variation over time in different landscape positions. The studies were conducted in Vale dos Vinhedos, RS, Brazil. Both models performance were considered satisfactory and able to map soil attributes. The empirical models depended from soil samples and results varied conform the method chosen, the soil samples number and representativity. The mechanistic models showed complexity and it was important to identify soil thickness tendencies, despite the impossibility to model all the phenomena involved during the pedogenesis. It was less dependent from soil samples and allowed a better understanding about the elements behavior involved. Both models can be used in digital mapping of soil attributes, considering their advantages and respecting each technique limitations.

[1/] D.Sc. Thesis in Soil Science. Programa de Pós-Graduação em Ciência do Solo, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul. Porto Alegre. (110 p.) February, 2017. Research supported by CAPES.

SUMÁRIO

RELAÇÃO DE TABELAS

RELAÇÃO DE FIGURAS

1. INTRODUÇÃO GERAL

O mapeamento digital de solos tem se tornado uma das mais importantes ferramentas na predição de classes e atributos ou propriedades de solos. O uso de técnicas aprimoradas de processamento de dados, junto a utilização de imagens com informações sobre a superfície do terreno (como os modelos digitais de elevação e imagens espectrais) tem facilitado e difundido o seu uso, trazendo uma importante ferramenta para os pedólogos na compreensão dos processos pedológicos e seu mapeamento. Apesar de sua importância, estudos no Brasil ainda são escassos, principalmente os relacionados à predição e mapeamento de propriedades de solo, o que denota a necessidade de um entendimento mais aprofundado de suas técnicas e funcionalidades.

As técnicas envolvidas no mapeamento pedológico com o uso de meios digitais estão em constante evolução e podem-se distinguir dois ramos em relação ao seu objetivo: o mapeamento digital de classes de solos, com vistas à predição de unidades discretas, e o mapeamento digital de atributos de solos, cujas estimativas aparecem como dados contínuos. No Brasil, o uso do mapeamento digital de classes de solos antecede o de atributos e atualmente apresenta um maior número de estudos. Este trabalho se dedicou ao mapeamento digital de atributos de solos, um ramo de pesquisa ainda pouco difundido em território nacional.

O objetivo geral do estudo foi descrever e avaliar diferentes métodos de mapeamento digital de atributos de solos, a partir de modelos empíricos e mecanísticos, aplicando as metodologias em uma microbacia no Rio Grande do Sul. O estudo se justifica devido à escassez de trabalhos no Brasil utilizando essa linha de pesquisa e à possibilidade de trazer informações úteis a continuidade dos estudos em diferentes regiões. A metodologia pode ser

aplicada em diversas áreas e a difusão das técnicas permite uma avaliação das reais facilidades e dos obstáculos que ocasionalmente possam ser encontrados quando utilizada sob diferentes condições.

Os objetivos específicos consistiram em: i) avaliar diferentes técnicas estatísticas utilizadas em modelos empíricos para predição de carbono orgânico em diferentes profundidades do solo, ii) propor método de predição de incerteza em modelo empírico e iii) avaliar a possibilidade de uso de modelo mecanístico para a predição de profundidade de solo.

No primeiro capítulo, é apresentada uma revisão bibliográfica dos dois principais modelos de estimativas de atributos de solos. Modelo empíricos são apresentados primeiramente, descrevendo suas principais metodologias e estudos realizados. Em seguida, o modelo mecanístico é apresentado, destacando-se os principais processos pedológicos e condições determinantes na contrução do modelo.

No segundo capítulo, é apresentado um estudo do mapeamento digital de concentração e estoque de carbono orgânico no solo (COS), utilizando-se de um modelo empírico. A variação espacial de COS no solo é analisada e quantificada em diferentes profundidades, com o uso de métodos de mapeamento digital. Quatro métodos de predição são testados e comparados. Os valores estimados são utilizados no cálculo e na elaboração de mapas de estoque de carbono orgânico no solo. É desenvolvido e avaliado um método para estimar a incerteza das predições.

No terceiro capítulo, modelos mecanísticos são utilizados para a predição de profundidade do solo. Foram propostos modelos que utilizam a interação entre funções de produção de solo e equações de evolução da paisagem. Um primeiro modelo foi proposto para estimar a profundidade atual do solo. Um segundo modelo foi elaborado visando estudar a evolução da espessura do solo no decorrer de 100 mil anos, em diferentes posições da paisagem. Foram testados e avaliados diferentes cenários, variando-se as funções de produção de solo de acordo com o índice de umidade topográfica.

2. CAPÍTULO I – REVISÃO BIBLIOGRÁFICA

The digital mapping of soil properties attempts to offer information about the spatial distribution of soil properties over an area, as soil thickness, organic carbon, nutrients, etc. The technique allows to predict, by interpolation methods, values of soil properties to an entire area, using data obtained by soil survey in specific points on landscape. The digital soil mapping uses the Jenny (1941) study as an important base knowledge, comprising the soil formation based on five attributes: climate (cl), organisms (o), relief (r), parent material (p) and time (t), what is summarized by "*clorpt*" model of soil formation. This model has been extended to a soil predicting model "*scorpan*" (McBratney et al, 2003), which added information about soil (*s*), when some soil properties can be predicted by its class or other soil properties, or about sample spatial position (*n)*.

The digital map of soil properties can be elaborated using different models. The most common approaches are empirical models, which use statistical techniques to find relations between data from soil survey. Soil properties data obtained from soil samples are correlated with environmental covariates and the correlation allows to extrapolate the predictions to a whole study area. Less common approaches are mechanistic models, which prefer to explain why and how the phenomenon occurs, rather than use statistical correlations. The mechanistic models use several equations to determine the physical process involved in each system component. More complex methods are used, but with the advantage of being less dependent of soil samples and, additionally, stimulating knowledge about pedogenesis, trying to explain how the nature works and to discover the elements and processes involved.

The prediction accuracy varies depending on the technology involved in each method applied. Make this information available might help to compare different methods and to evaluate the results. This chapter aims to give a brief

explanation of some empirical and mechanistic methods available to estimate soil properties to a study area. The method descriptions were based in a compilation of studies from several authors.

2.1 Empirical models

The empirical models, when used in digital soil mapping, consider that soil properties can be predicted by different regression methods. The techniques involve knowledge of different sciences, as statistics and geostatistics. Soil properties sampled by pedons are correlated with terrain covariates derived from environmental data (digital elevation model, spectral images, geology maps, landform maps, etc.). The correlation is used with the appropriated technique to estimate the soil properties where they are unknown. Samples needs to be representative of the study area, and enough to produce significant relations between soil properties and terrain covariates. The Digital Elevation Model (DEM) resolution needs to be consistent with the map scale, and as error-free as possible. GIS (Geographic Information Systems) software are used to derivate the covariates from the DEM or other images or maps. The strength of the relationship between the covariates and the soil properties will reflect on the predictions uncertainty.

2.1.1 Multiple regression analysis

Multiple regression analysis (MLR) is a general statistical technique used to analyze the relationship between a single dependent variable and several independent variables (Hair et al., 2014), represented by: $Y_1 = X_1 + X_2 + \cdots X_n$, where Y and X are metric variables.

In digital soil mapping, the MLR is used to predict the soil properties based on several covariates derived from a DEM, landforms, soil maps, geology maps and others. Discrete variables, as units in soil maps, might be codified on numeric variables to allow the regression analysis.

The methodology consists of extracting statistical parameters by correlating the environmental attributes with the representative samples. The regression produces a multivariate equation with coefficients for each

independent variable. The equation can be applied to estimate the dependent variable, e.g. soil organic carbon, extrapolating the results to the whole research area, allowing to produce a soil property map.

An approach commonly used is the stepwise MLR. In this method, the model includes the variables strongly correlated and does not include the variables with weak correlation. The regression is elaborated step by step, including the most important variables. In the first step, the independent variable (environmental covariate) with the highest correlation with the dependent variable (soil property) will be selected. The other variables are inserted at each step, if it is significant to explain the variance (Hair et al., 2014).

Models using a MLR need data with distribution close to normality, as it is an underlying assumption of the linear models. However, an alternative is a Generalized Linear Model (GLM), that can deal with non-normal distributions (Lane, 2002; McBratney et al., 2003), modifying the model rather than transforming the data to a normality condition.

Among the studies applying MLR in soil science, García et al. (2016) used a multiple linear regression to estimate the soil moisture in a flat area in Argentina, using images from radar and meteorological data as input variables. This study showed the simplicity of elaborating a model applying hydrological principles to SAR images, obtaining acceptable results ($R2 = 0.6$). Lentzsch et al. (2005) compared multiple regression and neural network to assess soil microbial biomass in sites located in Brandenburg, Germany. He concluded that both model are promising tools for soil microbial biomass prediction at landscape scale but the models are specific for the respective region. Using models from other regions or applying the elaborated model in other area, it can lead to significant different validation. Nunes et al. (2012) used a stepwise multiple regression to estimate the soil nitrate concentration, in Southern Portugal. The model improved when data were used grouped by clusters, mainly where water and temperature were not limiting factors for nitrification. A poorest performance was observed in soils under canopy areas. Qiu et al. (2010) predict soil moisture in a plateau China, using multiple-linear regressions. The results showed better adjust when using a generalized multiple linear regression and a stepwise method was most effective, when using few variables. Zornoza et al. (2007) evaluated soil quality using multiple linear regression, based on different soil physical, chemical and

biochemical properties in forest sites, Spain. The multiple linear regression showed to be a good tool, due the predictions reflect the balance among soil properties.

2.1.2 Random Forest

Random Forest uses classification trees to determine the relation between independent and dependent variables. A classification tree comprises leaf nodes, corresponding to an output variable, and branches, which have a set of rules. Ensemble methods work with many classifiers and aggregate their results, tending to be more accurate than base classifiers (Han et al., 2012). Random Forest is an ensemble method and grows many classification trees, with the growth of each tree often governed by random vectors. Breiman (2001) defines random forest as a classifier consisting of a collection of tree-structured classifiers $\{h(\boldsymbol{x}, \Theta_k), k = 1, \ldots\}$ where the $h(\boldsymbol{x}, \Theta_k)$ is a classifier, Θ_k are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input $\boldsymbol{x}$.

In random forest, according to Breiman (2001), bagging is used in tandem with random feature selection. Bagging uses a bootstrap method, in which the training subsets are randomly selected from the training data (Han et al., 2012). At each time of selection, the data may be selected again and re-added to other training subsets. In that random selection with replacement, the same data may occur more than once and some data may not be included in the training subsets. Then, a classifier model is built over each training subset, reffered as bagging. When predicting, each classifier return its prediction, which counts as one vote. The response is chosen based on the most voted among the classifiers. When applying to continuous variables, bagging takes the average value of each prediction (Han et al., 2012). In random forest, an additional layer of randomness is added to bagging (Liaw & Wiener, 2002). This randomness acts during the split of the tree nodes. Attributes are randomly choosen to split each node, differently from classification trees, which used all the input attributes.

In Random Forest, trees are grown without pruning (Breiman, 2001). The data portion used as training subset is known as the “inbag” data, whereas the rest is called the “out-of-bag” data (Ließ et al., 2012). The error and

uncertainty depends primarily on the correlation between any two trees in the forest and the strength of the individual classifiers. If the correlation is high, the error rate will increase. In case of large strength of each individual tree, the forest error rate will decrease. The output presents better validation when the forest maintain the strength of individual classifiers without increasing their correlation (Breiman, 2001; Han et al., 2012).

Two input methods can be distinguished when elaborating a Random Forest. The Forest-RI has a random input selection, where attributes are randomly selected as candidates for split at the node. The Forest-RC uses random linear combination of the input variables. Instead of randomly selecting a subset of the attributes, it creates new attributes (or features) that are a linear combination of the existing attributes (Han et al., 2012).

The Random Forest is considered relatively resistant to overfitting, with relative robustness with respect to noise features, deals with categorical and continuous predictors and it does not require data standardization of normalization, since it is relatively insensitive to value range (Breiman, 2001; Liaw & Wiener, 2002; Gambill et al., 2016). Despite giving measures for variable importance, the disadvantage is related to the limited interpretability, due the relationship between the predictors and the responses cannot be examined individually for each tree in the forest (Grimm et al., 2008; Chagas et al., 2016). Also, the model is usually only effective within the range in covariate values exhibited by the training data (Hengl et al., 2015).

When using the package *randomForest* in R (R Core Team, 2014), three parameters must be defined: the number of trees in the forest "n_{tree}"(the standard number is 500), the minimum amount of data per terminal node "nodesize" (the standard is five for each terminal node) and the number of variables used per tree "m_{try}" (the standard is one third of the total number of predictor variables) (Liaw & Wiener, 2002; Chagas et al., 2016). The size of the subset of variables used to grow each tree (m_{try}) is a sensitive parameter, defining the strength of each individual tree in the forest and the correlation between any two trees in the forest. Increasing the size, the strength of each individual tree is increasing, however the correlation between trees increases also. Tree strength improves model performance, while correlation amongst trees weakens it (Ließ et al., 2012).

Chagas et al. (2016) estimated soil surface texture in a semiarid region in Brazil using random forest and multiple linear regression. Products derived from remote sensing were used as environmental covariates. Random forest showed the best estimations for sand and clay and multiple linear regression showed better performance to predict silt. Grimm et al. (2008) used random forest to predict the spatial distribution of soil organic carbon (SOC) by 4 depths, in an island, Panama. The random forest presented good performance, showing no sensitivity to overfitting or noise features. Hengl et al. (2015) mapped soil properties of Africa. Random forest and linear regression were compared as prediction methods. The random forest performs better than linear regression. To increase accuracy, the random forest model needed quality-controlled point as input data and took considerably more computation time. Ließ et al. (2012) provided a spatial prediction of soil texture in the southern Ecuadorian Andes, using random forest and regression tree, and the random forest showed the best performance. Wiesmeier et al. (2011) found good results using random forest to spatially predict stocks of soil organic carbon, total carbon, total nitrogen and total sulphur, in a semiarid catchment in Northern China.

2.1.3 Model Trees

Model Trees has been developed as the M5 model and its extension Cubist. In a Model Tree, a tree is grown where the output leaves contain linear regression models. Also, there are intermediate linear models at each step of the tree. A prediction is made using the linear regression model at the leaves, considering the prediction from the linear model in the previous node (Quinlan, 1992). The tree comprises a set of rules, which initially are paths from the top of the tree to the bottom. Rules are eliminated via pruning and/or combined for simplification (Kuhn et al., 2012).

Model Trees predict continuous value rather than discrete class (Quinlan, 1992; Minasny & Mcbratney, 2008). In a similar method called regression tree, the predictions have discrete values as output. Model Trees, instead, presents a linear regression model at the leaves. Different linear models are able to capture local linearity in the predictor variable space, this leading to smaller trees and better prediction accuracy when compared with regression

trees (Quinlan, 1993; Akpa et al., 2016). Model trees have advantages over regression trees in both compactness and prediction accuracy, attributable to the ability of model trees to exploit local linearity in the data (Quinlan, 1992).

The process starts using a linear least-squares regression and has a set of rules when it ends, each rule associated to a multivariate linear model (Quinlan, 1992; Minasny & Mcbratney, 2008). The linear regression model at each terminal node is used to predictions. Cubist model also provides the relative importance of each variable in the model (Kuhn et al., 2012; Peng et al., 2015). The Cubist model can also use *committees* where iterative model trees are created in sequence. The first tree follows the main procedure. Subsequent trees are created using adjusted versions to the training set outcome. (Kuhn et al., 2012).

Akpa et al. (2016) estimated the SOC concentration for Nigeria using random forest, cubist and boosted regression tree (BRT). Random forest and Cubist exhibited similar and better performance than BRT. Minasny & McBratney (2008) compared Cubist, partial least-square regression and regression tree to predict soil properties from diffuse infrared reflectance spectra. The authors concluded that Cubist provided greater accuracy, with simpler and more parsimonious performance. The Cubist model still produces comprehensible equations, an optimal variable selection, and respects the upper and lower limits of the data. Similar studies, using Cubist to model soil properties from spectral data, were carried out by Peng et al. (2015) and Viscarra Rossel et al. (2016) to estimate soil organic carbon.

2.1.4 Neural networks

Artificial neural networks (ANNs) are a standard technique in the range of artificial intelligence and data mining, designed to learn rules from examples (Zell, 1994; Behrens et al., 2005). It consists of connected processing elements, called neurons, each having weighted inputs, transfer functions and output (Agatonovic-Kustrin & Beresford, 2000; Schmidhuber, 2015). An ANN is formed by hundreds of single units and a simple processing units communicate by sending signals to each other over a large number of weighted connections (Agatonovic-Kustrin & Beresford, 2000; Kröse & Van Der Smagt, 1996).

ANN is able to learn and generalize from experimental data. This ability allows this computational system to learn constitutive relationships of materials directly from the result of experiments (Tiwari et al., 2015). An ANN first develops a memory associating input patterns with outputs through training on examples and then can produce an output when given an input pattern (Zhu, 2000).

Each unit of an ANN receives inputs from neighbors or external sources and uses this to compute an output signal which is propagated to other units. The system is inherently parallel in the sense that many units can carry out their computations at the same time (Kröse & Van Der Smagt, 1996). The inputs multiplied by the connection weights (adjusted) are first summed (combined) and then passed through a transfer function to produce the output for that neuron. The activation function is the weighted sum of the neuron's input (Agatonovic-Kustrin & Beresford, 2000).

A neural network is trained to map a set of input data by iterative adjustment of the weights. The ANN reads the input and output values in the training data set and changes the value of the weighted links to reduce the difference between the predicted and target values, using the backward propagation of the error (Agatonovic-Kustrin & Beresford, 2000). The error in prediction is minimized across many training cycles until network reaches specified level of accuracy. If a network is left to train for too long, however, it will overtrain and will lose the ability to generalize (Agatonovic-Kustrin & Beresford, 2000). When the system reaches an ideal condition, the weighted links between simple units are saved and then can be used to make predictions using a new input data.

Agatonovic-Kustrin & Beresford (2000) reported significant processes involving ANN. Regarding the inputs, artificial neurons can receive two types, either excitatory or inhibitory. Excitatory inputs cause the summing mechanism of the next neuron to add while the inhibitory inputs cause it to subtract. Feedback is another type of connection where the output of one layer routes back to the input of a previous layer, or to same layer. Different types of architecture may be identified according to the absence or presence of feedback connection in a network (Agatonovic-Kustrin & Beresford, 2000).

There are two approaches to training the ANN, the supervised and the unsupervised. Supervised learning depends on the inputs and outputs of the training sets. In unsupervised training, the network is provided with inputs but not with desired outputs. The system itself must then decide what features it will use to group the input data. The most often used ANN is a fully connected, supervised network with backpropagation learning rule (Agatonovic-Kustrin & Beresford, 2000). The package *neuralnet* is often used in R to develop ANN's (Günther & Fritsch, 2010).

Aitkenhead and Coull (2016) used a neural network model to map soil organic matter, bulk density and soil organic carbon content at different soil depths across Scotland. Tiwari et al. (2015) used soil spectra data and ANN to predict soil organic carbon. The authors conclude that ANN model was a potential tool in predicting SOC distribution in agricultural field using hyperspectral remote sensing data at different scales. Zhu (2000) applied neural network for providing information on the detailed spatial variation of soil properties for hydroecological modeling, in a watershed in western Montana, US. The approach revealed greater spatial detail and had higher quality than that derived from the conventional soil map.

2.1.5 Geostatistics

Geostatistics consists in a set of statistical methods considering the spatial dependence between georeferenced databases, used to make predictions and data interpolation. It is a subset of statistics specialized in analysis and interpretation of geographically referenced data (Goovaerts, 1997; Hengl, 2009). A data spatial continuity is considered, which is not seen in the most of regression techniques. The basic assumption is that two data close to each other are more likely to have similar values than two data that are far apart (Isaaks & Srivastava, 1989).

Geostatistics considers $Z_{(s)}$ as random variables that assumes different values Z depending on position in a region S. The set of variables $Z_{(s)}$ in the whole area S is considered as a random function (Isaaks & Srivastava, 1989). A random function $Z_{(s)}$ is said to be first-order stationary if its expected value is the same at all locations throughout the study region, corresponding to the mean.

In that hypothesis, the distribution of a random function is constant to the whole area (Trangmar et al., 1985):

$$E\left[Z_{(s)}\right] = E\left[Z_{(s+h)}\right] = m$$

Where h is the vector of separation between sample locations and m is the mean of the samples.

The second-order stationary applies if the spatial covariance $C(h)$ of each $Z(s)$ and $Z(s+h)$ pair is the same (independent of position) throughout the study region and depends on h (Trangmar et al., 1985):

$$C(h) = E[Z(s) - m][Z(s+h) - m]$$

As h gets larger, $C(h)$ decreases and the spatial covariance decays. For the second-order stationary hypothesis, the variance of the increment $Z(s) - Z(s+h)$ is finite and independent of position within the region, for all vectors of h (Trangmar et al., 1985):

$$VAR[Z(s) - Z(s+h)] = E[Z(x) - Z(x+h)]^2 = 2y(h)$$

The result $2y(h)$ dividing by two yields the semi-variance statistic $y(h)$.

Kriging is a technique of making optimal, unbiased estimates of regionalized variables at unsampled locations using the structural properties of the semi-variogram and the initial set of data values (Trangmar et al., 1985). A useful feature of kriging is that an error term (estimation variance) is calculated for each estimated value, providing a measure of the reliability of the interpolation (Trangmar et al., 1985). The simplest forms of kriging involve estimation of point values (punctual kriging) or areas (block kriging) and assume that the sample data are normally distributed and stationary (Henley, 1981; Trangmar et al., 1985).

The types of kriging can be simple kriging, ordinary kriging, kriging with a trend model, block kriging, factorial kriging, dual kriging (Goovaerts, 1997).

Kriging is based on proximity of the samples and the ordinary kriging is the standard method. As described by Lark (2012), the ordinary kriging assume an intrinsic stationarity requiring that:

$$E\left[Z_{(s)} - Z_{(s+h)}\right] = 0$$

Where $Z_{(s)}$ is a random variable at location s and h is a lag vector. If the data exhibit a strong trend the Universal kriging offer a solution, but is necessary to model the covariance of the residual process (Lark, 2012).

The predictions are based on the model (Hengl, 2009):

$$Z_{(s)} = \mu + \varepsilon'_{(s)}$$

Where μ is the constant stationary function and $\varepsilon'_{(s)}$ is the spatially correlated stochastic part of variation. The predictions are made as (Hengl, 2009):

$$Z_{OK}(s_0) = \sum_{i=1}^{n} w_i(s_0) \, . \, z(s_i) = \lambda_0^T . \mathbf{z}$$

Where λ_0 is the vector of kriging weights (w_i), $\mathbf{z}$ is the vector of n observations at primary locations. To estimate the weights, reflecting the spatial autocorrelation structure, it implemented the semivariograms. The semi-variance $y(h)$ describes the spatially dependent component of the random function Z. It is equal to half the expected squared distance between sample values separated by a given distance h (Isaaks & Srivastava, 1989; Trangmar et al., 1985):

$$y(h) = \frac{1}{2} E[Z(s) - Z(s+h)]^2$$

The semi-variance at a given lag h is estimated as the average of the squared differences between all observations separated by the lag (Trangmar et al., 1985):

$$y(h) = \frac{1}{2N(h)} \sum_{i=1}^{N} [Z(s_i) - Z(s_i + h)]^2$$

Where there are $N(h)$ pairs of observations. The semi-variogram for a given direction is usually displayed as a plot of semi-variance $y(h)$ versus distance h.

The range of dependence (*sill*) is the approximately constant value raised by the shape of the semi-variogram representing the semi-variance increasing with distance between sample locations (Trangmar et al., 1985). The sill approximates the sample variance σ^2 for stationary data. Samples separated by distances closer than the range are spatially related. The range also defines the maximum radius from which neighboring samples are drawn for interpolation by kriging (Trangmar et al., 1985).

The non-zero semi-variances occurred as h tends to zero is called nugget effect. The sum of the nugget variance C_0 and the spatial covariance C approximately equals the sill or sample variance σ^2 for stationary data (Trangmar et al., 1985).

The semi-variogram can be modelled by a spherical, exponential, Gaussian, linear function and in all directions (omnidirectional) or in a specific direction, analyzing its anisotropy (Isaaks & Srivastava, 1989; Goovaerts, 1997).

Grego et al. (2006) used geostatistical techniques to assess the variability for soil moisture content at depths of 20, 60 and 90cm, in an area cultivated under no tillage system, in Brazil. The higher spatial dependency was found close to 60 cm depth, with a range of 90 to 110 m. Snepvangers et al. (2003) compared a spatial-temporal ordinary kriging with a spatio-temporal kriging with external drift (net-precipitation was used as secondary information) to interpolate soil water content in a grassland in Southern Netherlands. The technique using a secondary information had a clear advantage, resulting in a decrease in prediction uncertainty and in a more realistic behavior. Souza et al. (2010) assessed the spatial variability of soil physical attributes and soil organic matter in an area under agriculture of sugarcane.

2.1.6 Regression-kriging

Regression-kriging (RK) is a technique for spatial data, which adds together the regression value of the variables and the kriging value of the residuals of the regression (Sun et al., 2012)

The methodology of regression-kriging in digital soil mapping firstly uses a regression technique (correlating input covariates derived from terrain with an output variable) to make prediction to the entire area. The errors are then computed by the difference between predicted and observed values. The errors might preserve other relation not used in regression, a spatial dependence. A geostatistical approach is used to compute the spatial correlation between errors. Then the error results from geostatistical approach is summed to previous results from regression. This technique is often named regression-kriging. The model ‘scorpan’ can be used to predict the soil properties of interest using environmental variables (McBratney et al., 2003), by regression techniques, and kriging can be used on the residuals.

The components of spatial variation can be modeled separately (Hengl, 2009):

$$\hat{z}(s_0) = \hat{m}(s_0) + \hat{e}(s_0) = \sum_{k=0}^{p} \hat{\beta}_k \,.\, q_k(s_0) + \sum_{i=1}^{n} \lambda_i \,.\, e(s_i)$$

Where $\hat{m}(s_0)$ is the fitted regression model, $\hat{e}(s_0)$ is the interpolated residual, $\hat{\beta}_k$ are estimated model coefficients ($\hat{\beta}_0$ is the estimated intercept), $q_k(s_0)$ are the values of the variables at the target location, λ_i are kriging weights determined by the spatial dependence structure of the residual and where $e(s_i)$ is the residual at location s_i. The regression coefficients $\hat{\beta}_k$ can be estimated from the sample by some fitting method, e.g. ordinary least squares (OLS) or Generalized Least Squares (HENGL, 2009). The regression component can also be one of the methods presented here previously as Random Forest, ANN, Model Trees or others. Once the part of variation from a regression has been estimated, the residual can be interpolated with kriging and added back to the values estimated by regression.

Goovaerts (1997) refers to the same technique as simple kriging with varying local means. The known stationary mean m may be replaced by known varying means $m_{SK}^{*}(u)$, leading to the simple kriging with varying local means. Different estimates of the primary local mean can be used, depending on the secondary information available. In the case of a secondary continuous attribute y, the primary local mean can be a function (linear or not) of the secondary attribute value at u:

$$m_{SK}^{*}(u) = f(y(u))$$

The function $f\big(y(u)\big)$ can be a multiple linear regression or other.

Goovaerts (1997) also describes another similar approach, a kriging with an external drift, when the mean is modeled as a linear function of a smoothly varying secondary (external) variable $y(u)$. However, unlike the simple kriging with varying local means, the mean $m(u)$ is not estimated through a calibration or regression process prior to the kriging of z.

Andrade and Mendonça-Santos (2016) predicted soil fertility in regions of Rio de Janeiro State, Brazil, using a multiple linear regression and a kriging of the residuals. The model indicated minimum CEC variation between the areas studied. Adhikari and Hartemink (2015) mapped topsoil SOC content in an area of Wisconsin, USA, using Cubist and residual kriging. A large number of variograms were simulated and the parameters were estimated using a restricted maximum likelihood. Bonfatti et al. (2016) compared different methods

to predict soil carbon content by five depths, in a valley area of Rio Grande do Sul State, Brazil. The method presented better results when using a multiple linear regression. A residual kriging was applied to previous model to improve the soil carbon prediction. Dorji et al. (2014) used Cubist and residual kriging to estimate soil organic carbon stocks by different depths under different land use and land cover in a montane ecosystem of Bhutan. The regression-kriging method is, according to the authors, an approach robust, quick and easy to produce soil property maps by using less data points of the target variable and several environmental covariates. Ge et al. (2011) used regression-kriging to relate soil properties to remote sensing images in a cotton field in Mississipi, USA. A stepwise multiple linear regression was used and the results presented higher accuracy when incorporating the spatial correlation of regression residuals, approximately improve 50% when predicting soil sodium concentration. Vanwalleghem et al. (2010) used a regression kriging approach to predict soil horizon thickness in loess-derived soils in natural forest areas in Central Belgium. The high spatial randomness in most horizons depth resulted in poor predictions, except for the upper eluvial E horizon.

2.2 Mechanistic models

Mechanistic models, when applied in soil mapping, search for physical explanation about the components and process involved in pedogenesis. Mechanistic models intend to explain why and how the phenomenon occurs, in general or specific cases, and involve representing and reasoning about nature (Bechtel, 2005). It is based on scientific considerations, rather than statistical convenience (Bretó et al., 2009; Bunge, 2004). Despite less common than an empirical approach, current studies in soil mapping has given attention to mechanistic models due to permit a real understanding about the soil formation and how it interacts with the landscape evolution. The models complement the currently studies based on empirical approaches (Minasny et al., 2015).

The complexity of the mechanistic models lies in the number of components and processes involved in pedogenesis, being impossible to model all the elements and interactions. The models need to have strong assumptions, to deal with the randomness and to disregard the elements with minor influences.

When generalizing the equations, it's important to identify similarities and differences between mechanisms operative in different circumstances, not just applying the same law to different conditions (Bechtel, 2005). Herein, we will give a brief explanation about the main process involved in mechanistic models of soil properties. Chapter 4 has an example of a mechanistic model applied to estimate soil thickness.

2.2.1 Cumulative and non-cumulative soils

The elements of pedogenesis might be originated form the process occurring in the place where the soil occurs or might have influence on the material transported from upslope areas. Considering the landforms, a mechanistic model need to be able to differentiate each condition and to choose the more appropriate method to model the soil attribute on each position on landscape.

Material accumulation on depressions close to streams is often noticed on landscapes and reproduced by models of landscape evolution. The downslope sections receive constantly sediments from upslope, transported mainly by gravity or runoff, and equations applied to models can simulate erosion and depositions, on uplands and lowlands. However not all deposit thickness, generally greater on lowlands, can be considered completely as soil thickness. Soil profiles can vary its characteristics depending on how fast is the deposition or erosion, including the soil thickness originated by pedogenical process (Birkeland, 1999).

The soils formed by sediments depositions and pedogenical process occurring concomitantly are considered cumulative soils (Birkeland, 1999; Nikiforoff, 1949; Schaetzl and Anderson, 2005). Their features are partly sedimentologic and partly pedogenic. In such soils, the A horizon evolves accumulating parent material deposited and can eventually become a B horizon, as sediments continue being deposited on surface while soil formation is going on.

Some positions in topography are favorable to cumulative soil profile formation. They are especially common in colluvial and fan deposits at the base of hillslopes (Birkeland, 1999). The properties of this soils are not consistent with

those in the surrounding area. Overthickened A horizon are common, due either to deposition of organic-matter-rich material from upslope, to organic-matter accumulation at the site while sediment is accumulating, or to a combination of both processes. It also may contain more clay to a greater depth than adjacent noncumulative soils. This occurs because the clay deposit from upslope areas and the clay formation or translocated in the profile (Birkeland, 1999).

The differentiation between cumulative and non-cumulative soils is important to model several soil properties, considering the influence from the other areas in landscape. The soil depth, for example, is influenced not only by the time of soil formation, but also by the amount and fastness of sediment deposition.

2.2.2 Time of soil formation

The time or soil age influences the degree of rock and mineral decomposition and the amount of soil formed. Young or immature soils have properties more similar to parent material and old or mature soils had more time to form new minerals and to lose the main characteristics of original rocks. The magnitude of any soil property is related to time (Jenny, 1941). A mechanistic model needs to consider the time of soil formation to estimate how is the current conditions and distribution of soil properties. Some properties, like organic matter content, form rather rapidly, whereas others, like high clay contents, take a much longer time (Birkeland, 1999).

Yaalon (1971) grouped the soil properties according the time necessary to reach a steady-state. The first group is of a relatively rapidly adjusting features, which rapidly ($< 10^2$ to 10^3 years) approach dynamic equilibrium with their environment, as the organic matter, nitrogen content, acidity or some types of structure. The soil properties might alter rapidly if the environmental condition change. The second group is of slowly adjusting features, which approach the steady state at a very slow rate ($> 10^3$ to 10^4 years), as a pedoturbation in a vertisol, redox processes in a pseudogley, the formation and destruction of clay coatings in an argillic horizon, the balance between weathering and erosion on a catenary slope. The development of soil property is slow and resistant to alterations. The third group is of self-terminating process,

including irreversible reactions, as the transformation of primary minerals and other weathering processes which involve the loss of material, e.g., color development of B horizon (Bockheim ,1980), oxic, petrocalcic, placic and petrogypsic horizons, plinthite, laterite, clay mineralogy and strongly develop argillic and natric horizon (Schaetzl and Anderson, 2005).

Birkeland (1999) compared soils and horizons formation time. A horizon qualified as diagnostic epipedons, probably require less than 5 kyr to form. Cambic horizon form rapidly and by 10 kyr should have good expression. Calcic (Bk) horizons might form as rapidly as the cambic horizon, and many environments would have them by 100 kyr. Argillic (Bt) horizons probably take between 100 kyr and 1 Myr. Kandic horizons would take longer than argillic horizons to form in a particular environment, because a longer duration of leaching is required to meet the key properties. Petrocalcic horizons of the K horizon variety take 100+ kyr to form. Oxic horizons take the longest to form from the average parent material, and they are put close to 1 Myr (Birkeland, 1999).

Regarding the soil classes, Ultisols might be formed on deposits and landscapes of more than 100 kyr, Oxisols 1 Myr or more, Entisols perhaps in a century. Histosols and Vertisols between 1 and 5 kyr. Inceptisols could form in that time, or slightly longer. Both Mollisols and Aridisols would require variable times to form because they include both cambic and argillic horizons; those with cambic horizons probably could form in 5 kyr and those with argillic horizon about 10 kyr. Alfisols, because an argillic horizon is required, probably also require about 10 kyr in wet climates and perhaps the same time in dust-influenced dry climates (Birkeland, 1999).

For a mechanistic model to have a reasonable soil properties predictions, it needs consider the time enough to soil or horizon formation. When it considers shorter times, a steady state might not be reached, making unrealistic predictions. Longer time are preferable, however it increases the model computation time, which imposes a limit.

2.2.3 Landscape and soil evolution models

Simulated models, based on physical knowledge, can model the erosion and deposition occurring in landscapes and to estimate future scenarios. A Landscape Evolution Model (LEM) can determine the elevation changes based on differences in sediment erosion and deposition over time. The LEM has wide use in Earth science, as point out by Willgoose (2005), and has being used as experimental tools to understand (a) the spatial organization of runoff generation and channel networks, (b) the remote sensing of erosion process parameters from digital elevation models, (c) the linking climate and tectonic processes with detailed stratigraphy, (d) spatial distribution of soils and vegetation and (e) the impact of continental-scale erosion and deposition on basin development and large-scale crustal dynamics feedbacks. In modern LEM's, there are spatial distribution and involvement of hillslope, channel and tectonic processes (Willgoose, 2005).

The landscape evolution influences soil evolution. The aggrading or degrading of material will change the thickness of sediment available to soil formation. So, the soil depth depends, between other factors, on the soil position in landscape. Soils in upland areas tends to be different than soils in bottom valleys or hills. Soils which sediment depositions and pedogenical process occurring concomitantly are considered cumulative soils (Birkeland, 1999; Nikiforoff, 1949; Schaetzl and Anderson, 2005). On stable surfaces, traditional pedogenic processes are operative and are less influenced by additions of material that is called topdown pedogenesis (Almond and Tonkin, 1999).

Simultaneously to the aggradation and degradation of sediments, there is a weathering of bedrock and thickening of sediment layer, which will be subject to pedogenic processes. It can be modelled by soil production functions. The Soil Production Function (SPF) is the functional dependence of bedrock conversion to soil on the overlying soil depth (Heimsath et al., 2000, 1997). The soil production rates decline exponentially with increasing soil depth, varying primarily according to lithology and climate zones.

There is a strong link between LEM and SPF. The spatial and magnitude variability of erosion, which model the landscape, will be influenced by differences in soil texture and soil organic carbon content in the course of soil

formation (Minasny et al., 2015), as well soils with different development will lead to the variation of vegetation and shaping the landscape form. For soil persisting it must be replenished at a rate equal to or greater than that of erosion. The rate is higher on superficial soils and declines with increasing soil mantle thickness (Heimsath et al., 1997). If local soil depth is constant over time (steady state condition), the soil production rate equals the erosion rate, which equals the lowering rate of the land surface.

It's important to consider the steady state condition to model soil depth and landscape. We can found in literature distinct concepts of steady state for soils and steady state for landscape. In a steady state soil thickness, the variation of soil depth is zero, and the soil production rate equals the erosion rate (Heimsath et al., 2000). In landscape, topographic and denudational steady-state is defined as a delicate balance of erosion and (constant) rock uplift such that a statistically invariant topography and constant denudation rate area maintained (Whipple, 2001). As initially low-relief landscapes are uplifted, erosion rates steadily increase over time in response to steepening of river profiles and adjacent hillslopes, further enhanced by orographic precipitation. Whipple (2001) point out that a steady-state topography and denudation are likely to prevail during periods of climatic stability, but rapid climatic fluctuation in the Quaternary appear to preclude the attainment of steady-state condition in modern orogens.

Few models have integrated landscape and soil formation, making possible evaluate its interaction, and the most have worked with hypothetical landscape (Minasny and McBratney, 2001; Saco et al., 2006; Temme and Vanwalleghem, 2015). Saco et al. (2006) used the SIBERIA model combined with soil production rate, evaluating the use of spatially varied soil moisture. Results were consistent with Heimsath et al. (1997) showing exponential decline in soil production rate with soil depth. The effect of spatial variability driven by subsurface water availability was explored. The LORICA model (Temme and Vanwalleghem, 2015) was built based on the landscape evolution LAPSUS and the soil formation MILESD. First results show soil-landscape interactions, where the surface changes in the landscape depend on soil development, and soil changes depend on landscape location.

Models using real landscapes have been driven mainly by its geomorphological and hydrological charateristics. The mARM model used a soil

grading, topographic and meteorological characteristics in Ranger Uranium Mine (Australia) and simulate temporal and spatial varying soil erosion and armour development (Cohen et al., 2010, 2009). Vanwalleghem et al. (2013) used the model MILESD on an area in Australia. The model includes physical and chemical weathering, clay migration and neoformation, bioturbation and carbon cycling. The results showed the importance of considering the soil-forming processes interacting with soil redistribution by erosion and deposition. This model predicts trends in total soil thickness along a catena and predicted soil texture and bulk density with errors in the order of 10%. Nicótina et al. (2011) included surface hidrology in the model, for an area in Idaho (USA) and found trend in soil depth distribution and a scatter graphic between modeled and observed soil depth. Catani et al. (2010), studying an area in Italy, obtained good results do predict soil thickness using an empirical geomorphology-based model (Geomorphologically Indexed Soil Thickness – GIST) dealing with curvature, position along the hillslope profile and slope gradient.

2.3 Summary points

Important methods to digital mapping of soil properties were presented in this study. There are several approaches and derivations, and it's important to know how the models are implemented, aiming to choose the most appropriated. The potential of each method becomes clear as the environmental covariates or physical equations can be used to predict soil properties, inferring properties in places not sampled. Table 1 shows studies using the models presented.

There is no general best method for predictions, based on validations in several studies. Regarding empirical models, regression-kriging has been showed as a preferred method, producing results more consistent with the spatial dynamics of soil properties. Some empirical methods have limitations relating to overfitting and this should be verified to avoid interpretation errors. A good practice is to use an independent set of representative samples to validate the models. Mechanistic models use complementary equations and the challenge is to discover the correct equations and how it would be implemented in a complex model. In practice, it is impossible to model all the natural phenomena occurring in landscape and the technique needs to evaluate the most important processes.

Generally, empirical models may present better results when the samples are dense and representative in the study area. If it is not the case, mechanistic models may produce a more appropriate response to the predictions of soil properties, besides providing pedogenic knowledge by each soil property.

Table 1. Studies using empirical and mechanistic models in soil science.

Model	Method	Estimated soil characteristic	Reference
Empirical	Multiple Linear Regression	Soil moisture	García et al. (2016)
		Soil microbial biomass	Lentzsch et al. (2005)
		Soil nitrate concentration	Nunes et al. (2012)
		Soil moisture	Qiu et al. (2010)
		Soil quality	Zornoza et al. (2007)
	Random Forest	Soil texture	Chagas et al. (2016)
		Soil organic carbon	Grimm et al. (2008)
		Soil organic carbon, pH, sand, silt and clay fractions, bulk density, cation-exchange capacity, total nitrogen, exchangeable acidity, Al content and exchangeable bases	Hengl et al. (2015)
		Soil texture	Ließ et al. (2012)
		Soil carbon stocks, total soil carbon, total soil nitrogen, total soil sulphur	Wiesmeier et al. (2011)
	Model Trees	Soil organic carbon concentration	Akpa et al. (2016)
		Soil properties from diffuse infrared reflectance	Minasny & McBratney (2008)
		Soil properties from spectral data	Peng et al. (2015)
		Soil properties from spectral data	Viscarra Rossell et al. (2016)
	Neural Network	Soil organic matter, soil bulk density and soil organic carbon content	Aitkenhead & Coull (2016)
		Soil organic carbon	Tiwari et al. (2015)
		Soil similarity and soil A horizon depth	Zhu (2000)
	Geostatistics	Soil moisture	Grego et al. (2006)
		Soil water content	Snepvangers et al. (2003)
		Soil physical attributes and soil organic matter	Souza et al. (2010)

Table 1. Continuation...

	Regression-kriging	Soil fertility	Andrade & Mendonça-Santos (2016)
		Soil organic carbon content	Adhikari & Hartemink (2015)
		Soil carbon content	Bonfatti et al. (2016)
		Soil organic carbon stocks	Dorji et al. (2014)
		Relating soil properties to remote sensing images	Ge et al. (2011)
		Soil horizon thickness	Vanwalleghem et al. (2010)
Mechanistic	SIBERIA model combined with SPF	Soil production rate and soil thickness	Saco et al. (2006)
	SPF and equations of geomorphological processes	Soil production rate and soil thickness	Heimsath et al. (1997)
	LORICA model	Soil erosion and deposition, physical and chemical weathering, fine clay, soil thickness, bioturbated mass and organic matter input	Temme and Vanwalleghem (2015)
	mARM model	Soil erosion and armour development	Cohen et al. (2010, 2009)
	MILESD	Soil thickness, soil texture and soil bulk density	Vanwalleghem et al. (2013)
	SPF and equations of geomorphological processes	Soil thickness	Nicótina et al. (2011)
	GIST	Soil thickness	Catani et al. (2010)

3. CAPÍTULO II – ESTUDO 1: DIGITAL MAPPING OF SOIL CARBON IN A VITICULTURAL REGION OF SOUTHERN BRAZIL [1]

3.1 Introduction

Assessing the amount and distribution of soil organic carbon (SOC) levels is important as it provides information about soil fertility, rates of sequestration of carbon, recovery of degraded soil, or the impact of land use changes. Mapping the SOC concentration and stocks is challenging because of the considerable variation and dynamics. Spatial and temporal SOC changes are affected by natural and anthropic factors including management practices and land use changes.

Several recent studies have predicted and mapped SOC (Adhikari et al., 2014; Padarian et al., 2012; Kirsten et al., 2015; Malone et al., 2009; Mendonça-Santos et al., 2010; Ross et al., 2013; Zhang and Shao, 2014) and the estimation is based on relation between covariates (land use, soil type, slope, aspect, etc.) and SOC levels. Different covariates were found in models to explain SOC distribution. Thompson and Kolka (2005) found that more than 71% of SOC variation could be explained by slope, aspect, curvature, topographic wetness index and overland flow distance. Wiesmeier et al. (2014) found that the most important factors to predict SOC stocks were land use, soil type, soil moisture and climate. Adhikari et al. (2014) predicting SOC concentration, at different soil depths, reported that the importance of variables differed by depth. Minasny et al. (2013) synthesized a large number of digital SOC mapping studies and concluded that different covariates could explain the variation of SOC depending on the complexity of the landscape.

1. Adapted from article published in Geoderma Journal.
http://dx.doi.org/10.1016/j.geoderma.2015.07.016. 0016-7061/

The majority of SOC inventory assessments to date focused the 0-20 cm or 0-30 cm surface layers, whereas considerable amounts of SOC may be present deeper in the soil profile (Lal, 2005; Rumpel and Kögel-Knabner, 2011; Minasny et al., 2013; Boddey et al., 2010). Sisti et al. (2004) studied SOC stocks down to 100 cm depth with zero tillage and conventional tillage and found, in rotations with vetch planted as a winter green-manure crop, significantly higher soil carbon and nitrogen concentrations under zero tillage, with most of the differences occurring at 30-85 cm depth. Angers & Eriksen-Hamel (2008) showed different interpretation of SOC stocks when considering different depths, in no till and full-inversion tillage. Full-inversion tillage could accumulate more carbon at the bottom of the plow layer, but the SOC does not completely offset the gain under no till in the surface horizon. The authors highlight the importance of taking into account the whole profile to understand the distribution of SOC stocks.

Land use has major impacts on SOC concentration and stocks. However, these effects are also affected by soil class and depth (Hartemink and McSweeney, 2014; Nieder & Benbi, 2008). Changes in land use impacts the SOC levels and modifies soil characteristics. Several studies explained the changes of SOC with land use change. Conant et al. (2001), reviewing 115 studies, found that conversion from native land (mostly rain forest) to pasture increased soil C content for nearly 70% of the studies. Guo & Gifford (2002), compiling 74 publications, found that SOC stocks declined after land use changed from pasture to plantation (-10%), native forest to plantation (-13%), native forest to crop (-42%), and pasture to crop (-59%). However, the SOC stocks increased when the native forest was converted to pasture (+8%), crop to pasture (+19%), crop to plantation (+18%), and crop to secondary forest (+53%). Cerri & Andreux (1990) showed that C levels after 50 years of sugarcane cultivation, in São Paulo, Brazil, were 46% of the levels under primary forest.

Although there is a considerable body of research on the digital mapping of SOC in temperate regions, few studies have been conducted in the tropical and subtropical areas. Examples include Berhongoray et al (2013) estimating SOC stocks in Argentine Pampas, Cheng et al. (2004) predicting SOC concentration in a subtropical area in China, Vasques et al. (2010) estimating SOC stocks in a subtropical watershed in Florida. Digital soil mapping has been used in Brazil (Giasson et al., 2006; Mendonça-Santos & Santos, 2007) and

examples of SOC predictions include the studies by Mendonça-Santos et al. (2010) whom used regression-kriging for evaluate the SOC stocks in Rio de Janeiro State, and Souza et al. (2014) using regression-kriging to predict SOC and clay content in Rio Doce Basin (Minas Gerais State). There have been other studies (e.g., Cerri et al., 2007; Tornquist et al., 2009b) where ecosystem models such as Century or Rothamsed C Model were applied to estimate SOC dynamics in the upper soil layers from different areas in Brazil.

The present study aimed to analyze the distribution of SOC in the grape growing region of Vale dos Vinhedos, in Rio Grande do Sul State, Brazil. The objectives were: (i) to compare different methods to predict SOC content, (ii) to quantify and understand the spatial variation of SOC concentration by depth through digital soil mapping, and to assess the uncertainty, (iii) to quantify and map SOC stocks, and (iv) to estimate SOC changes due to land use change.

3.2 Materials and methods

3.2.1 Study area

The study was conducted in the Vale dos Vinhedos (Vineyard Valley) which is a wine production region in northeastern Rio Grande do Sul State (Fig. 1). The study area covered 8,118 hectares. The climate is classified as Cfb, subtropical with a mild summer, mean annual temperatures of 17.2°C and 1,777 mm annual rainfall (EMBRAPA, 2017). The dominant lithology is effusive rocks mostly from the Mesozoic Era (IBGE, 1986). Lower sequence comprises mostly basalts and diabase dikes, whereas the upper sequence has predominantly acid effusive rocks such as rhyolite and dacites.

Average soil depth is 150 cm (range 25 to > 250 cm) and many soils are stony and rocky (average 4.5% of fragments > 20 mm in diameter). According the *Soil Taxonomy* map, in the study area, Inceptisols cover about 44%, Ultisols 29% and Mollisols almost 15% (Fig.2). Mollisols are mostly present at lower elevations close to valley bottoms in the northern part of the study area.

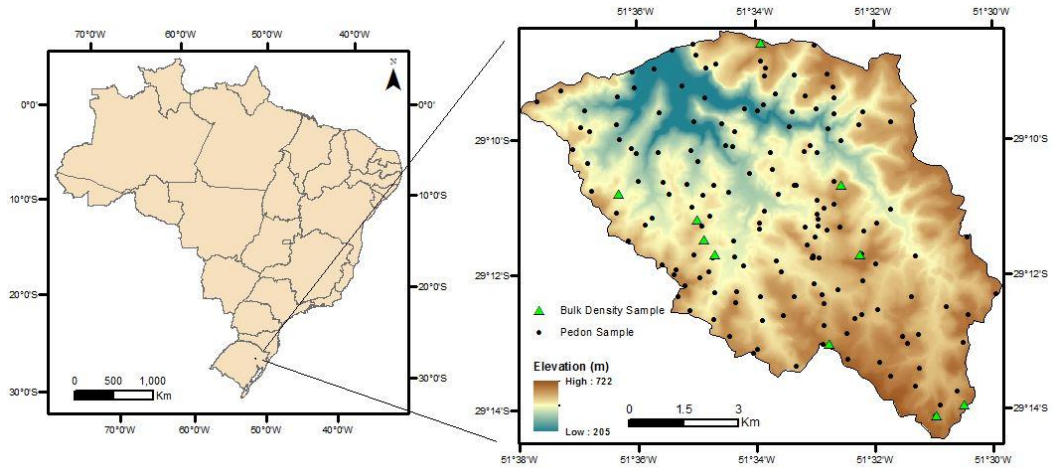


Figure 1. Study area (Vale dos Vinhedos) in Rio Grande do Sul, Brazil (8,118 ha) and location of the 163 pedons and 10 bulk density sampling points

Soils in the western part of the study area are mainly Argissolos (Ultisols and Alfisols), Chernossolos (Mollisols), and Neossolos (Entisols and Mollisols). The eastern part has more rugged terrain and the dominant soils are Neossolos (Entisols) and Cambissolos (Inceptisols), with association of Argissolos (Ultisols and Alfisols), Latossolos (Oxisols) and Nitossolos (Oxisols and Ultisols) (Flores et al., 2012). Forest (44%) and Vineyard (31%) are the dominant land use in the study area, according the land use classification map (Fig. 2).

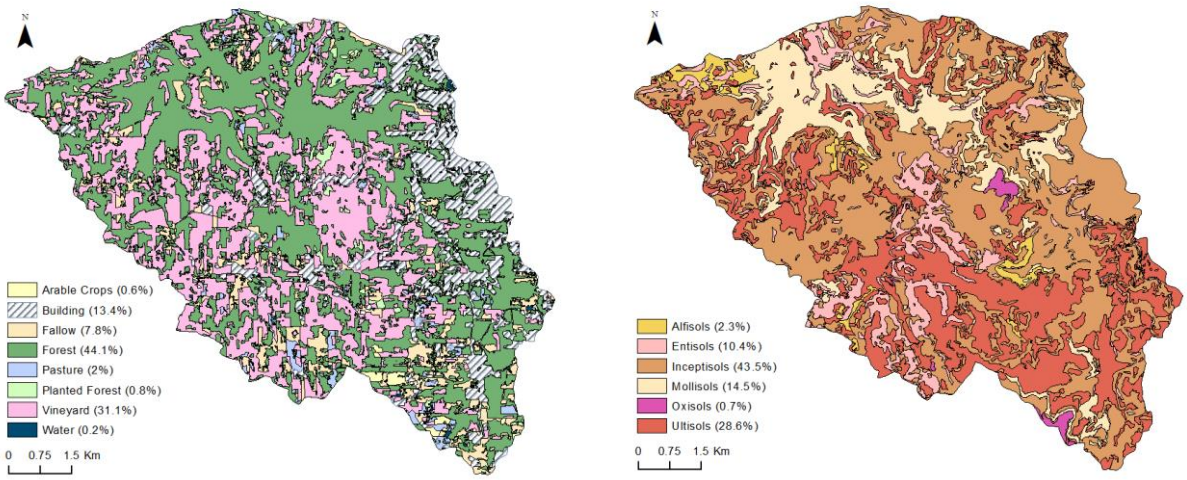


Figure 2. Land use and *Soil Taxonomy* map of Vale dos Vinhedos in Rio Grande do Sul, Brazil. Percentages of different land use and soil order classes in parentheses.

3.2.2 Soil and environmental data

The soil data were obtained from the soil survey project “Os Solos do Vale dos Vinhedos” (Flores et al., 2012). Sample points were selected along predefined paths representing different landscape units (Flores et al., 2012). Sampling was done with 163 total pedons, comprising 580 soil horizons. The soils were analyzed following Brazilian standard methods (Santos et al., 2006), SOC analysis by Walkley-Black wet oxidation. The soil pedons distribution by soil order classes and land use is in Table 2.

Table 2. Distributions of soil pedons by soil order classes and land use.

Soil Order	Arable Crops	Fallow	Forest	Pasture	Planted Forest	Vineyard
Alfisol	0	0	0	2	0	4
Entisol	0	1	6	1	1	7
Inceptisol	3	4	7	2	2	46
Mollisol	1	2	1	0	0	13
Oxisol	0	0	1	0	0	0
Ultisol	2	9	4	1	2	41

Additionally, in 2014, samples were obtained from 10 pedons (34 horizons) for an estimate of soil bulk density of the Flores et al. (2012) soil survey, allocated by contrasting land uses (vineyard, forest/planted forest, pasture, arable crops, and fallow) and soil classes. The 10 measured bulk density were used to evaluate three pedotransfer functions, which were chosen based on studies that include data from subtropical soils. Table 3 lists measured bulk density. Once the bulk density was calculated, the values were splined to derive bulk density for the 5 GlobalSoilMap standard depths. These values were then attributed to each map unit of Flores et al. (2012) soil map (scale 1:10.000) considering the reference soil profiles, extrapolating then to the whole study area.

On SOC concentration and soil depth predictions the following data layers from Flores et al. (2012) were used: 5x5 m grid resolution DEM, a soil map (scale 1:10,000) and orthorectified aerial imagery. The DEM was upscaled to 15 m grid cell size.

Table 3. Bulk density (t/m3) for different land use and soil depths (cm), obtained from field measurements (10 soil pits).

Land Use	Depth 1	Depth 2	Depth 3	Depth 4
Vineyard	1.17 (11 cm)	1.20 (20 cm)	1.22 (35 cm)	-
Vineyard	1.14 (7 cm)	1.17 (16 cm)	-	-
Vineyard	1.13 (9 cm)	1.21 (34 cm)	1.22 (60 cm)	1.25 (81 cm)
Vineyard	1.16 (13 cm)	1.35 (35 cm)	1.17 (60 cm)	-
Forest	0.97 (25 cm)	1.07 (44 cm)	1.13 (63 cm)	1.23 (85 cm)
Forest	1.02 (20 cm)	1.08 (45 cm)	1.28 (75 cm)	-
Planted Forest	1.09 (15 cm)	1.27 (50 cm)	1.40 (82 cm)	1.33 (124 cm)
Pasture	1.15 (10 cm)	1.16 (33 cm)	1.25 (51 cm)	-
Arable Crops	1.10 (7 cm)	1.55 (30 cm)	1.44 (45 cm)	1.28 (73 cm)
Fallow	1.29 (40 cm)	1.33 (59 cm)	1.21 (94 cm)	1.16 (118 cm)

The original soil legend of the Flores et al. (2012) survey, published according to the Brazilian soil classification (Embrapa, 2013), was converted to *Soil Taxonomy* (Soil Survey Staff, 2014) using pedon data (clay content, pH, thickness, carbon content, texture, color, clay skins and drainage) and additional guidance from the correlation table proposed by Anjos et al. (2012).

A land use map was made using the orthorectified mosaic of aerial images from November 2005 (Flores et al., 2012). Initially, a supervised classification was performed after the images were filtered 3 times (3x3, 5x5, 7x7) using the mean. The supervised classification identified land uses for approximately 50% of the area particularly in the forested areas. Land use in the other half area was delimited manually. The final land use map contains 8 classes namely vineyard, forest, planted forest, pasture, arable crops, fallow, building and water bodies. Building and water bodies were masked (Fig. 2).

A set of terrain attributes was derived from the DEM including Slope, Aspect, Valley Depth, Topographic Wetness Index, Overland Distance to Channel Network and others. A map with 13 landform classes was made in LandMapR software using the DEM (MacMillan, 2003). The covariates used for predicting the SOC levels and soil depth are presented in Table 4.

Table 4. Variables used in the prediction of SOC content (g/kg) and soil depth of the study area in Vale dos Vinhedos in Rio Grande do Sul, Brazil

Variables	Data descriptions	Type	Mean (Min-Max)	Soil carbon	Soil depth
Digital Elevation Model – 15m	Elevation above mean sea level	Numeric	541.49 (206.15 -723.05)	X	
Coordinate X	UTM Latitude	Numeric	445052 (438242 -451863)	X	
Coordinate Y	UTM Longitude	Numeric	6770978 (6765083 - 6776873)	X	
Slope	Local hill slope gradient	Numeric	13.85 (0 - 81.01)	X	X
Aspect	Slope aspect	Numeric	180 (0 - 360)	X	X
Analytical Hillshading	Angle between the surface and the incoming light beams	Numeric	0.93 (0 - 2.74)	X	X
TWI	Topographic Wetness Index	Numeric	3.12 (0.27 - 8.76)	X	X
LS Factor	Slope length factor	Numeric	3.71 (0 - 95.9)	X	X
Vertical Distance to Channel Network	Altitude above channel network	Numeric	26.86 (0 - 259.84)	X	X
Valley Depth	Relative position of the valley	Numeric	24.48 (0.36-275.18)	X	X
Slope Height	Vertical distance from the base of the slope to the crest	Numeric	24.91 (0.03 - 309.36)	X	X
Normalized Height	Height position within a reference area	Numeric	0.49 (0 - 1)	X	X
Mid Slope Position	Cover the warmer zones of slopes	Numeric	0.52 (0 - 1)	X	
Flow Direction	Direction of the flow	Numeric	35.06 (1 - 255)	X	X
MRVBF	Identifies the depositional areas	Numeric	0.21 (0 - 4.92)	X	X
Overland Flow Distance to Channel Network	Distance from non-channel cells to channel cells	Numeric	204.24 (0 - 1385)	X	X
Direct Insolation	Potential incoming solar radiation	Numeric	3.29 (0 - 5.71)	X	
Soil Order	Soil order map	Categorical	6 classes	X	X
Land Use	Land use map	Categorical	6 classes	X	
Convexity	Terrain surface convexity	Numeric	0.51 (0.26 - 0.78)		X
Topographic Position Index (TPI)	Compare elevation of each cell to the neighborhood	Numeric	-0.05 (-11.14 – 21.97)		X
Mass Balance Index (MBI)	Balance between soil mass deposited and eroded	Numeric	0.16 (-0.81 - 1.61)		X
Plan Curvature	Curvature in a horizontal plane	Numeric	0 (-0.16 – 0.31)		X
Vector Ruggedness Measure (VRM)	Measures terrain ruggedness	Numeric	0.01 (0 - 0.12)		X
LandMapR	Landform classification	Categorical	12 classes		X

3.2.3 Prediction models

Following GlobalSoilMap specification (Arrouays et al., 2014) until 1 m depth, equal area splines were used to harmonize the SOC concentration and bulk density data for 5 depth intervals: 0-5, 5-15, 15-30, 30-60, and 60-100 cm. The smoothing parameter lambda chosen was 0.1 (Malone et al., 2009).

The splined data were randomly splitted into 75% (122 pedons) for training the model, and 25% (41 pedons) for validation. The training data was used to predict SOC concentrations and all the pedons were used to predict soil depth. Four different regression models were tested: Multiple Linear Regression (MLR), Stepwise Multiple Linear Regression (SMLR), Cubist, and Random Forest.

In MLR, each independent variable is weighted by the regression to ensure maximal prediction from the set of independent variables (Hair et al., 2009). The weights denote relative contribution of the independent variables and facilitate to know the influence of each variable. However, correlation among independent variables need to be considered. In SMLR, each variable is considered to be included prior to developing the equation. The independent variable with the greatest contribution is added first, followed by the variables selected based on their incremental contribution over the variables already in the equation (Hair et al., 2009). The Cubist model is based on the M5 algorithm of Quinlan (1992). The M5 algorithm builds tree-based models, which may have multivariate linear models at their leaves (Quinlan, 1992). It first partitions the data into subsets within which their characteristics are similar with respect to the target variable and the covariates. There are several rules arranged in hierarchy. The Random Forest is an ensemble learning method for classification (and regression) that operate by constructing a multitude of decision trees at training time which are later aggregated to give one single prediction for each observation in a data set. For regression, the prediction is the average of the individual tree outputs (Breiman, 2001; Malone, 2013).

The soil depth map had values ranges between 0 to 250 cm. The map was sliced into 5 layers: 0-5cm, 5-15cm, 15-30cm, 30-60cm, 60-100cm, creating 5 sets of thickness data. The thickness data were used to calculate SOC stocks, by each depth intervals.

The gravel and stone contents were obtained from the 163 pedons (Flores et al., 2012) and the distribution by depth was estimated by equal area splines. The values were extrapolated to the entire area through reference profiles of soil map units (Flores et al., 2012).

To calculate SOC stocks in t/ha, SOC concentrations in mass fraction were multiplied by bulk density previously calculated and mapped (section 2.2) and thickness for each depth, and corrected for gravel and stone contents, according to equation:

$$SOC\left[\frac{ton}{ha}\right] = \left[\frac{SOC\left[\frac{g}{kg}\right] \times BD\left[\frac{g}{cm^3}\right] \times thickness[cm]}{10}\right] \times \left(1 - \frac{gravels[\%]}{100}\right) \quad \text{Equation (4)}$$

To compare SOC stocks in soils under different land use and soil types, the results needed to be corrected by mass, avoiding carbon stock variation due to bulk density changes. The cumulative mass approach should be preferred as the basis for carbon stock accounting on a fixed mass per unit area (Minasny et al., 2013). The approach from Gifford and Roderick (2003) was used to calculate the cumulative mass and SOC stocks down to 1 m profile. This approach corrects SOC stocks using a reference cumulative mass. Soil mass of the forest areas were chosen as reference, as represent mass of soils under natural vegetation, and were calculated by the measured bulk density splined and the respective thickness. The reference mass by interval depth were 4.95 g/cm^2 for 0-5cm, 9.9 g/cm^2 for 5-15cm, 15.15 g/cm^2 for 15-30cm, 33.6 g/cm^2 for 30-60 cm and 47.6 g/cm^2 for 60-100 cm. For the whole area, the soil mass for each depth was calculated by the bulk density maps and the thickness layers derived from the soil depth map. Then, the reference soil mass, the soil mass for each depth of the entire study area, and the previously calculated SOC stocks were each one cumulatively summed.

The cumulative corrected SOC stocks to the entire area, for each depth, were calculated through the equation applied to each pixel:

$$c_s(t) = c_s(z_a) + \frac{c_s(z_b) - c_s(z_a)}{m_s(z_b) - m_s(z_a)}\left(m_s(t) - m_s(z_a)\right) \quad \text{Equation (5)}$$

Where $c_s(t)$ is the value of cumulative SOC stocks corrected by mass; $c_s(z_a)$ and $m_s(z_a)$ are the value of cumulative SOC stocks and mass, respectively, from the lower boundary of the layer above it; $c_s(z_b)$ and $m_s(z_b)$ are the cumulative SOC stocks and mass of the lower boundary of the current layer; $m_s(t)$ is the cumulative soil mass from the lower depth of the reference layer.

The SOC stocks for each interval depth was calculated by subtracting the cumulative SOC stocks of the lower and upper limit from respective layer.

.

3.2.4 Prediction evaluation

The SOC concentration models were validated with 25% of the data, and the soil depth model with the whole dataset, using 4 statistical parameters: RMSE, ME, R^2 and CCC. The R^2 is the coefficient of determination of linear regression, between the observed values and predicted values. RMSE correspond to root mean square error, ME to the mean error, and CCC to the Lin's Concordance Correlation Coefficient. The R^2 was obtained directly from the model in R, whereas other parameters were calculated as follows:

$$\mathrm{ME} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}} \hat{\mathrm{z}}(\mathrm{x_i}) - \mathrm{z}(\mathrm{x_i}) \qquad \text{Equation (6)}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}[\mathrm{z}(\mathrm{x_i}) - \hat{\mathrm{z}}(\mathrm{x_i})]^2} \qquad \text{Equation (7)}$$

$$\mathrm{CCC} = \frac{2*\rho*\sigma_{\hat{\mathrm{z}}(\mathrm{x_i})}*\sigma_{\mathrm{z}(\mathrm{x_i})}}{\sigma^2_{\hat{\mathrm{z}}(\mathrm{x_i})} + \sigma^2_{\mathrm{z}(\mathrm{x_i})} + (\overline{\hat{\mathrm{z}}(\mathrm{x_1})} - \overline{\mathrm{z}(\mathrm{x_1})})^2} \qquad \text{Equation (8)}$$

Where n is the number of the validation sample points, $\mathrm{z}(\mathrm{x_i})$ is the observed value, $\hat{\mathrm{z}}(\mathrm{x_i})$ is the predicted value, $\sigma^2_{\mathrm{z}(\mathrm{x_i})}$ and $\sigma^2_{\hat{\mathrm{z}}(\mathrm{x_i})}$ are the variances, and ρ is the correlation coefficient between the predictions and observations. The impact of each variable was measured by the absolute value of *t*-statistics for each model parameter obtained through MLR.

3.2.5 Uncertainty and probability maps

Estimating uncertainty is complex considering all the sources of uncertainty. There are a number of approaches to estimate the uncertainty and Malone et al. (2011) and Shrestha and Solomatine (2006) suggest the empirical approach. In this approach, the residuals between modelled outputs and corresponding observed data are used to formulate prediction intervals (PI's). The uncertainty is expressed in the form of two quantiles of the underlying distribution of model error (residuals). The PI takes into account all sources of uncertainty and circumvents attempts to separate out the contribution of each source of uncertainty (Malone et al., 2011; Shrestha and Solomatine, 2006; Solomatine and Shrestha, 2009). The methodology is independent of the prediction model structure, as it requires only the model outputs.

We used the empirical approach estimating PI's through the residuals of SOC predictions. Between several methods for modeling the distribution of residuals, we chose sequential Gaussian geostatistical simulations because it is more related to the spatial method used for SOC prediction.

Firstly, the residuals from MLR prediction of SOC, at each 5 standard depth, were simulated with 100 iterations and then the outputs were added back to the predicted SOC concentration. For each predicting pixel, we considered the two percentiles, lower 5% and upper 95%, covering the 90% PI, as suggested in GlobalSoilMap specifications (Arrouays et al., 2014). Lower and upper limits were mapped (Fig.5.) and the uncertainty models were evaluated on the 25% validation dataset (Fig.9.).

With the 100 values of SOC concentration, we applied Eq. 4 for produced SOC stocks and Eq. 5 to correct by cumulative mass. The results were used to produce maps of probability of total SOC stock (0 – 100 cm) that exceed a threshold of 184 t C/ha. The value of 184 t C/ha is based on the averaged SOC stocks under forest for the entire study area. Areas with high probability of exceeding this limit are likely to have the same SOC stocks as under forest. The number of times that pixels values exceeded the threshold, between 100, was counting and recording for producing the maps. The following probabilities were considered for mapping: 20%, 40%, 60% and 80%.

3.2.6. SOC changes

SOC changes due to land use changes were estimated using Projected Natural Vegetation Soil Carbon (PNVSC) approach (Waring et al., 2014). PNVSC is considered a projected SOC that could be present today if the area was under natural vegetation.

The PNVSC maps were elaborated re-applying the equations produced by MLR models for SOC concentration (section 3.2.3), nevertheless with the coefficients for land use types other than forest set to zero. The produced maps are hypothetically representing soil carbon which could be observed today if the whole study area remained under natural vegetation. The predicted SOC concentration can now to be compared with PNVSC by Eq. 9 to estimate SOC changes due to land use change (e.g. Adhikari and Hartemink, 2015):

$$SOC_{changes} = SOC_{predicted} - PNVSC \quad \text{Equation (9)}$$

Negative SOC change indicates that the soil has less SOC compared to the projected natural vegetation (forest) whereas positive SOC change indicate that the soils have accumulated SOC.

3.3 Results

3.3.1 SOC prediction and model comparison

While comparing four prediction methods using the training data, Cubist and Random Forest showed a high R^2 (>0.92) and CCC (>0.8) for SOC prediction at all depths, compared to MLR and SMLR with lower values (R^2 <0.51 and CCC<0.64). Similarly, both Cubist and Random Forest had lower RMSE (<6.4 g/kg) than MLR and SMLR (RMSE >6.7), when comparing all the soil depths.

However, when using the validation samples (25% of pedons), MLR had a higher R^2 and CCC and a lower RMSE than the other models (Fig. 3b). Comparing the distribution of observed and predicted values (Fig. 3a), MLR had less spread of points. The MLR model suffered from the higher bias (ME)

compared to all other methods. RMSE did not differ much between methods with Cubist showing the highest RMSE at all depths.

Based on R^2 and CCC values, and considering the RMSE not so different between methods, MLR was the best model to estimate the SOC concentration. The high bias indicates that MLR might overestimate the predicted values and, therefore, it should be considered when interpreting the results. Based on these findings, we assumed that MLR could be the appropriate method for SOC concentration and soil depth prediction.

After the predictions were made by the MLR models, the distribution of prediction residuals and their spatial dependence was analyzed and plotted (Fig. 4). Spatially, the residuals were poorly auto-correlated for the top three depths (0-5, 5-15, and 15-30 cm), whereas a better spatial structure was observed below 30 cm soil depth. Residuals of soil depth prediction were normally distributed with no spatial auto-correlation as suggested by the pure nugget effect of the variogram.

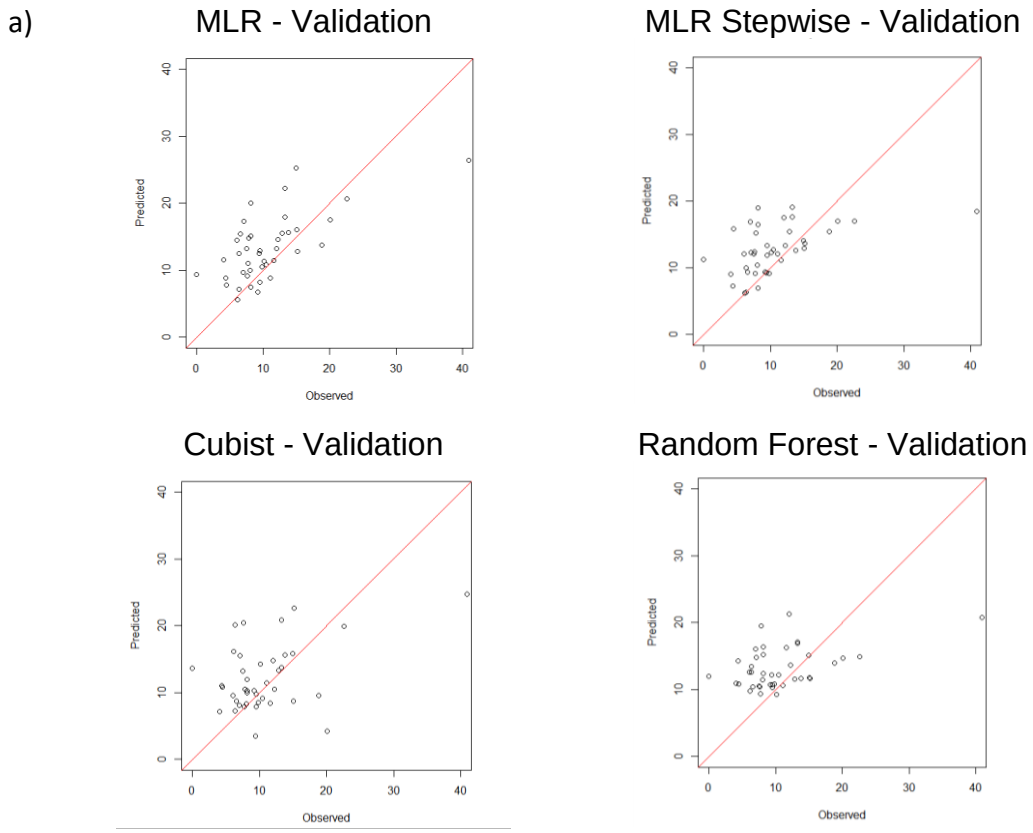


Figure 3. a) Example of distribution of predicted x observed values, depth 30 to 60cm

b)

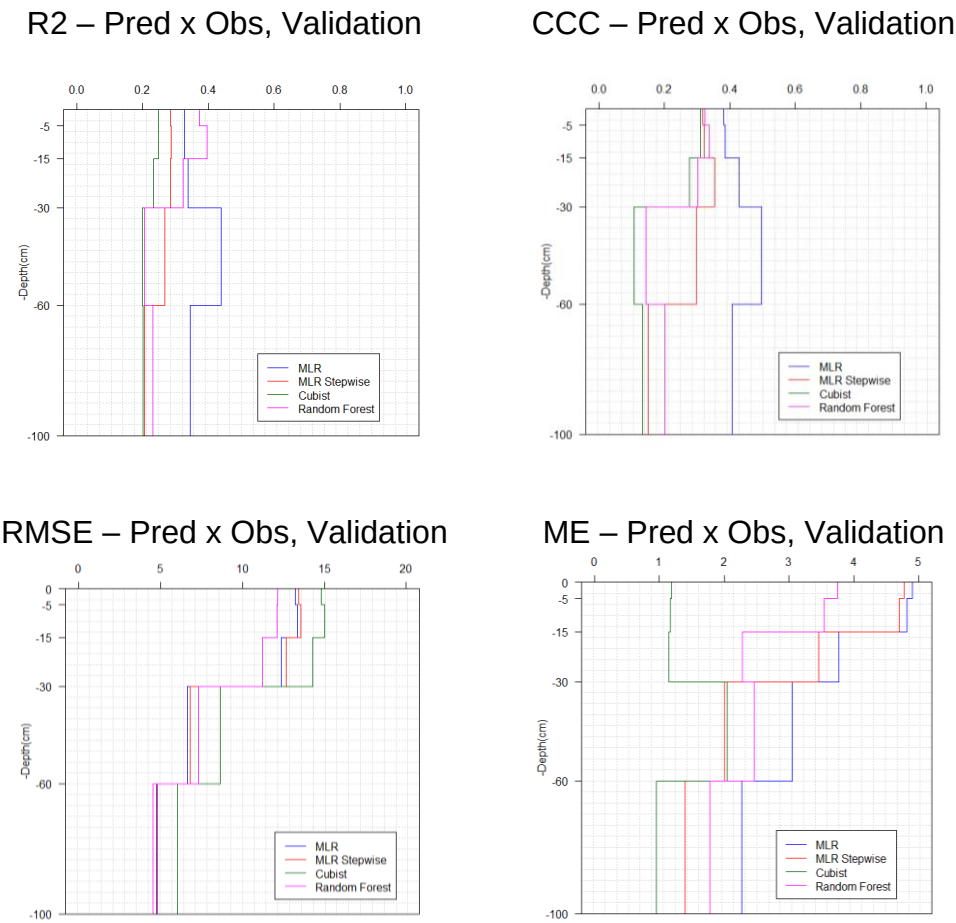


Figure 3. b) validation of SOC concentration prediction for all depth intervals, by 4 different methods.

3.3.2. Evaluation of MLR model for SOC and soil depth prediction

Descriptive statistics of training data, validation data, and the estimates are shown in Table 5. The mean SOC concentration for the validation are slightly lower than for training data. The estimated data have higher mean, median and maximum values than the training data which suggests that the MLR model can overestimate SOC concentration. The bias of MLR is depicted in Fig .3b. and in Table 6, showing higher values at 0-5 cm soil depth and decreasing until depth 60-100 cm. For the soil depth, the mean of estimated data was similar to training data.

Validation results of the MLR model is shown in Table 6. The R^2 between predicted and measured values differed by depth and was highest at 30-60 cm, with value of 0.44, and CCC of 0.49. For all other depths, R^2 was between 0.33 and 0.34, and values of CCC were between 0.38 and 0.43. The

RMSE and ME decreased with increasing soil depth. When the residuals were added to the MLR predictions, the R^2 and CCC increased and RMSE and ME decreased (Table 6).

For the soil depth prediction model, CCC of 0.59, R^2 of 0.43, and RMSE of about 34.8 cm were observed. The high values of R^2 and CCC were probably due to the use of same samples for model training and validation.

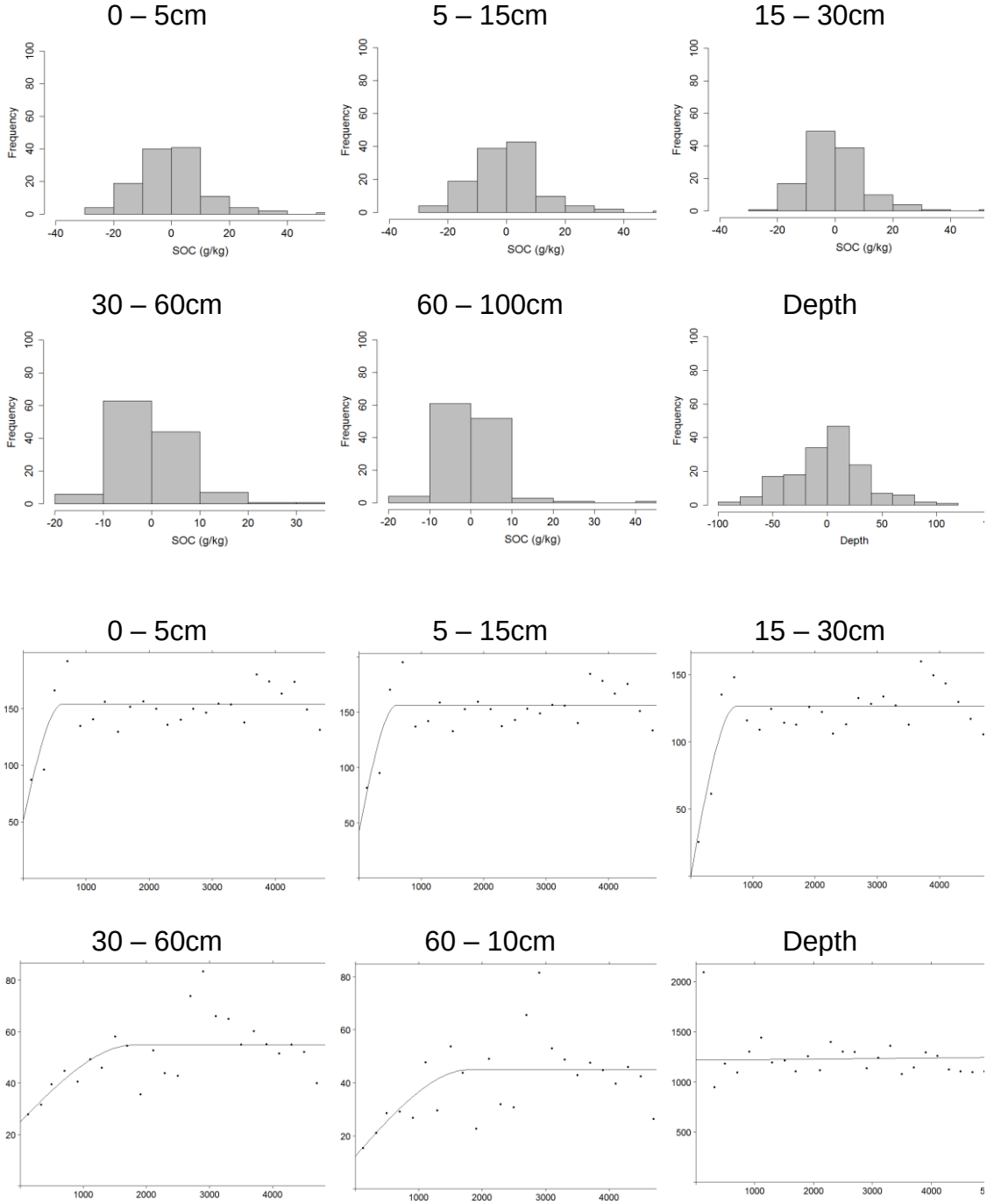


Figure 4. Histograms and variograms of residuals (observed – predicted), from SOC concentration and depth predictions.

Table 5. Descriptive statistics of training, validation, and estimated SOC content

			Training Data	Validation Data	Estimates
			n = 122 points	n = 44 points	n = 312790 pixels
SOC Content	0 – 5cm	Mean	27.5	24.7	31.5
		Median	23.8	22.7	31.1
		Min	4.6	7.0	0
		Max	93.5	61.0	103.8
	5 – 15cm	Mean	27.5	24.8	31.6
		Median	23.5	23.0	31.2
		Min	4.7	7.0	0
		Max	95.3	61.6	108.6
	15 – 30cm	Mean	23.3	21.9	26.5
		Median	19.9	20.7	25.9
		Min	4.6	7.0	0
		Max	90.3	59.6	176
	30 – 60cm	Mean	12.2	10.7	13.1
		Median	10.1	9.4	12.5
		Min	0	0	0
		Max	59.5	40.9	60.1
	60 – 100cm	Mean	7.7	6.4	9.1
		Median	6.1	5.9	8.6
		Min	0	0	0
		Max	59.5	26.2	59.5
			n = 163 points	n = 163 points	n = 312790 pixels
Soil depth		Mean	147.1	156.3	148.7
		Median	150	150	150
		Min	35	25	0
		Max	250	250	250

Table 6. Validation of the Multiple Linear Regression and Regression Kriging for predicting SOC content (g/kg).

Soil	Multiple Linear Regression				Regression Kriging			
Depth	R^2 ŁŁ	RMSE ŁŁ	ME ŁŁ	ccc ŁŁ	R^2 ŁŁ	RMSE ŁŁ	ME ŁŁ	ccc ŁŁ
0 - 5	0.33	13.23	4.89	0.38	0.34	12.82	4.09	0.39
5 - 15	0.33	13.36	4.80	0.38	0.33	13.00	3.97	0.39
15 - 30	0.34	12.37	3.76	0.43	0.35	12.00	3.10	0.45
30 - 60	0.44	6.62	3.04	0.49	0.48	5.80	2.44	0.58
60 - 100	0.34	4.77	2.26	0.41	0.41	4.44	1.87	0.52
Depth ŁŁ ŁŁ	0.43	34.78	0	0.59	-	-	-	-

ŁŁ R^2 = Coefficient of Determination, RMSE = Root Mean Square Error, ME = Mean Error, CCC = Lin's Concordance Correlation Coefficient.

ŁŁ ŁŁ Depth model used the whole data samples, and the validation was made based on leave one cross out.

3.3.3. Spatial predictions and variable importance

In general, SOC levels differed by depth, soil order and by land use type. SOC concentration decreased below 15 cm depth (Table 7 and Fig. 5). The mean values (Table 7) vary between 5.8 g C/kg, from vineyard areas in Alfisols at 60-100 cm depth, to 43.9 g C/kg, from pasture areas in Entisols at 15-30 cm depth. Entisols have the highest mean SOC concentration, 39.1 g C/kg at 5-15cm depth and Alfisols the lowest, 7.1 g C/kg at depth 60-100cm (Table 7). Similarly, forest has the highest mean SOC concentration, 36.1 g C/kg at 5-15cm depth, and arable crops the lowest, 6.7 g C/kg, at depth 60-100cm.

The importance of the variables for SOC concentration prediction differed by depth. The relative importance of the 15 main variables in SOC concentration and the soil depth model is presented in Fig. 5. Up to 30 cm soil depth, the most important variable was Soil Order (Entisols), coordinate X, Aspect and DEM. Below that soil depth the important variables were: Overland Flow Distance to Channel Network, Aspect, Soil Order (Entisols and Oxisols), coordinate Y, and Normalized Height. Overall, the Entisols soil order was a good predictor.

Descriptive statistics of soil depth data and its prediction are shown in Table 5, and the predicted map in Fig. 7. Soils shallower than 70 cm occupied 1% of area (81 ha) and most of them were Entisols (65%) and Mollisols (33%). Soils deeper than 200 cm occupied 5% of area (439 ha) and most of them were Ultisols (54%) and Mollisols (20%). Soils between 70 and 200 cm occupied the

largest area (94%) and most of them were Inceptisols (43%), Ultisols (28%), and Mollisols (16%). The deepest soils were Ultisols (169 cm), followed by Oxisols (158 cm), Inceptisols (150 cm), Alfisols (142 cm), Mollisols (134 cm) and Entisols (110 cm). Soil depth increased in the northern part of the study area and it varied mainly with slope as shallower soils were found on steeper slopes. Pedotransfer function estimated bulk density for all the pedons, with average of 1.17 g/cm3 (0-5 cm), 1.18 g/cm3 (5-15 cm), 1.19 g/cm3 (15-30cm), 1.26 g/cm3 (30-60cm) and 1.27 g/cm3 (60-100 cm). The values ranged between 0.54 – 1.4 g/cm3 (0-5 cm), 0.56 – 1.4 g/cm3 (5-15 cm), 1.19 – 1.4 g/cm3 (15-30 cm), 1.26 – 1.47 g/cm3 (30-60 cm) and 1.27 – 1.47 g/cm3 (60-100 cm).

Table 8 lists pedotransfer functions and validation using root mean square error (RMSE). Based on the lower value of RMSE (0.11), the simplified equation of Benites et al. (2007) - equation (2) in Table 8 - was chosen to extrapolate the bulk density for the whole dataset, producing 163 bulk density estimates. This function was developed from a large compilation of pedons from the Brazilian soil survey databased maintained by EMBRAPA (Empresa Brasileira de Pesquisa Agropecuária) that include pedons in Rio Grande do Sul State (Tornquist, 2009a; Benites et al.,2007).

Table 7. Predicted SOC content (g/kg) by soil order and land use types, from the study area in Vale dos Vinhedos in Rio Grande do Sul, Brazil

Soil Order	Land Use	0 – 5 cm	5 – 15 cm	15 – 30 cm	30 – 60 cm	60 – 100 cm
Alfisol	Arable Crops	26.1 (±1.4)	26.0 (±1.4)	19.3 (±2.2)	10.2 (±2.4)	7.8 (±1.6)
	Fallow	27.2 (±5.1)	26.7 (±5.2)	17.5 (±5.8)	10.5 (±5.2)	7.9 (±4.2)
	Forest	28.1 (±5.4)	27.9 (±5.5)	18.8 (±5.7)	8.7 (±4.8)	8.5 (±4.2)
	Pasture	18.8 (±6.2)	19.0 (±6.1)	18.3 (±5.7)	12.0 (±5.0)	9.2 (±4.6)
	Planted Forest	-	-	-	-	-
	Vineyard	20.5 (±5.3)	20.2 (±5.4)	13.7 (±5.5)	7.9 (±4.5)	5.8 (±3.8)
Entisol	Arable Crops	41.3 (±0.7)	41.7 (±0.7)	36.1 (±0.7)	16.2 (±1.2)	13.0 (±1.0)
	Fallow	37.9 (±8.7)	38.0 (±8.9)	33.1 (±9.1)	17.3 (±5.0)	11.0 (±4.3)
	Forest	43.1 (±8.7)	43.7 (±8.9)	40.9 (±10.4)	16.9 (±6.4)	11.9 (±5.4)
	Pasture	39.5 (±8.2)	40.4 (±8.3)	43.9 (±8.2)	24.7 (±4.5)	15.4 (±3.7)
	Planted Forest	30.8 (±5.3)	31.2 (±5.4)	31.1 (±5.6)	16.8 (±6.9)	9.2 (±5.5)
	Vineyard	31.5 (±6.9)	31.7 (±7.1)	29.1 (±7.8)	13.8 (±5.3)	7.9 (±4.2)
Inceptisol	Arable Crops	29.2 (±4.9)	29.3 (±5.0)	24.8 (±5.5)	13.4 (±4.9)	7.8 (±3.8)
	Fallow	33.6 (±6.3)	33.2 (±6.3)	25.1 (±6.4)	13.8 (±5.7)	8.9 (±5.8)
	Forest	37.0 (±7.0)	37.0 (±7.1)	29.7 (±7.0)	14.2 (±6.1)	11.8 (±6.0)
	Pasture	33.2 (±6.4)	33.7 (±6.4)	34.2 (±6.3)	21.8 (±5.3)	15.3 (±4.6)
	Planted Forest	24.9 (±4.1)	25.0 (±4.2)	23.4 (±4.2)	17.6 (±4.2)	12.0 (±4.3)
	Vineyard	27.4 (±7.0)	27.3 (±7.1)	21.9 (±6.9)	12.2 (±5.6)	7.8 (±4.9)
Mollisol	Arable Crops	27.5 (±9.3)	27.5 (±9.4)	22.0 (±8.4)	10.6 (±4.5)	5.7 (±3.5)
	Fallow	33.7 (±7.4)	33.4 (±7.7)	25.2 (±8.8)	12.9 (±5.0)	6.8 (±4.0)
	Forest	37.0 (±8.5)	37.3 (±8.7)	31.4 (±9.0)	11.1 (±5.3)	7.5 (±3.7)
	Pasture	32.4 (±11.3)	32.6 (±11.5)	31.4 (±10)	17.3 (±3.0)	9.8 (±2.4)
	Planted Forest	19.7 (±3.2)	19.9 (±3.2)	19.8 (±4.1)	12.6 (±3.3)	7.3 (±2.5)
	Vineyard	29.4 (±8.9)	29.3 (±9.1)	23.5 (±9.3)	11.7 (±5.4)	6.4 (±3.8)
Oxisol	Arable Crops	32.1 (±4.8)	32.2 (±5.0)	33.4 (±5.5)	28.8 (±3.6)	9.1 (±2.8)
	Fallow	38.6 (±3.3)	38.3 (±3.5)	34.9 (±4.7)	29.8 (±3.0)	13.5 (±2.2)
	Forest	40.2 (±6.9)	40.1 (±6.9)	37.1 (±6.1)	28.5 (±4.7)	12.2 (±3.6)
	Pasture	34.2 (±4.6)	34.4 (±4.7)	39.9 (±4.9)	34.6 (±3.7)	13.4 (±3.3)
	Planted Forest	-	-	-	-	-
	Vineyard	33.5 (±4.7)	33.5 (±4.7)	33.7 (±4.4)	29.9 (±3.3)	11.8 (±2.7)
Ultisol	Arable Crops	21.6 (±6.7)	21.8 (±6.7)	19.6 (±6.1)	11.9 (±5.1)	5.6 (±3.4)
	Fallow	27.7 (±5.8)	27.5 (±5.9)	22.4 (±6.4)	13.5 (±4.4)	8.6 (±3.8)
	Forest	30.8 (±7.1)	31.0 (±7.3)	26.3 (±7.8)	12.5 (±5.2)	9.8 (±4.6)
	Pasture	25.5 (±6.9)	26.1 (±7.0)	29.0 (±6.6)	19.5 (±3.7)	12.5 (±3.0)
	Planted Forest	17.5 (±3.7)	17.7 (±3.7)	19.5 (±3.6)	16.1 (±3.2)	9.9 (±2.6)
	Vineyard	21.6 (±6.9)	21.7 (±7.1)	19.3 (±7.4)	11.3 (±4.7)	6.8 (±3.9)

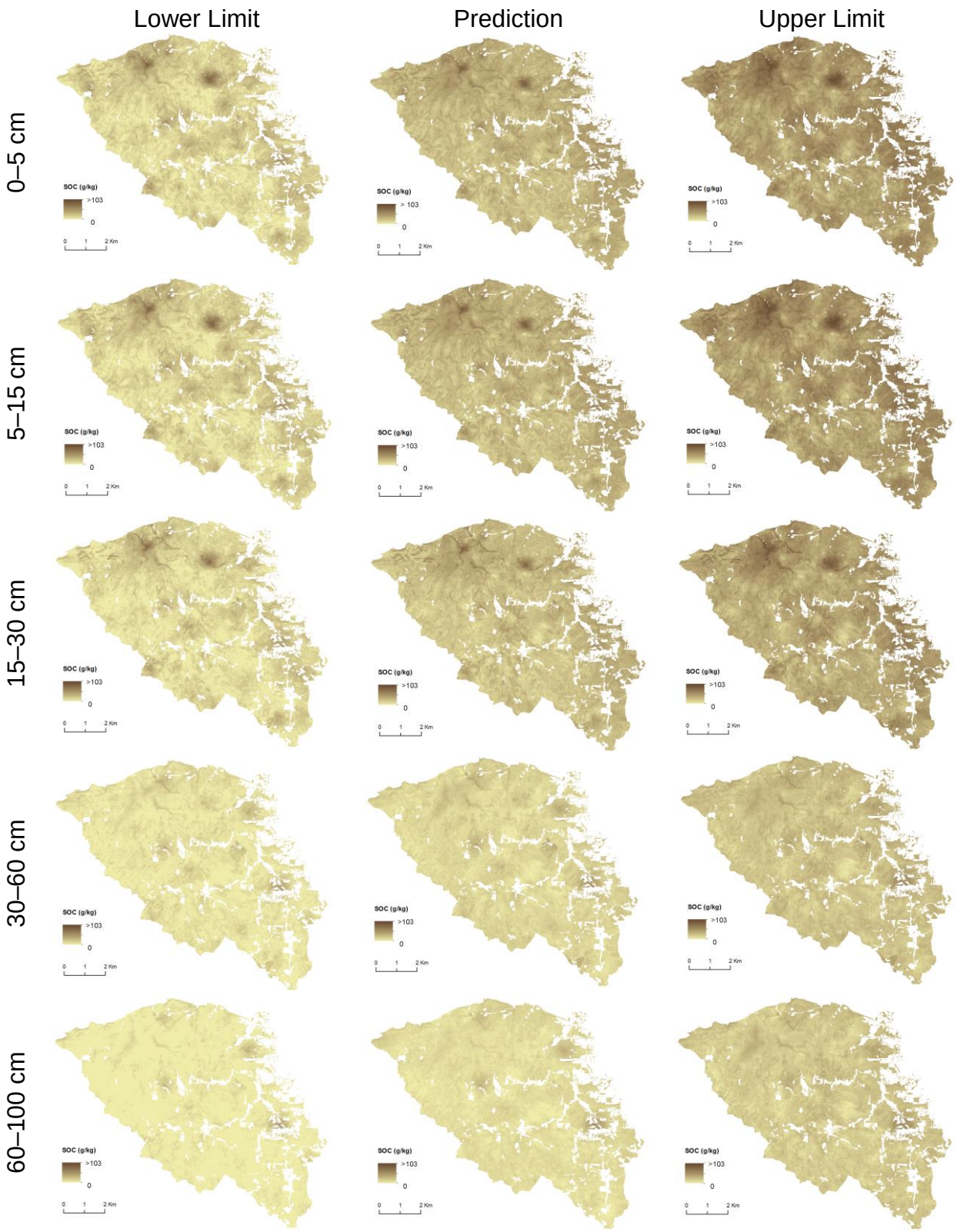


Figure 5. Prediction of SOC content (g/kg) and lower (5%) and upper limit (95%) for five soil depths of Vale dos Vinhedos in Rio Grande do Sul, Brazil.

Table 8. Pedotransfer functions to estimate soil bulk density.

*Pedotranfer functions	Reference	RMSE
(1) $p_m = 1.35 + 0.0045 * sand + 6 * 10^{-5} * (44.7 - sand)^2 + 0.060 * \log depth$ $p_b = \frac{100}{\left(\frac{OM(\%)}{p_{OM}}\right)+\left(\frac{100-OM(\%)}{p_m}\right)}$	Tranter et al (2007)	0.16
(2) $p_b = 1.5688 - 0.0005 * clay - 0.009 * OC$	Benites et al. (2007)	0.11
(3) 30 – 30cm: $p_b = 1.5544 - 0.0004 * clay - 0.01 * OC + 0.0067 * SB$ 30 – 100cm: $p_b = 1.5674 - 0.0005 * clay - 0.006 * OC + 0.0076 * SB$	Benites et al. (2007)	0.13

*p_b = bulk density (g/cm^3); p_m: mineral bulk density (g/cm^3); p_{OM} = organic matter bulk density = 0.224 g/cm^3; $sand$ (dag/kg); $depth$ (cm). OC(g/kg): organic carbon; SB (cmolc/kg)- sum of basic cations (Ca^{2+},Mg^{2+} and K^{+}); $clay$ (g/kg).

Figure 8 shows SOC stock maps for each 5 depth, and the total stock for 0-100 cm. Overall, it appeared that the spatial distribution of SOC stocks was similar to SOC concentration. Values were higher on the valley banks and bottom valley, which were under forest and with reduced agricultural use. Total SOC stocks were highest in Oxisols (230 - 280 t C/ha) and lower in Alfisols (104 - 143 t C/ha), as in Table 9. Soils under pasture areas had the highest SOC stocks (139 - 280 t C/ha) and soils under planted forest areas the lowest SOC stocks (116 - 174 t C/ha). Oxisols under pasture areas had the highest SOC stocks (280 t C/ha) and Alfisols under vineyard the lowest (104 t C/ha).

Table 9. Calculated SOC Stocks (t C/ha) by soil order and land use types for the study area in Vale dos Vinhedos in Rio Grande do Sul, Brazil

Soil Order	Land Use	0 – 5 cm	5 – 15 cm	15 – 30 cm	30 – 60 cm	60 – 100 cm	Total 0-100cm
Alfisol	Arable Crops	13 (±1)	25 (±2)	30 (±3)	40 (±7)	34 (±9)	143 (±18)
	Fallow	12 (±2)	23 (±4)	26 (±7)	39 (±17)	38 (±19)	137 (±46)
	Forest	11 (±3)	23 (±5)	27 (±7)	36 (±14)	39 (±19)	135 (±43)
	Pasture	8 (±3)	16 (±6)	25 (±9)	45 (±15)	45 (±19)	139 (±50)
	Planted Forest	-	-	-	-	-	-
	Vineyard	9 (±2)	17 (±5)	20 (±7)	30 (±14)	29 (±18)	104 (±40)
Entisol	Arable Crops	17 (±1)	34 (±3)	45 (±4)	46 (±10)	54 (±10)	197 (±15)
	Fallow	14 (±4)	28 (±8)	40 (±10)	51 (±16)	37 (±18)	170 (±39)
	Forest	16 (±4)	33 (±8)	51 (±15)	58 (±28)	36 (±20)	194 (±57)
	Pasture	14 (±4)	28 (±8)	49 (±12)	71 (±28)	50 (±23)	212 (±54)
	Planted Forest	12 (±3)	24 (±6)	37 (±9)	47 (±30)	28 (±22)	147 (±60)
	Vineyard	12 (±3)	23 (±6)	35 (±9)	42 (±22)	28 (±18)	140 (±44)
Inceptisol	Arable Crops	13 (±3)	26 (±6)	35 (±9)	47 (±15)	40 (±18)	160 (±44)
	Fallow	14 (±3)	29 (±7)	35 (±9)	48 (±18)	43 (±27)	170 (±55)
	Forest	15 (±4)	31 (±8)	40 (±10)	51 (±19)	52 (±26)	189 (±56)
	Pasture	14 (±4)	29 (±8)	46 (±10)	73 (±18)	72 (±22)	235 (±55)
	Planted Forest	11 (±3)	22 (±5)	32 (±8)	55 (±16)	54 (±18)	174 (±43)
	Vineyard	11 (±3)	23 (±7)	30 (±9)	42 (±17)	38 (±21)	144 (±49)
Mollisol	Arable Crops	11 (±3)	22 (±6)	27 (±9)	29 (±12)	22 (±13)	111 (±38)
	Fallow	12 (±3)	24 (±6)	29 (±9)	42 (±18)	30 (±19)	137 (±44)
	Forest	11 (±4)	22 (±9)	30 (±11)	45 (±17)	36 (±17)	144 (±43)
	Pasture	11 (±3)	23 (±6)	34 (±8)	51 (±12)	36 (±18)	155 (±36)
	Planted Forest	7 (±2)	14 (±5)	21 (±7)	42 (±10)	33 (±11)	116 (±27)
	Vineyard	11 (±4)	21 (±8)	27 (±12)	38 (±18)	27 (±16)	124 (±49)
Oxisol	Arable Crops	16 (±2)	32 (±5)	50 (±8)	95 (±12)	37 (±14)	230 (±38)
	Fallow	19 (±2)	38 (±3)	53 (±7)	100 (±10)	59 (±10)	268 (±30)
	Forest	20 (±3)	40 (±7)	56 (±9)	95 (±15)	53 (±18)	263 (±49)
	Pasture	17 (±3)	33 (±5)	59 (±9)	113 (±14)	58 (±15)	280 (±40)
	Planted Forest	-	-	-	-	-	-
	Vineyard	17 (±2)	33 (±5)	51 (±7)	100 (±11)	51 (±13)	251 (±33)
Ultisol	Arable Crops	10 (±3)	21 (±6)	29 (±8)	43 (±16)	31 (±16)	135 (±44)
	Fallow	13 (±3)	26 (±6)	34 (±9)	50 (±14)	44 (±18)	167 (±44)
	Forest	14 (±3)	28 (±7)	38 (±10)	49 (±16)	47 (±21)	176 (±50)
	Pasture	12 (±3)	24 (±6)	41 (±9)	70 (±13)	64 (±14)	212 (±39)
	Planted Forest	8 (±2)	17 (±4)	28 (±6)	56 (±10)	53 (±13)	162 (±29)
	Vineyard	10 (±3)	20 (±6)	28 (±10)	42 (±16)	36 (±19)	136 (±47)

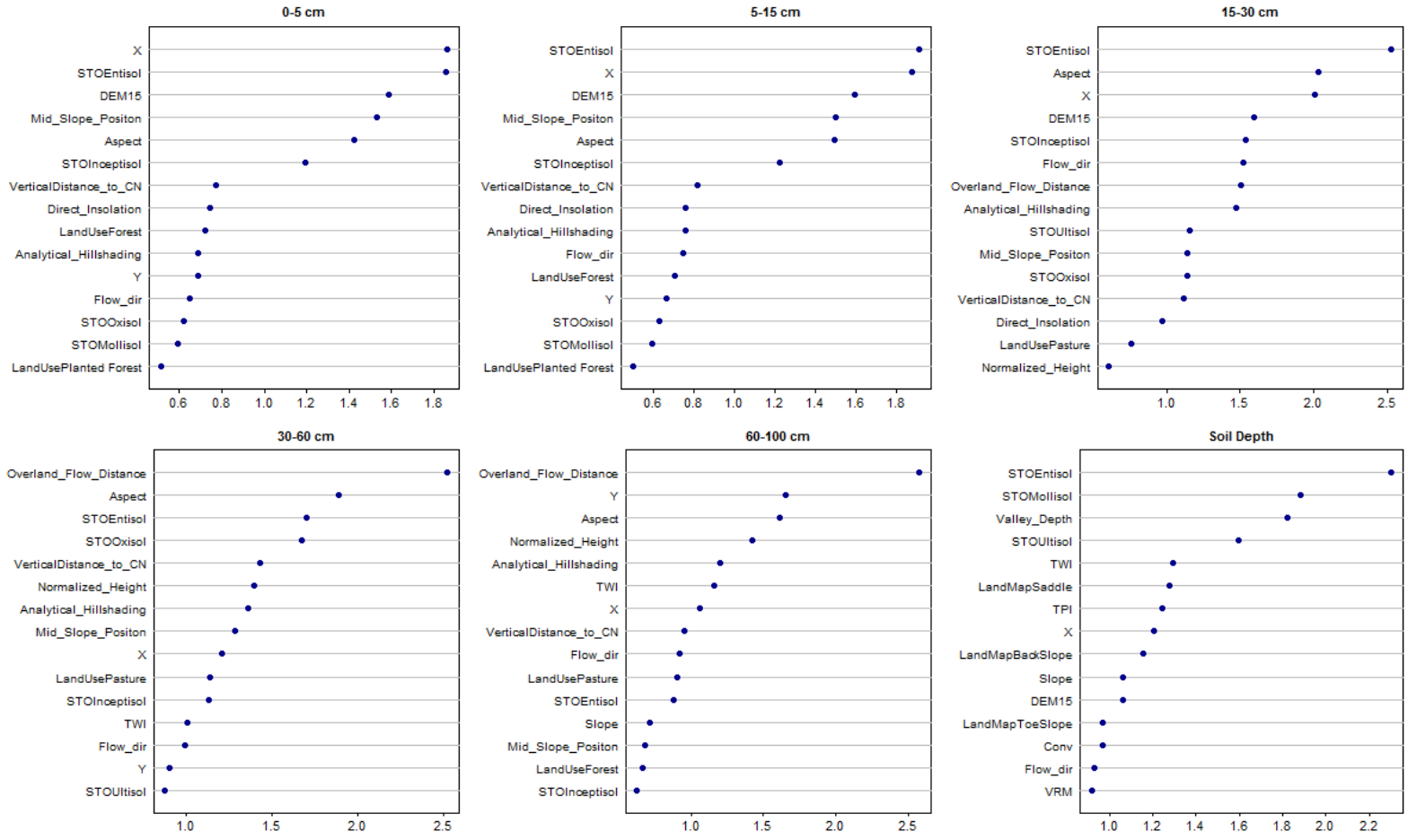


Figure 6. Relative importance of the 15 variables used for predicting SOC content at each soil depth. The importance is calculated based on the absolute value of the t-statistics for each model parameter (see Table 4 for a description of the variables).

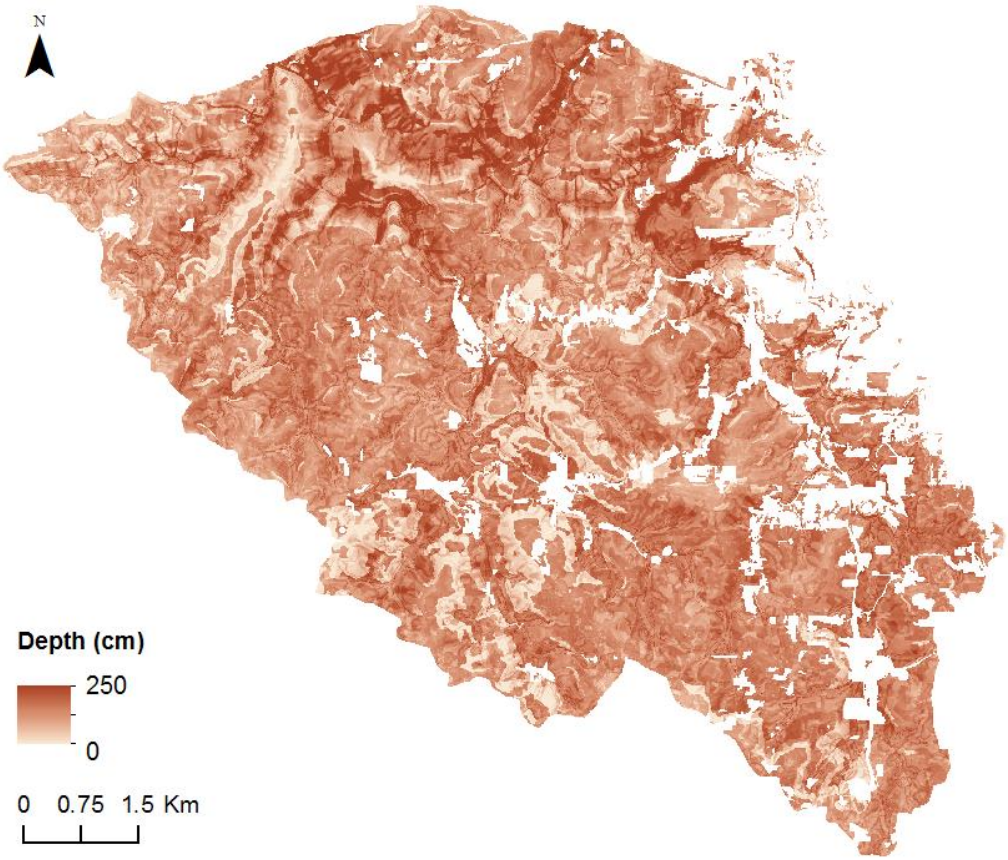


Figure 7. Soil depth map produced using a regression-kriging method, in Vale dos Vinhedos, Rio Grande do Sul State, Brazil.

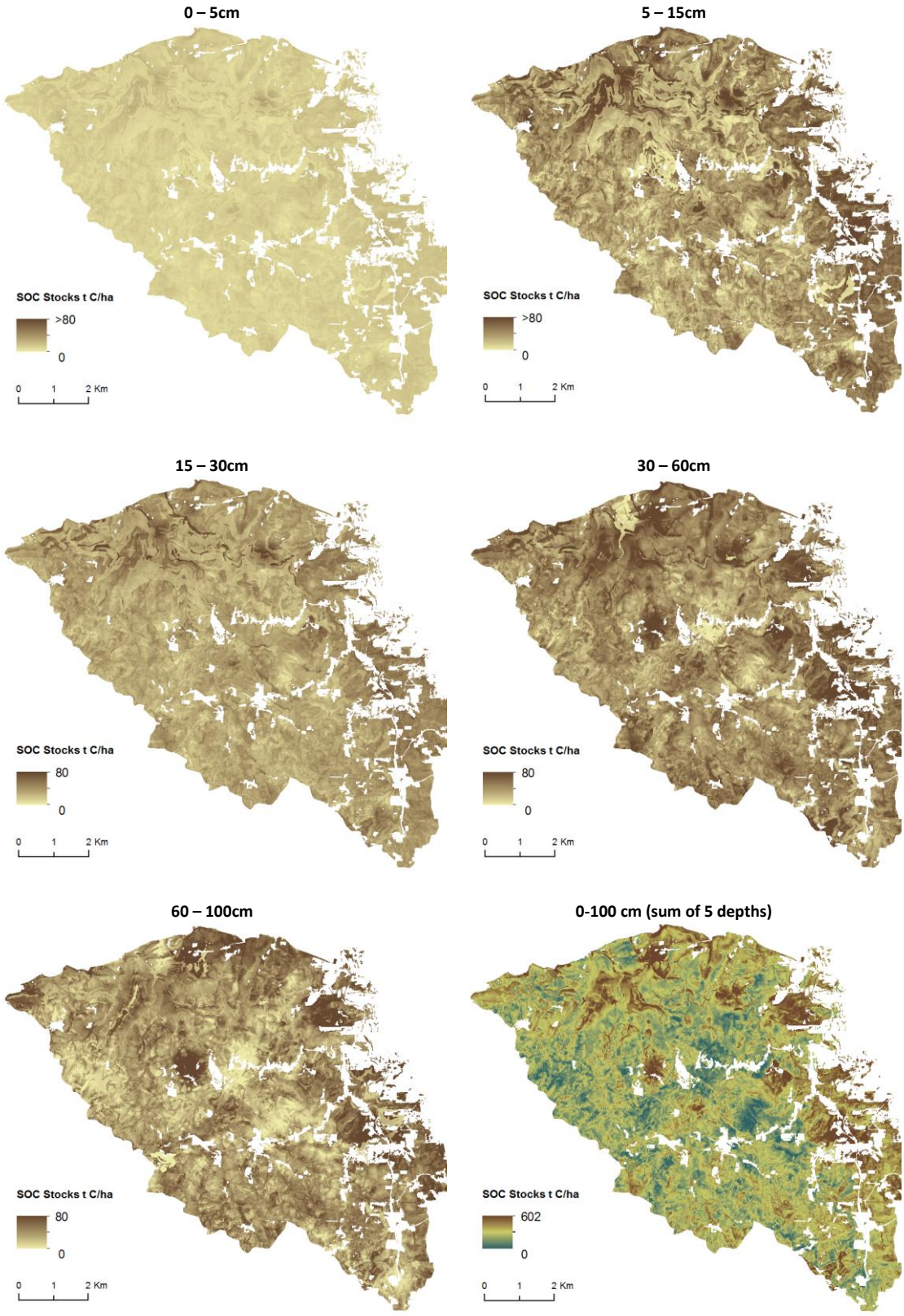


Figure 8. Estimation of SOC stocks (t C/ha) of the study area in Vale dos Vinhedos in Rio Grande do Sul, Brazil.

3.3.4. Uncertainty and probability maps

For uncertainty of SOC concentrations prediction, calculated by empirical approach and mapped (Fig.5.), the averages for the lower limit of prediction decreased from 10.5 g C/kg at 0-5 cm soil depth to 1.5 g C/kg at 60-100 cm depth. The means for upper prediction limit decreased from 53.8 g C/kg at 0-5 cm soil depth to 19.4 g C/kg at 60-100 cm soil depth. A similar trend was found for the difference in the lower and upper limits.

The uncertainty values of nine sample points, from the validation dataset (25% of pedons), are shown in Fig. 9. The blue line represents estimated value, and red lines represent the low (left) and the upper limit (right) for the 5 depths. Bars represent the splined SOC values, harmonized by depth of GlobalSoilMap. The prediction intervals are higher in upper layers than in lower layers. Of the 41 validation samples the following number were within prediction intervals: 36 for 0-5 cm depth, 34 for 5-15 cm depth, 34 for 15-30 cm depth, 38 for 30-60 cm depth, and 39 for 60 – 100 cm soil depth. More than 90% of the validation samples were within the prediction intervals derived from residuals with higher spatial covariance (30-60 cm and 60-100 cm).

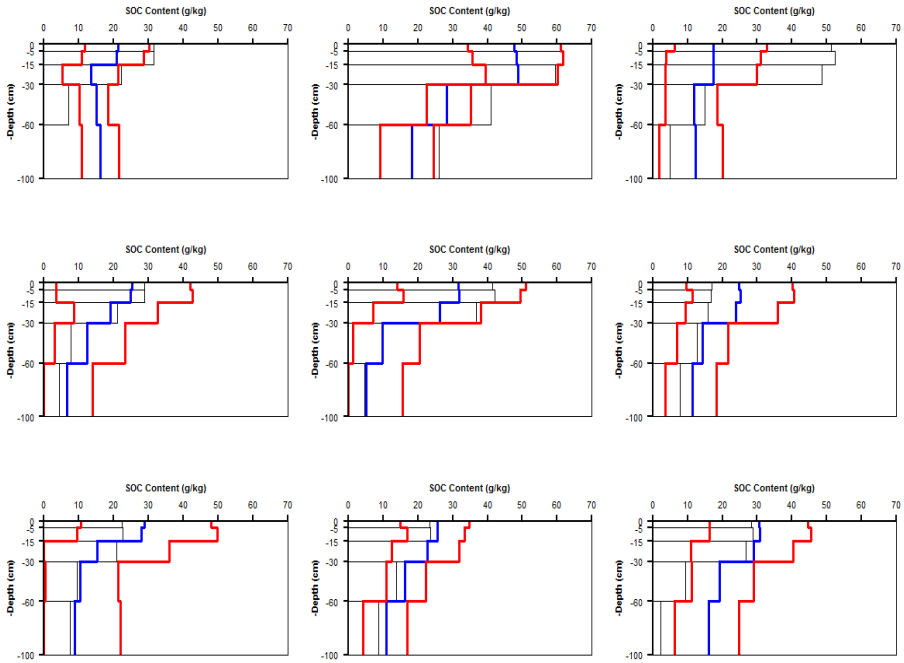


Figure 9. Examples of uncertainty prediction intervals, for SOC levels, of 9 independent validation points. Predicted SOC values are shown in blue, and the lower (5%) and upper limit (95%) in red. Bars represent the values from splines.

For SOC stocks, the probability maps (Fig.10.) show areas where the SOC stock exceeds the threshold value at 20, 40, 60 and 80% probabilities. The probability for SOC stocks exceeding the limit is highest in the valley bottoms and in the eastern part of study area. There is an 80% probability of SOC stocks to exceed 184 t C/ha in about 13% (1029 ha) of the area.

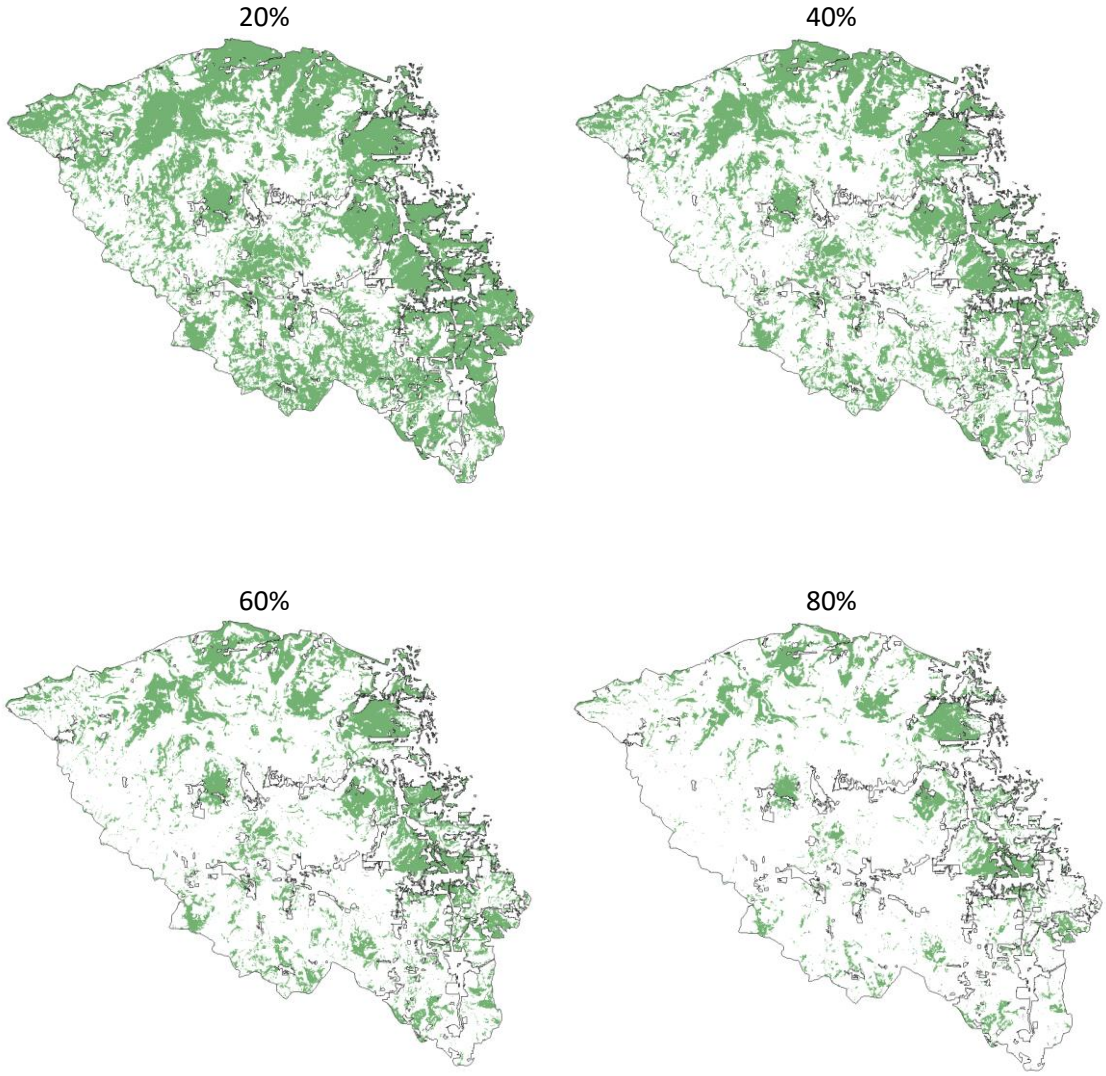


Figure 10. Maps of different probabilities that the soil contain at least 184 ton SOC/ha in Vale dos Vinhedos in Rio Grande do Sul, Brazil

3.3.5. SOC changes

The mean values of SOC predictions and PNVSC values are given in Table 10 where the data were aggregated by soil order and land use. Areas where SOC has been lost as compared to the same soils under forest are given in bold. SOC has been lost at 0-5 and 5-15 cm soil depth for all soil orders and

land use types (except forest which was used as a reference). This loss is also observed at 15-30 cm and 60-100 cm depth, except for Oxisols and pasture. At 30-60 cm soil depth SOC levels has been increased in all soil orders and land use types. The maps of PNVSC and SOC changes are given in Fig. 11.

3.4 Discussion

This study predicted SOC concentration and SOC stocks in a subtropical area under different land use and a range of soil orders. The impact of land use on SOC was evaluated by comparing the SOC concentration under current use with a projected SOC that could be present today if the area was under natural vegetation. In this discussion, we shall focus on the methods of prediction, the effect of the variables used for prediction, predictions of SOC concentrations and stocks, uncertainty and probability maps and the distribution of SOC under different land uses and soil types.

3.4.1. Prediction model

The different methods tested for regression showed that model evaluation is more reliable when using a separate validation dataset. Predicted values might be very similar to observed values, when considering the training model. This model may overfit the data and the performance can be poor using validation data. Minasny and McBratney (2013), observing the behavior of a random forest model, concluded that it can easily overfit the data. In our study, the Cubist and Random Forest models seem to overfit the data, whereas MLR produced estimation closer to validation data.

SOC concentrations were predicted based on Regression Kriging Model (MLR and kriging of residuals). The prediction showed the variation in SOC concentration spatially and by depth, land use and soil order. The model explained only part of the variation and when comparing the estimated mean, median and maximum values, our estimation from the model produced slightly higher SOC concentration than training data. This can be explained by the biased estimate when using a non-probability sample to calibrate the model or also some regions of the feature space to be over or under –represented in the training data.

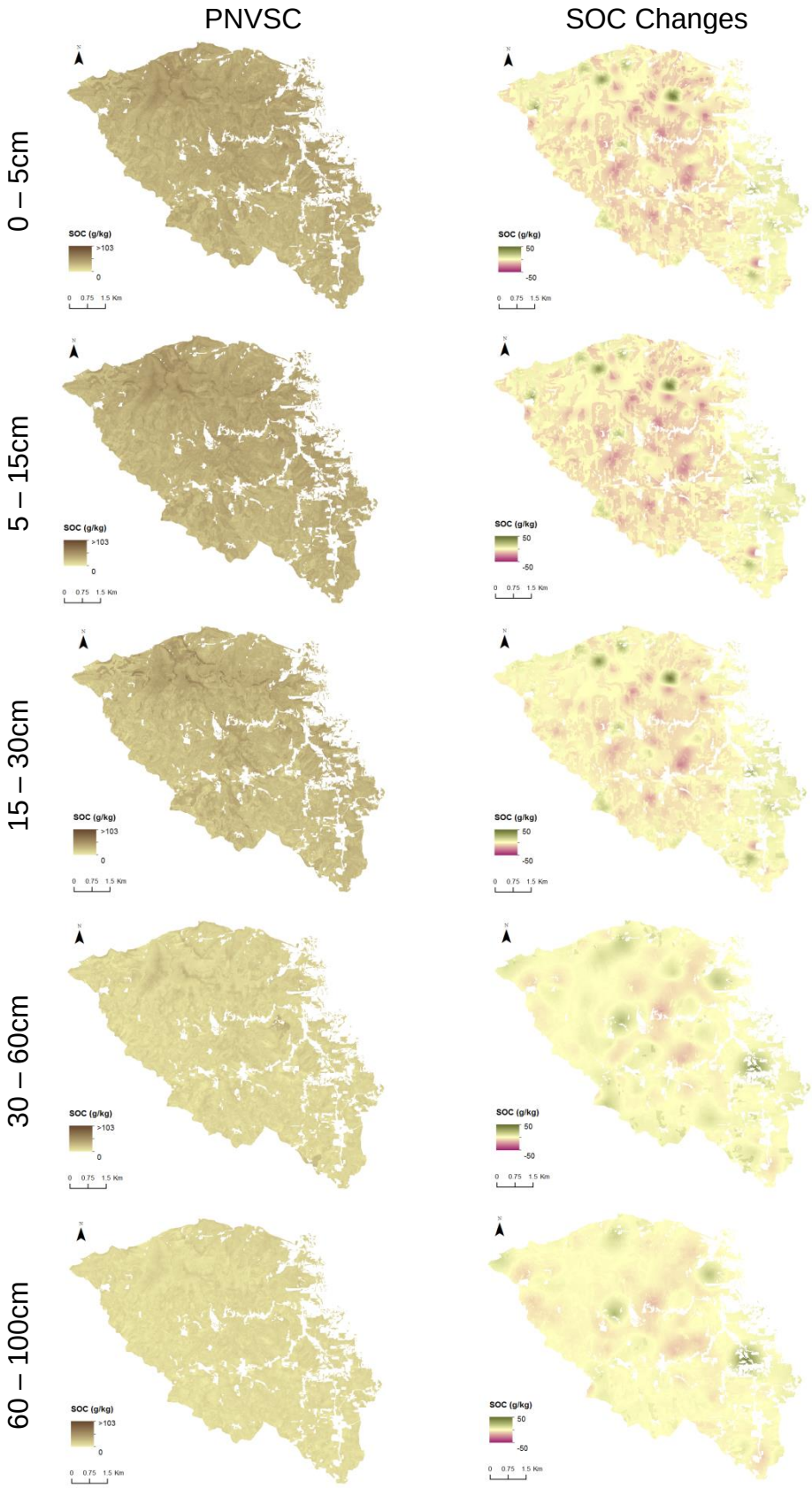


Figure 11. Maps of Projected Natural Vegetation Soil Carbon (PNVSC) and changes in SOC in Vale dos Vinhedos in Rio Grande do Sul, Brazil.

Table 10. Predicted SOC content and Projected Natural Vegetation Soil Carbon (PNVSC), by soil order and land use. Figures in bold indicate that SOC was lost based on the PNVSC approach (SOC content predicted – PNVSC).

	SOC and PNVSC (g/kg) - mean values (± standard deviation)									
Soil Order	0 – 5cm	PNVSC	5 – 15cm	PNVSC	15 – 30cm	PNVSC	30 – 60cm	PNVSC	60 – 100cm	PNVSC
Alfisol	**24.2 (±6.5)**	27.0 (±5.0)	**23.9 (±6.6)**	26.8 (±5.0)	**16.2 (±6.1)**	17.6 (±3.8)	8.4 (±4.7)	7.2 (±2.8)	**7.1 (±4.2)**	7.3 (±2.7)
Entisol	**38.7 (±9.7)**	42.1 (±6.6)	**39.1 (±10.0)**	42.6 (±6.7)	**36.4 (±11)**	38.5 (±5.9)	15.9 (±6.2)	15.5 (±4.8)	**10.5 (±5.3)**	10.8 (±4.0)
Inceptisol	**32.3 (±8.3)**	35.9 (±6.4)	**32.2 (±8.4)**	35.9 (±6.5)	**26.0 (±7.9)**	28.0 (±5.2)	13.5 (±6.0)	12.6 (±4.7)	**9.9 (±5.8)**	10.3 (±4.6)
Mollisol	**35.0 (±9.2)**	36.6 (±6.0)	**35.2 (±9.4)**	36.8 (±6.1)	**29.3 (±9.7)**	29.9 (±5.3)	11.4 (±5.3)	11.0 (±3.5)	**7.3 (±3.8)**	7.8 (±2.6)
Oxisol	**35.1 (±6.1)**	38.2 (±3.1)	**35.1 (±6.1)**	38.1 (±3.1)	35.7 (±5.7)	35.5 (±4.0)	30.3 (±4.4)	26.9 (±3.0)	11.8 (±3.3)	11.7 (±1.8)
Ultisol	**26.4 (±8.1)**	29.7 (±5.8)	**26.5 (±8.2)**	29.9 (±5.8)	**23.0 (±8.1)**	24.8 (±4.9)	12.4 (±5.1)	11.3 (±3.5)	**8.5 (±4.4)**	8.9 (±3.0)
Land Use										
Arable Crops	**25.4 (±7.8)**	29.5 (±4.3)	**25.5 (±7.9)**	29.5 (±4.3)	**23.3 (±8.1)**	24.4 (±3.7)	15.2 (±8.1)	13.9 (±3.1)	**6.7 (±3.7)**	9.1 (±2.0)
Fallow	**30.9 (±7.2)**	32.2 (±3.9)	**30.7 (±7.3)**	32.3 (±3.9)	**24.2 (±7.4)**	25.9 (±3.5)	13.7 (±5.1)	11.8 (±3.1)	**8.6 (±4.7)**	9.3 (±3.0)
Forest	35.9 (±6.5)	35.9 (±6.5)	36.1 (±7.0)	36.1 (±7.0)	30.2 (±7.2)	30.2 (±7.2)	13.1 (±5.2)	13.1 (±5.2)	10.1 (±4.7)	10.1 (±4.7)
Pasture	**30.1 (±8.5)**	32.8 (±4.2)	**30.5 (±8.6)**	32.8 (±4.2)	32.3 (±8.1)	26.3 (±3.8)	21.1 (±5.8)	12.7 (±2.8)	13.3 (±4.1)	9.5 (±2.1)
Planted Forest	**23.8 (±6.0)**	36.6 (±2.5)	**23.9 (±6.1)**	36.6 (±2.5)	**23.5 (±5.6)**	29.5 (±2.4)	16.9 (±4.6)	14.2 (±2.6)	**10.8 (±4.3)**	11.7 (±2.6)
Vineyard	**26.0 (±7.9)**	33.9 (±4.8)	**26.0 (±8.0)**	33.9 (±4.9)	**21.8 (±8.0)**	26.6 (±4.5)	12.0 (±5.5)	11.4 (±3.9)	**7.4 (±4.4)**	8.9 (±3.3)

The values of validation parameters such as R^2 and CCC were higher for 30-60 cm soil depth and were lower at other depth intervals. The validation results, on Table 6, are comparable to most recent studies predicting SOC. For example, on temperate areas, Adhikari et al. (2014) found the model could explain 43% of variation in validation data, whereas Malone et al. (2009) found R^2 values of validation points ranging between 20% and 27%. Other studies present similar results (Brogniez et al., 2014; Collard et al., 2014; Forges et al., 2014; Wiesmeier et al., 2014).

The soil depth model could explain 43% of the variation, using all data for estimation and validation. The soil depth map followed the topographic variation, showing the deeper soils in valley bottom. The calculated bulk density varied between 0.54 and 1.47 g/cm^3 and were similar to the values found by Tornquist et al. (2009a), between 0.4 and 1.4 g/cm^3.

3.4.2. Importance of predictor variables

The relative importance of each variable was evaluated by absolute t-values. The t-value is model dependent, which means that if two or more variables are correlated with SOC concentration, and also correlated with each other, then only one of these intercorrelated variables may appear with a high t-value.

We noted that variable importance differed by soil depth. Up to 30 cm soil depth, the covariates Soil Order, coordinate X, Aspect and DEM were good predictors. Soil Order (Entisols) contributed mainly due to its consistent higher SOC values (Table 7). There was a decrease in SOC concentration towards the west (Fig. 5) and hence coordinate X was important to identify this variation in east-west direction. There was a higher SOC concentration in the soils of the north in the valley bottom, and coordinate Y identifies this variation. In correlation analysis, it was noted that Y has a correlation of -0.47 with X, which is fairly high compared to other covariates. Both coordinates could explain the spatial variation of SOC concentration, although only X showed high t-value, and to separate individual effect in prediction is not straightforward (Hair et al., 2009). The north-facing slopes receive more solar radiation, and as a result possibly enhanced SOC decomposition and lower SOC levels. This effect can be seen at slightly

higher SOC values in the northeast (slope south-facing) compared to the southwest (slope north-facing). This variation could be identified by the Aspect covariate.

At lower elevation (200 – 350 m in the study area), temperatures increase and likely the soils contain less carbon due to higher rates of decomposition. However, the elevation was a proxy for deposited material and areas at lower elevation had deeper soils with more SOC. There was a relatively high and negative correlation (-0.55) between DEM and Valley Depth. Only the covariate DEM is showed with high t-value, but both explain the SOC variation related with elevation.

For the layers below 30 cm soil depth the covariates Overland Flow Distance, Aspect, Soil Order, coordinate Y, and Normalized Height were important predictors for SOC concentration. Overland flow distance to channel network indicate that the SOC concentration is higher closer to channel network, possibly because of organic material deposits under dense vegetation. Libohova et al. (2014) found that areas with water accumulation for longer time periods stored 50-68% more total SOC compared to drier areas. Noticeable influence of soil orders covariates (Entisol, Oxisol or Inceptisol) in SOC prediction was found up to 60 cm depth but not below this depth. The coordinate Y is consistent with the valley bottom in north direction. The Normalized Height indicates the height relatively to the highest and lowest position within an area (Dietrich and Böhner, 2008) and this covariate correlates with Overland Flow Distance to Channel Network (0.63).

For prediction of soil depth, Soil Order (Entisol, Mollisol and Ultisol) and Valley Depth proved good predictors. The Entisols are shallower soils (mean depth 110 cm) and Mollisols and Ultisols are the deeper soils. Although Oxisols are also deep soils, it had no significant impact on the soil depth model possibly because of the limited number of samples.

3.4.3. SOC concentration and stocks

The SOC concentration predicted for soils under forest and pasture differed by depth. In the upper layers, soils under forest had higher values whereas soils under pasture had more SOC with depth (Table 7 and Table 10).

Forest has larger amounts of litter and organic material, which is incorporated into the soil. Aboveground input and relatively low rates of decomposition generally increases topsoil SOC levels compared to grasslands (Don et al., 2011; Guo and Gifford, 2002; Jobbágy and Jackson, 2000). For pasture, deep roots contribute to the accumulation of SOC with depth (Guo and Gifford, 2002).

Soils under arable crops and vineyard had the lowest SOC concentration and stocks as a result of reduced organic matter input and enhanced decomposition (Elliott, 1986; Sanford, 2014; Schrumpf et al., 2013), but SOC levels could improve with careful soil management (Lal, 2006). Soil erosion may decrease SOC stocks in agricultural systems (Don et al., 2011) whereas leaving the land fallow may increase SOC levels depending on the length of the fallow. Hartemink (1998) found, in Papua New Guinea, that SOC concentration changed from 51 g C/kg to 36 g C/kg after 17 years of sugarcane cultivation.

Planted forests in Vale dos Vinhedos are mostly pinus or eucalyptus, and the soils generally had a low SOC concentration and SOC stocks. It is known that coniferous and broadleaf trees can have different carbon accumulation (Guo and Gifford, 2002) but we were not able to distinguish these forest types. Planted broadleaf trees accumulate SOC levels comparable to natural forests. SOC stocks under plantation forest could be restored to the original level under native forest, but it may requires several decades (Guo and Gifford, 2002; O'Brien and Jastrow, 2013). As planted forests are harvested there may be considerable soil erosion and loss of topsoil carbon (Hartemink, 2003).

The SOC concentration and stocks differed by soil order. Until 30 cm soil depth, Entisols have a higher SOC concentration but with depth Oxisols have the highest SOC concentration. Most Entisols (58%) are under forest which explains some of the higher SOC concentrations. Oxisols are deeper soils and have possibility of long-term accumulation of SOC with depth. Many of the Oxisols are under pasture (16%), whereas other soil orders have less than 3% of their area under pasture. Pasture has generally higher SOC accumulation with depth. Alfisols are mostly under vineyard which can explains their lower SOC levels. About two-third of the Mollisols are under forest, accumulating more SOC in upper layers. Most of the Inceptisols are under forest (39%) and vineyard (35%).

SOC stocks were calculated and corrected based on equivalent soil mass (Gifford and Roderick, 2003; Lee et al., 2009; Ellert and Bettany, 1995). We found corrected SOC stocks varying from 104 t C/ha in vineyards in Alfisols to 280 t C/ha in pasture areas in Oxisols, with an average of 161 t C/ha. Results of SOC stocks for 100 cm depth (Table 9, Fig. 7) are comparable to other studies (Table 11). This can be attributed to the relatively high SOC concentrations. About 16% of the SOC concentration values between 60 and 100 cm depth exceeded 10 g C/kg, and considering 40 cm thickness it explains the relative high SOC stocks with depths. Environmental conditions in the study area favor SOC accumulation, due the high precipitation and relatively low temperature. SOC stocks average for soils under arable crops is 163 t C/ha, for fallow is 175 t C/ha, for pasture is 205 t C/ha, vineyard is 150 t C/ha, for planted forest is 149 t C/ha and 184 t C/ha for soils under forest. Other studies in Brazil found similar values such the studies by Boddey et al. (2010), Sisti et al. (2004) and Jantalia et al. (2007) (Table 11).

Tornquist et al. (2009a) found in Rio Grande do Sul State, for SOC stocks to 30 cm soil depth of non-sandy and non-wet soils, mean values of 77 t C/ha for Alfisols, 66 t C/ha for Entisols, 83 t C/ha for Inceptisols, 76 t C/ha for Mollisols, 77 t C/ha for Oxisols and 48 t C/ha for Ultisols. These stocks are comparable to the current study in Vale dos Vinhedos, based on equivalent soil mass, of 57 t C/ha for Alfisols, 85 t C/ha for Entisols, 76 t C/ha for Inceptisols, 60 t C/ha for Mollisols, 106 t C/ha for Oxisols and 67 t C/ha for Ultisols. Bernoux et al. (2002) found for non-sandy or non-wet soils, in areas with mixed forest, SOC stocks (0-30 cm) between 61 to 128 t C/ha. These values are comparable to 84 t C/ha found in forest areas in Vale dos Vinhedos at the same depth based on the equivalent soil mass. Wasige et al (2014), studying SOC in Rwanda until 50 cm depth, found under forest, stocks ranged 295 t C/ha in Cambisols to 487 t C/ha in Histosols. These values were not corrected by mass and are higher than the 115 t C/ha found in Vale dos Vinhedos (Table 9), for forest until 60 cm depth, corrected by mass. For agriculture areas (main crops are tea, coffee, maize and banana), the values were between 114 t C/ha in Acrisols (Ultisols) to 169 t C/ha in Ferralsols (Oxisols). These SOC stocks are similar to found in arable crops (126.3 t C/ha) and vineyard (115 t C/ha) areas in our study up to 60 cm soil depth.

Table 11. SOC stocks (t/ha) under different land use and in different soils.

Location	Land Use, Soil Type	Depth	SOC stocks (t C/ha)	Reference
Brazil - Distrito Federal	Tillage – 6 treatments	100 cm	171	Jantalia et al. (2007)
Brazil - Rio Grande do Sul	3 different crop rotation in: Zero till Conventional tillage	100 cm	 175.2 163.8	Sisti et al. (2004)
Brazil - Rio Grande do Sul	Rotations with intercropped or cover-crop legumes in: Zero Till Conventional tillage	100 cm	 154 - 172 132 - 163	Boddey et al. (2010)
Brazil – Rio Grande do Sul	Alfisols Entisols Inceptisols Mollisols Oxisols Ultisols	30 cm	77 66 83 76 77 48	Tornquist et al. (2009a)
Brazil	Mixed Ombrophyllous forest	30 cm	61 - 128	Bernoux et al. (2002)
Brazil - Amazon	Forest on Arenosol Forest on Histosol	100 cm	40 724	Batjes and Dijkshoorn (1999)
Spain – Canalda river basin	Cropland (mainly cereals and potatoes) Forest Grazing	100 cm	63 116 89	Simó et al. (2014)
USA	-	100 cm	345	Wills et al. (2014)
USA	Forest Pasture Crops (82 row crops)	100 cm	76.8 74.9 107	Bliss et al. (2014)
China	Forestland Grassland Farmland	100 cm	143.3 82.4 92.2	Yu et al. (2007)
Rwanda – Rukarara river catchment	Forest Agriculture (tea, coffee, maize and banana)	50 cm	295 - 487 114 - 169	Wasige et al. (2014)

3.4.4 Uncertainty and probability maps

It was found that 88% of validation values are within the prediction intervals. Malone et al. (2011) found similar result using an empirical uncertainty method based in distribution of prediction errors. However, for depth with higher spatial covariance of residuals (30-60 cm and 60 – 100 cm) more than 90% of the values were within the prediction intervals. Our results suggest that the methodology adopted to calculate uncertainty depends of the spatial covariance of the residuals. The limited accuracy may be related to variation in environmental

conditions between the training and validation data, lower spatial relation found in the most of interval depths, and errors in measures of the training or validation samples.

The SOC stocks probability maps (Fig. 10) reflect SOC stocks exceeding 184 t C/ha. Such areas are found in valley bottoms due the sediment accumulation and reduced drainage, and in Entisols, Mollisols under forest because of higher production of organic material and lower rates of decomposition. The low probability values are mostly in soils under vineyard or arable crops (mainly Inceptisols and Ultisols). The maps show that only about 13% of the area has 80% of probability for exceeding the 184 t C/ha. These areas may have the same or more SOC than the soils under original land use. The 20% probability map shows that non-colored areas have 80% of probability to be able to stock more SOC. About 42% of soils of study area (3,374 ha) could sequester more carbon if occupied by natural forest.

3.4.5. SOC changes

The PNVSC analysis showed that the topsoils could accumulate more SOC if they were under forest (Table 10) because of increased organic material addition and reduced decomposition. Below 15 cm depth, soils under pasture have a higher capacity to accumulate SOC which is commonly found (Guo and Gifford, 2002; Lacoste et al., 2014; Nieder and Benbi, 2008.At interval depth 30-60 cm, regardless of soil type or land use, the soil accumulates more carbon than if the soil was under forest. A possible explanation is that there is storage in carbon in that depth after carbon being translocated from upper layers.

3.5 Conclusions

From this research the following can be concluded:

- The concordance of observed SOC values and SOC content estimated using Multiple Linear Regression was higher than the results estimated by Stepwise Multiple Linear Regression, Random Forest and Cubist.
- Forest accumulates more carbon in upper layers and pasture accumulates more carbon with depth.

- Oxisols and Entisols accumulate larger contents of SOC. Lower values for SOC were found in Alfisols, Ultisols, arable crops, vineyard and planted forest.

- The SOC stocks (down to 100 cm) were on average 166 t C/ha but varied between 107 t C/ha in vineyards on Alfisols, to 324 t C/ha in fallow areas on Oxisols.

- The uncertainty intervals had better estimations where regression residuals showed higher spatial dependency (30-60 cm and 60-100 cm depth).

- The PNVSC was able to estimate SOC changes due to land use change. The analysis showed that estimated SOC in current land use is less than the projected SOC under natural vegetation.

4. CAPITULO III – ESTUDO 2: A MECHANISTIC MODEL TO PREDICT SOIL THICKNESS IN A VALLEY AREA OF RIO GRANDE DO SUL, BRAZIL

4.1 Introduction

Physically-based models of landscape-scale erosion and deposition have largely advanced in the last decades. Landscape Evolution Models (LEMs) can simulate elevation changes based on differences caused by erosion and deposition, and have been validated in a wide range of environments. LEMs have been widely used in the Earth Sciences (Willgoose, 2005) as an experimental tool to investigate processes of landscape evolution. LEMs can also be used to study the spatial distribution of soils and vegetation (Saco et al., 2007).

Soil properties that control landscape evolution dynamics and the spatial variability and magnitude of erosion include soil texture, soil organic carbon content or surface stone cover (Minasny et al., 2015). Conversely, landscape evolution influences soil development, as erosion or deposition changes the thickness across the landscape. In an eroding landscape, the bedrock-saprolite contact is closer to the surface accelerating soil production, which deepens the bedrock-saprolite contact. The accelerated soil production could be due to the action of bioturbation or by the uprooting of bedrock material (e.g. Phillips and Marion, 2006), due to more intense chemical weathering as surface horizons are flushed by infiltrating rainwater (Maher, 2010). It could also be due to more active physical weathering as a result of frost cracking (Anderson et al., 2013).

Few models have integrated landscape and soil formation and most studies work with a hypothetical landscape and validation of soil-landscape models with limited field data (Minasny et al., 2015). Saco et al. (2006) used the

SIBERIA model, combined with a soil production rate, to evaluate the use of spatially varied soil moisture. Results were comparable to Heimsath et al. (1997) who showed an exponential decline in soil production rate with soil thickness. The LORICA model (Temme and Vanwalleghem, 2015) was built based on the landscape evolution LAPSUS and the soil formation MILESD (Vanwalleghem et al., 2013); it demonstrated soil landscape interactions, but the model was not validated with field data. Vanwalleghem et al. (2013) used MILESD on a test area within the Werrikimble National Park in NSW, Australia. The model included physical and chemical weathering, clay migration, neoformation, bioturbation and carbon cycling. The results showed the importance of the soil-forming processes interacting with erosion and deposition. The model predicted trends in total soil thickness along a catena, which were comparable to field observations. Soil thickness, texture and bulk density were predicted with errors in the order of 10%.

An important assumption in these studies is that soil production and transport are in steady-state (Heimsath et al., 1997). Heimsath et al. (2001) notes that deviations from such steady-state lead to an incorrect modelling of the exposure history. Therefore, most studies have been carried out in diffusion-dominated convex ridges (noses), hereby avoiding places dominated by landslides or other perturbations such as tree falls. Little or no research has been carried out in complex landscapes that include actively eroding areas and bottom valleys, influenced by sediment deposition. In lower-lying areas soil thickness increases and soil production is expected to be low according to the assumed exponential dependency between soil production and depth.

The objectives of this study were to assess the potential of a combined soil and landscape evolution model to (i) predict soil thickness and (ii) to estimate the trends of soil thickness variation over time, under different landscape positons. Different scenarios were developed and tested and the results were validated using measured soil thickness data in the region of Vale dos Vinhedos in Rio Grande do Sul State, Brazil. The study was conducted in two steps. Firstly, we evaluated four different combined soil production – landscape evolution models to predict soil thickness, considering soil wetness varying spatially using a topographic wetness index. In the second step, soil thickness evolution was analyzed, by using the predicted soil thickness as input in a LEM.

4.2 Materials and methods

4.2.1 Study area and soil data

The study was conducted in the Vale dos Vinhedos (Vineyard Valley) which is a wine production region in northeastern Rio Grande do Sul State (Fig. 12). The area covers 8,118 ha (29º08'15"S to 29º14'26"S, and 51º29'48"W to 51º37'55"W). The climate is classified as Cfb: subtropical with a mild summer, mean annual temperatures of 17.2°C and 1,736 mm annual rainfall (Embrapa, 2017). The dominant lithology is effusive rocks mostly from the Mesozoic Era (IBGE, 1986).

The geology is part of Bacia do Paraná, Formação Serra Geral, and is divided in two units: Unit of Gramado, in lowlands and Unit of Caxias, in uplands. The Unit of Gramado has basalt as a predominant parent material. Rhyodacite is predominant in the Unit of Caxias. The rocks are from Cretaceous, with approximately 132 Myr, and resulted from a succession of volcanic flows.

The topography is formed by steep and jagged edges, by a drainage system with high capacity of vertical erosion. Some upland areas are preserved and are a testimony of older geology. The relief was carved by the drainage system, sectioning the sequential volcanic flows, and forming stepped structural terraces.

The geology map available is at a scale 1:750,000 (CPRM, 2006) and the DEM had 5mx5m resolution. The DEM was upscaled to 15m grid cell size. A land use map (Bonfatti et al., 2016) in scale 1:10,000 are available. A map with 13 landform classes was elaborated using the DEM and the software LandMapR (MacMillan, 2003). The soil database consists of 163 pedons, with 32 soil properties and elemental concentrations (Flores et al., 2012).

The average soil thickness (depth to bedrock) is 150 cm (range 25 to > 250 cm) and many soils are stony and rocky (average 4.5% of fragments > 20 mm in diameter). In the study area, Inceptisols cover about 44%, Ultisols 29% and Mollisols almost 15%. Mollisols are mostly present at lower elevations close to valley bottoms in the northern part of the study area. Soils in the western part of the study area, classified using the Brazilian Soil Classification System (SiBCS), are mainly Argissolos (Ultisols and Alfisols), Chernossolos (Mollisols),

and Neossolos (Entisols and Mollisols). The eastern part has more rugged terrain and the dominant soils are Neossolos (Entisols) and Cambissolos (Inceptisols), with association of Argissolos (Ultisols and Alfisols), Latossolos (Oxisols) and Nitossolos (Oxisols and Ultisols) (Flores et al., 2012).

Forest (44%) and Vineyard (31%) are the dominant land use in the study area. Deciduous forest is the main vegetation in plateau rugged areas, and Araucaria forest in flatter areas (IBGE, 1986). A soil thickness map was produced by regression-kriging (Bonfatti et al., 2016) (Fig. 12).

Digital Elevation Model and pedon samples (Flores et al., 2012)

Soil Depth Map elaborated by regression-kriging (Bonfatti et al., 2016)

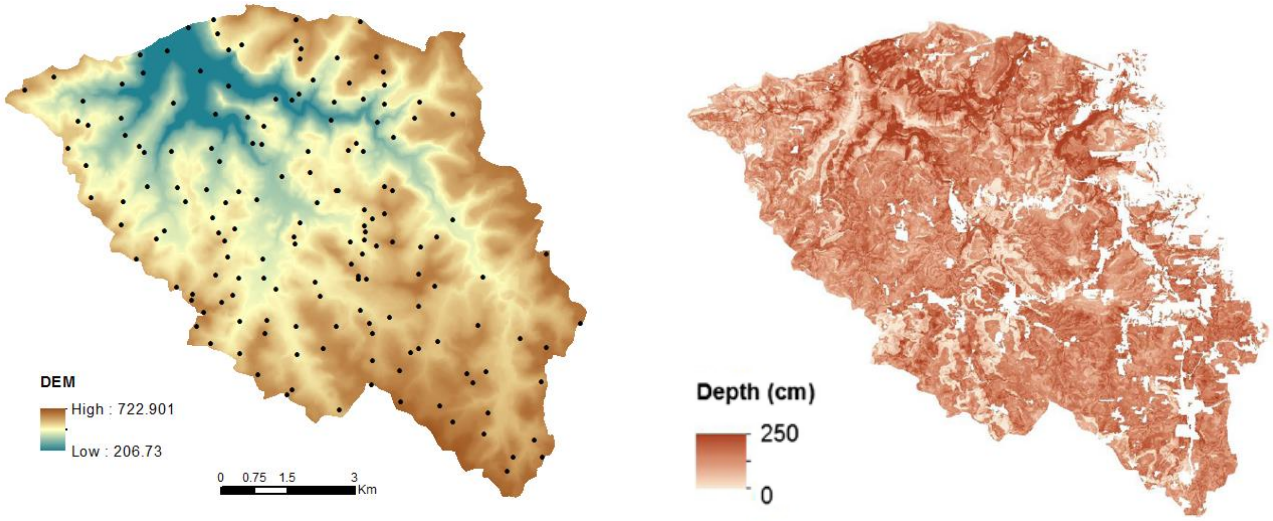


Figure 12. Study area (Vale dos Vinhedos) in Rio Grande do Sul, Brazil (8118 ha). Location of 163 pedons in a DEM (Flores et al., 2012) and a soil depth map produced by regression kriging (Bonfatti et al., 2016).

4.2.2 The mass continuity equation

The basis of LEM is the mass continuity equation (Carson and Kirkby, 1972; Dietrich et al., 1995; Heimsath et al., 2001b):

$$\rho_s \frac{\partial h}{\partial t} = \rho_r \frac{\partial e}{\partial t} - \nabla q_s \qquad (1)$$

Where h is the soil thickness, e is the elevation of the bedrock-soil interface, ρ_s is the soil density, ρ_r is the density of the rock, q_s is the flow of material and ∇ is the vector of partial derivatives.

In this equation, the variation in mass of soil thickness depends on the balance between the mass of soil produced (variation of bedrock elevation) minus the mass lost by erosion (or summed by deposition). Dividing by ρ_s, we adjusted the equation from mass to depth continuity (m):

$$\frac{\partial h}{\partial t} = \left(\frac{\rho_r}{\rho_s}\right)\frac{\partial e}{\partial t} - \frac{\nabla q_s}{\rho_s} \qquad (2)$$

The parameter $\partial e/\partial t$ is equivalent to the soil production rate and the q_s is the erosion/deposition rate. If soil production is higher than erosion, the soil thickness h will increase over time, otherwise, the soil thickness will reduce. There is a feedback mechanism between erosion and soil production. The erosion will approximate the weathering front to the surface, accelerating the soil production. When the soil production rate is similar to the erosion rate, the system approaches a steady-state, reaching a dynamic equilibrium (Pelletier and Rasmussen, 2009; Phillips, 2010). If we consider a stable topography, soil thickness will not change over time, thus $\partial h/\partial t = 0$. The steady-state soil thickness is a practical assumption to predict current soil thickness and used in several models (Heimsath et al., 1997; Nicótina et al., 2011; Saco et al., 2006). Actually, topography changes continuously and soil erosion differs at each condition. Thereby, if topography changes, the soil thickness will change over time.

4.2.3 Landscape Evolution Model (LEM) and the Soil Production Function (SPF)

The LEM used physical equations to calculate the sediment added or eroded to the entire catchment (Table 12). The sediment transport (right-hand term in equation 2) was computing by the equation 3 in inter-rills and equation 7 in concentrated flow areas (Table 12), based on approach of Willgoose & Rodriguez-Iturbe (1991). Equation 3 is the continuity equation for sediment transport and used parameters obtained by equations 4 to 6 (Table 12). Equation 4 represents the fluvial sediment transport, dependent on discharge q and slope S in the steepest downhill direction.

Table 12. Equations used to model landscape evolution and soil production functions.

Equations	Description	Reference
$\frac{\partial z}{\partial t} = \frac{1}{\rho_s}\left(\frac{\partial q_{sx}}{\partial x} + \frac{\partial q_{sy}}{\partial y}\right)$ (3)	Landscape evolution	Adapted from Willgoose & Rodriguez-Iturbe (1991) and Saco et al. (2006)
$q_s = \beta_1 q^{m_1} S^{n1}$ (4)		
$q = \frac{Q}{w}$ (5)		
$Q = \beta_2 A^{m_2}$ (6)		
$\frac{\partial z}{\partial t} = -\in Q^u S^v$ (7)	Stream incision law	Chen et al. (2014)
$SPF = P_0 e^{-mH}$ (8)	Soil Production Function	Heimsath et al. (1997)
$Y = E - D$ (9)	Sediment Yield	ASCE (1975)
$SDR = \frac{Y}{E} * 100$ (10)	Sediment Delivery Ratio	

z = elevation from the DEM
t = time
q = discharge per unit width
ρ_s = soil bulk density
x, y = horizontal directions
q_s = sediment flux, per unit width
S = slope
Q = discharge
A = contributing area
w = width of the flow, it was used cell size as equivalent
P_0 = rate of weathering of a bare bedrock surface
H = soil thickness
E = erosion
D = deposition
$\beta_1, \beta_2, m, m_1, n_1, m_2, \in, u, v$= constants

The sediment movement is thus limited by a channel's potential to transport material rather than the ability of the channel to detach material via erosional or weathering processes (Chen et al., 2014). The discharge and slope are controlled by exponential constants m_1 and n_1. The amount of sediment

transport is modulated by the constant β_1. Equation 6 was used to calculate the discharge. The upslope contributing area A is used as a proxy for the water flow. Equation 5 was used to calculate the discharge per unit width w. The stream incision law (Equation 7 in Table 12) was employed to predict erosion in concentrated flow areas. It states that the erosion rate in channel increases with the flux of water Q and with the local gradient S (Chen et al., 2014; Dietrich and Perron, 2006).

The model runs differently in unconcentrated and concentrated flow areas. Sediment transport in unconcentrated flow areas (inter-rills) was considered to be transport-limited. Under such conditions, the potential of soil detachment is exceeding the sediment transport capacity of overland flow. In this case, the equation of channel's potential to transport material (Equation 4 in Table 12) is used in Equation 3 to calculate the actual sediment transported in different directions. For the concentrated flow areas, a detachment-limited case was assumed whereby the transport capacity of overland flow exceeds its detachment capacity (Haan et al., 1994; Mitasova et al., 2013). In this case, the actual sediment transport is limited by the detachment capacity and it was calculated by the stream incision law (equation 7 in Table 12). The transport capacity might not be higher for all the stream courses. Sinks or barriers can reduce the transport capacity, consequently depositing sediments. This would affect, for example, the estimation of sediment delivery ratio. Since the aim of the model was to predict soil thickness, and the unconcentrated flow areas cover the major part of the study area, we considered this to be negligible.

The flow potential of transport material (equation 4 and 7) considered the discharge (derived from A) and the slope. High slope or discharge areas will have higher erosion than plain or low discharge places. However, the vegetation reduces erosion, minimizing the variation between different topographies. As forest is the dominant natural vegetation in our study area, a maximum sediment transport rate was used. As we assumed a soil thickness steady-state in Model 1 (section 4.2.5), the sediment transport was limited by the maximum rate of soil production, even if stream transport capacity or stream incision law was large. The system can therefore be kept in a dynamic equilibrium. In Model 2 (section 4.2.6), when soil thickness changes over time under a non-steady-state condition, the sediment transport was limited by two times the maximum soil

production rate. This value was arbitrarily chosen and allows occurrence of exposed bedrock without larger departure from the steady-state conditions.

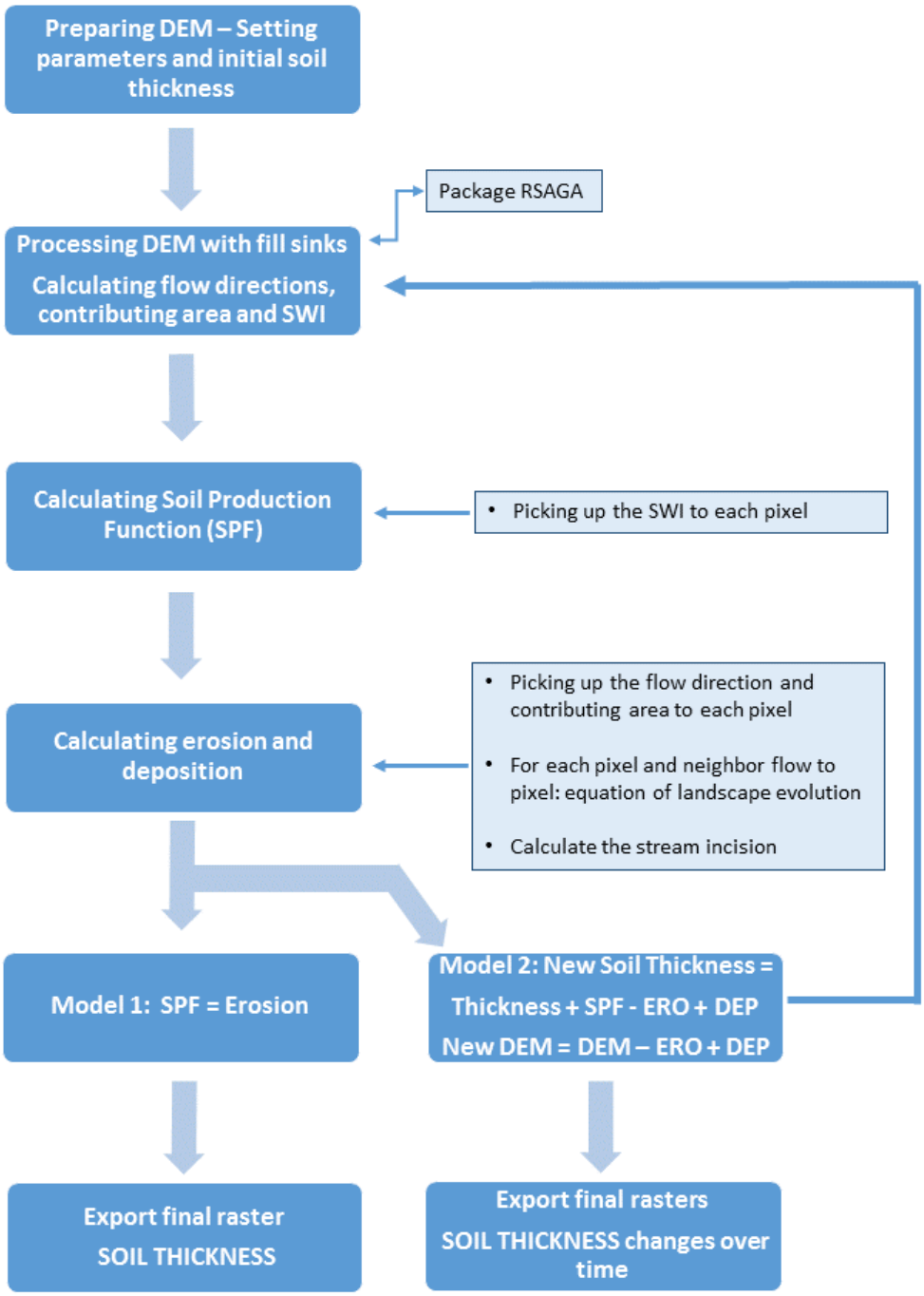
The SPF (Equation 8 in Table 12) was used to estimate the soil production rate, which varies with thickness of soil mantle overlying the bedrock (Heimsath et al., 1997; Stockmann et al., 2014). The function contributes to understanding the dynamics of bedrock to soil conversion, at different depths. When combined with a landscape evolution model, it allows to investigate how erosion or deposition influence soil thickness. If a steady-state is assumed, the models allows to estimate the current soil thickness.

The sediment yield (Equation 9) is the sediment outflow from a drainage area, and it was calculated as the difference between sediment eroded and sediment deposited within the catchment. It is dependent upon the rate of total erosion in the contributing basin and the efficiency of transport of the eroded materials. Deposition can occur at intermediate locations wherever the entraining runoff waters are insufficient to sustain transport (ASCE, 1975). The equation 10 calculates the sediment delivery ratio, which is the proportion of the sediment yield relative to total erosion, which is equivalent to soil loss in the watershed (ASCE, 1975; Haan et al., 1994).

4.2.4 Model implementation

Figure 13 shows the model implementation. The model used a digital elevation model (DEM) to determine the upslope contributing areas, flow direction and SWI (SAGA Wetness Index). The SPF and SWI are used to estimate soil production rate. Flow direction is used to estimate the direction of sediment transport for each pixel. Upslope contributing areas and slope were used to estimate soil erosion. The difference between erosion and sediment was used to calculate surface elevation changes. The model adjusted the soil thickness over time in response to sediments added or eroded and the soil production rate occurring in each pixel. It produced a new DEM and a new soil thickness map at each time step.

Figure 13. Diagram showing the model sequence to estimate soil thickness and soil thickness change over time.



Flow accumulation, determined by D8 algorithm (O'Callaghan and Mark, 1984), was used to separate unconcentrated flow from concentrated flow areas. Values higher than 100,000 in flow accumulation map were assigned to concentrated flow areas. Values lower or equal to 100,000 were set as unconcentrated flow areas. The flow accumulation using the D infinity algorithm (Tarboton, 1997) was used to determine the discharge Q (Equation 6).

Two models were implemented in the R statistical computing software to predict soil thickness (Model 1) and to simulate soil thickness changes over time (Model 2). Both used landscape evolution and SPF equations (Chen et al., 2014; Dietrich et al., 2003; Heimsath et al., 1997; Minasny et al., 2015; Saco et al., 2006; Stockmann et al., 2011; Willgoose & Rodriguez-Iturbe, 1991).

Model 1 aimed to predict soil thickness using the current topography. The model employed the exponential Soil Production Function (SPF) discussed by Heimsath et al. (2000, 1999, 1997), showing the decline of soil production rate with depth and a landscape evolution model simulating erosion and deposition processes. Erosion rates were calculated following Willgoose & Rodriguez-Iturbe (1991) and Saco et al. (2006) (Table 12). In addition, the influence of soil moisture accumulation in the landscape was evaluated by calculating a topographic wetness index, and investigating how it influences soil production rates. In total, four models were tested with different uses of topographic wetness index to predict the soil thickness to the catchment.A soil depth steady-state was assumed. In which, the soil production rate equals the erosion rate (Heimsath et al., 2000). In our study area the assumption is reasonable given the age of the surface (132 Myr) and the relatively stable climatic and tectonic conditions. While there has been a change in land use, accompanied by increasing erosion rates over the last couple of centuries, we assume that the cumulative effect of this anthropogenic phase on soil thickness is limited, and current soil profiles depths are close to their steady-state condition. Model 2 was elaborated based on calculating landscape evolution over time. Different from Model 1, no steady-state condition was imposed and both erosion and deposition of sediment were taken into account. The evolution of surface elevation and soil thickness were calculated from the net change between the erosion within each cell and the sediment deposited into each cell from upslope cells. In this case, soil thickness may increase even at a low soil production rate. This model allowed to investigate soil thickness evolution and to provide a provisional knowledge about areas in the landscape where steady-state does not apply and where the soil thickness is increasing or decreasing.

4.2.5 Model 1

Model 1 estimated the current soil thickness, based on potential erosion and under a steady-state condition. The current topographic surface conditions, assumed to reflect the long-term landscape and soil thickness steady-state, were determined by a current DEM and it was used to estimate erosion rates. The elaborated model (section 4.2.4) was run over 1 year, predicting erosion E and total deposition D to the whole catchment. The total erosion and deposition were used to determine the sediment yield and the sediment delivery ratio (Equations 9 and 10 in Table 12). To predict soil thickness, the erosion rates E were assumed to be equal to soil production rates. The SPF were then used inversely to calculate the H (soil thickness). Different SPFs were used to investigate four different scenarios, for the purpose of testing the soil moisture influence by distinct spatial functions.

The modulated SPF varied by a spatial function, as in Saco et al. (2006), formulated as:

$$SPF = P_0 e^{-mH}(1 + a * F(x,y)^d) \quad , \qquad (11)$$

where P_0 is the rate of weathering of a bare bedrock surface, m is the depth decay parameter and H is the soil thickness. The additional parameters are $F(x,y)$ representing a spatial varying SPF, a and d respectively a multiplicative and exponential constant of the spatial function.

At steady-state, $SPF = E$, thus the equation was adjusted to produce the values of H (soil thickness):

$$H(x,y) = \frac{-1}{m} \ln\left(\frac{E}{P_0(1+a*F(x,y)^d)}\right) \qquad (12)$$

Some results yielded negative values of *H* so the equation needed adaptation, as it is a logarithmic equation. Herein, the maximum value of erosion was limited to P_0, to avoid negative values of *H*. It is consistent with the maximum sediment transport considered (section 4.2.3). The values of zero erosion were changed to a minimum erosion of 10^{-4} mm/yr.

In the first scenario tested (Table 13, scenario 1) no spatial function was used, i.e., the value of a was set to zero and soil production only depends on local soil thickness.

Table 13. Soil production functions used to predict soil depth (up to BC horizon).

Function	Scenarios
$H = \frac{-1}{m}\, ln\left(\frac{E}{P_0}\right) \quad (13)$	1.Soil depth H from Soil Production Function
$H = \frac{-1}{m}\, ln\left(\frac{E}{P_0(1+a*\widetilde{SWI}^d)}\right) \quad (14)$	Soil depth H from Soil Production Function using a modified SWI (standardized $\widetilde{SWI}$): 2.a with coefficients a = 3 and d = 1 2.b with coefficient a = 10 and d = 5
$H = \sqrt{SWI} - C_1 \quad (15)$	3. Soil depth H from function using only SWI, C_1 = 1.3

H = Soil thickness
P_0 = Rate of weathering of a bare bedrock surface
E = Erosion
SWI = Saga Wetness Index
$\widetilde{SWI}$ = Modified Saga Wetness Index
m, a, d, C_1 = Constants

In scenarios 2.a and 2.b, soil moisture was included to modulate the soil production. A modified topographic Saga Wetness Index (SWI) was used. The SWI was based on a modified catchment area calculation. It predicted, for the cells in valley floors with a small vertical distance to a channel a more realistic and higher potential soil moisture compared to the standard Topographic Wetness Index calculation (Böhner and Selige, 2006). The parameters a and d were set to 3 and 1 on scenario 2.a. These values were based on the work of Saco et al. (2006). Based on iterative procedure, comparing to the observed *H*, parameters a and d were adjusted to 10 and 5 in scenario 2.b.

The SWI was obtained using the package RSAGA and rescaled, obtaining a modified SWI ($mSWI$) (Equation 16). In scenario 3, the SWI was used in its usual form. For rescaling the SWI, the average SWI plus the standard deviation was set to 1. The average SWI minus the standard deviation was set to

0. The minimum value was set to 1x10^{-5}. This calculation of a $mSWI$ was preferred due to perform better than just setting the maximum to 1 and minimum to 0, as in Saco et al. (2006). The equation used was:

$$mSWI = \frac{(maxSWI - SWI) * (oneSWI - zeroSWI)}{(maxSWI - minSWI) - oneSWI} \quad , \quad (16)$$

where SWI = value of SWI corresponding to each pixel; $oneSWI$ = 1; $zeroSWI$ = 0 $maxSWI$ = average (SWI) + standard deviation (SWI)and $minSWI$ = average (SWI) – standard deviation (SWI). At each iteration, the only parameter that varies is the SWI.

In scenario 3 only SWI was used as a predictor. A non-linear regression was used to relate SWI and H (Equation 15). A square root of SWI was chosen based on the validation parameters (section 4.2.8).

4.2.6 Model 2

Model 2 studied soil thickness changes over time without assuming a steady-state. The soil thickness predicted in Model 1 was used as initial values and the topography varies with sediment added or eroded at each cell by sequential iteration. Soil thickness was calculated based on variation caused by net sediment transport, added to soil production as in Equation 1. The erosion and soil production equations were implemented separately, and no steady-state was imposed, so the erosion rates were not necessarily equal to soil production. The soil production rate was controlled by the soil thickness at each time step using the classical exponential SPF (Equation 11).

In order to reduce the computational time, a 500 year time step was adopted for LEM, iterating up to 100,000 years (Fig. 13). A subroutine was elaborated to run SPF with a time step of 100 years. The changes in soil thickness, as compared to the current situation were modeled for 25 kyr, 50 kyr, 75 kyr and 100 kyr.

4.2.7 Model parameterization

Some equations parameters are specific for each condition and need to be adjusted as the environment change. Herein, the model parameters were chosen to reflect the characteristics of the study area and were taken from the literature in areas with similar environment. The parameters are shown in Table 14 and refer to equations in Table 12.

β_1 (Equation 4) is the erodibility factor (Willgoose, 2005). Peckham (2003) named this “transportability” of the surface material, which depends on soil type, land use, cohesiveness, vegetation, etc. It was set at 0.0004 per year, which produced average erosion close to average rate of soil production.

The power of q and S (m_1 and n_1 in Equation 4) are taken depending on the type of sediment transport modeled and values between 1 and 2 are typical for both (Peckham, 2003). Willgoose (2005) and Willgoose & Rodriguez-Iturbe (1989) used m_1 = 1.5 and n_1 = 0.7 and the same values were used here.

The parameterβ_2 is equivalent to runoff rate and the value was used based on coefficient of runoff in forested areas in Brazil, 0.35 (Pruski, 2009). Considering the amount of precipitation (1736 mm per year) β_2 was determined as 0.6 m yr^{-1} (35% of precipitation). The constant m_2 was 0.8 based on Willgoose (2005) and Willgoose & Rodriguez-Iturbe (1989).

The parameter for the incision law u and v was taken based on different literatures (Chen et al., 2014; Howard & Kerby, 1983; Kirkby, 1969). Values of 2/3 and 2/3 were chosen under the relation m/n = 1, following Seidl & Dietrich (1992). The $\in$ value was chosen to produce an average erosion in concentrated flow areas 6 times higher than in inter-rills, being $\in$ = 0.001 m yr^{-1}.

The soil bulk density ρ_s was determined as the average found on study area, of 1200 kg/m^3 (Bonfatti et al., 2016), considering bulk densities up to 100 cm depth.

The parameter P_0 , the rate of weathering of a bare bedrock surface, was set to 143 mm kyr^{-1}, based on a study in Australia with similar elevation (Heimsath et al., 2001a). The m value (4 m^{-1}) was chosen based on values found in Heimsath et al (2005, 2001a, 2000) and adjusted to produce soil thickness similar to the dataset.

Table 14. Variables and parameters used in equations for modeling landscape evolution and soil production functions.

Parameters	Description	Values
Landscape evolution		
q_s	Sediment flux per unit of width	Calculated
β_1	Rate constant for sediment transport by year	0.0004
q	Discharge per unit width	Calculated
m_1	Power of q in sediment transport equation	1.5
S	Slope between the cell and the neighbor	Calculated
$n1$	Power of S in sediment transport equation	0.7
Q_c	Discharge, in the channel	Calculated
w	Width of channels	15
β_3	Multiplicative constant, relating discharge to contributing area	0.6
A	Contributing area	Raster
m_3	Power constant, relating discharge to contributing area	0.8
$\in$	Rate constant for stream incision	0.001
u	Discharge per unit width in stream incision equation	2/3
v	Power of S in stream incision equation	2/3
ρ_s	Soil bulk density	1200
Soil Production Function		
P_0	Rate of weathering of a bare bedrock surface	0.000143
m	Depth decay parameter	4

4.2.8 Validation

The Model 1 was validated using the soil thickness values from the database of Flores et al. (2012). The predicted values were compared to the observed values until the horizon BC or B (in case the BC was absent), down to 150 cm depth. Soil thickness values from 42 profiles were selected.

For the validation the following parameters were used: COR (coefficient correlation), ME (mean error), RMSE (root mean square error) and CCC (Lin's coefficient). The CCC parameter was used to quantify the proximity

of the distribution (observed x predicted soil thickness) from the 1:1 line (Lin, 1989). The ME, RMSE and CCC (Lin, 1989) was implemented in R software:

$$\mathrm{ME} = \frac{1}{n}\sum_{i=1}^{n} \hat{z}(x_i) - z(x_i) \qquad (16)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} [z(x_i) - \hat{z}(x_i)]^2} \qquad (17)$$

$$\mathrm{CCC} = \frac{2 * \rho * \sigma_{\hat{z}(x_i)} * \sigma_{z(x_i)}}{\sigma^2_{\hat{z}(x_i)} + \sigma^2_{z(x_i)} + (\overline{\hat{z}(x_1)} - \overline{z(x_1)})^2} \qquad (18)$$

Where n is the number of validation sample points, $z(x_i)$ is the observed value, $\hat{z}(x_i)$ is the predicted value, $\sigma_{z(x_i)}$ and $\sigma_{\hat{z}(x_i)}$ are the standard deviation, $\sigma^2_{z(x_i)}$ and $\sigma^2_{\hat{z}(x_i)}$ are the variances, and ρ is the correlation coefficient between the predictions and observations.

4.3 Results

4.3.1 Model 1: Soil thickness prediction

Variation in soil depth was evaluated against different topographic variables. SWI and Slope correlated well with soil thickness, and were investigated as spatial condition to the soil production function (Table 15 and Figure 14). SWI had a positive correlation whereas Slope had a negative correlation. Thus, SWI was added in the SPF and it improved soil thickness prediction significantly. Slope, due to its interrelation with SWI, did not improve the model and it was not included. Land use and lithology did not correlate well with soil thickness (Figure 14).

Table 15. Correlation matrix of soil depth (Flores et al., 2012) and topographic covariates.

	Soil depth	LandUse	Geology	Slope	SWI	Aspect	Elevation
Soil depth	1	0.071	0.13	-0.36	**0.46**	-0.23	-0.0069
LandUse		1	-0.40	-0.19	0.087	-0.18	0.25
Geology			1	0.35	-0.13	-0.11	-0.67
Slope				1	-0.67	0.019	-0.36
SWI					1	-0.18	0.11
Aspect						1	0.16
Elevation							1

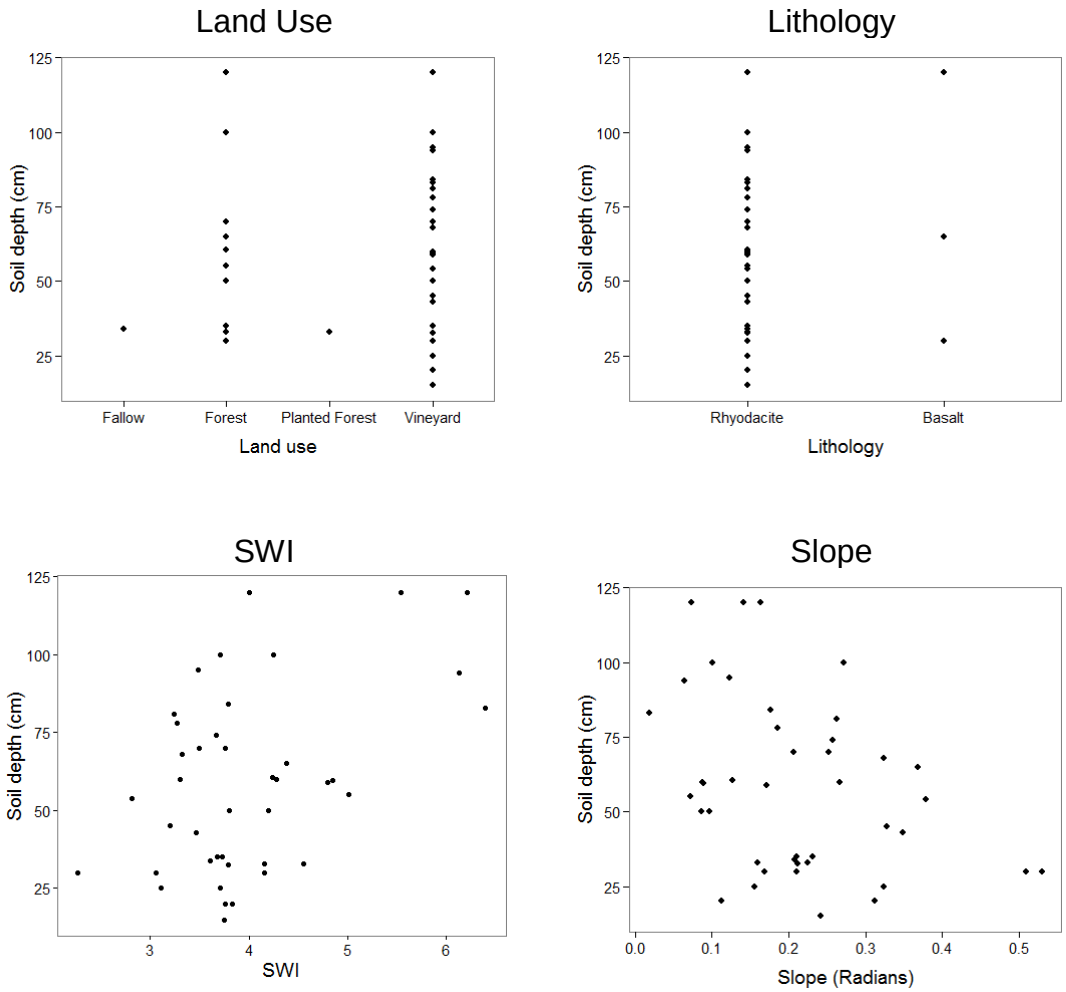


Figure 14. Distribution of soil depth values by land use, lithology, SWI and Slope.

The soil thickness prediction was evaluated under four scenarios (Table 13) and each scenario was validated using the field observations (Fig. 15).

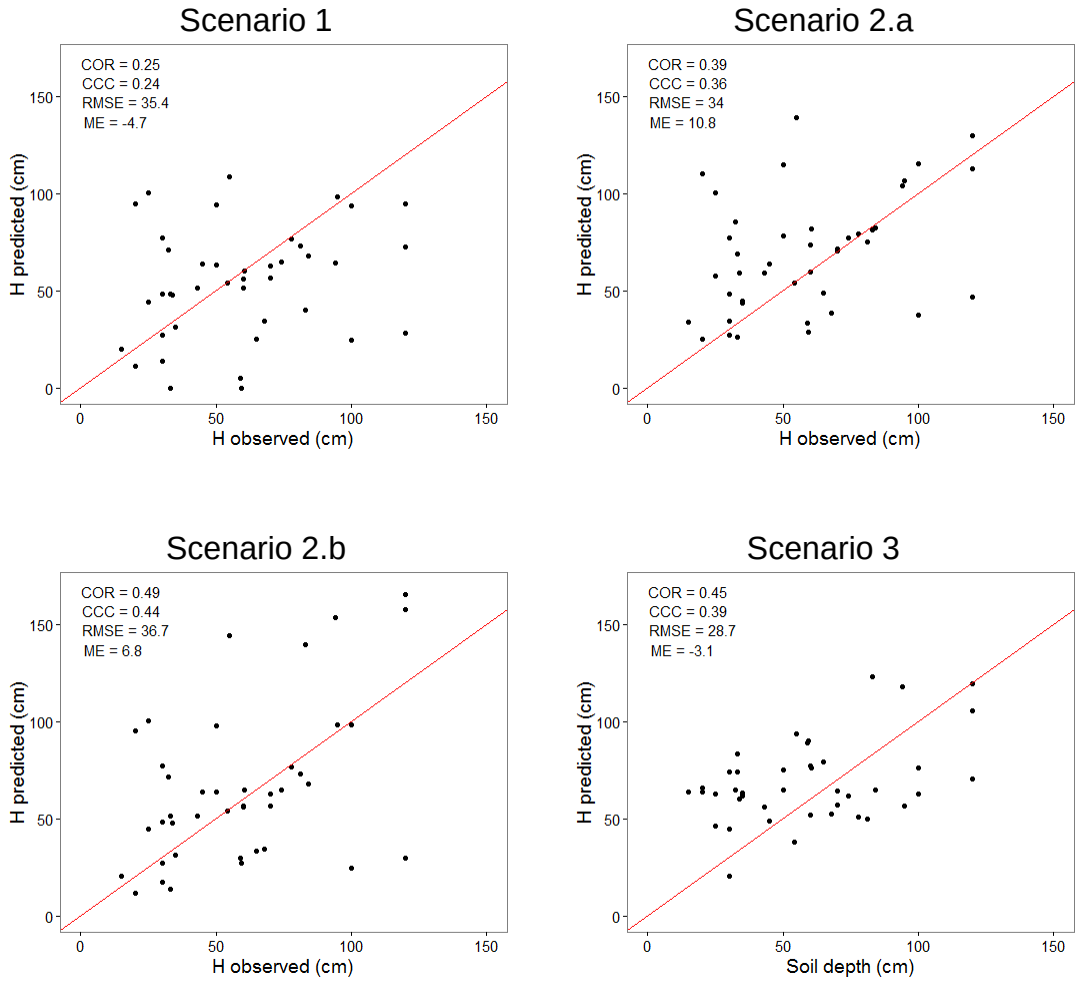


Figure 15. Distribution of H predicted and H observed, for the four scenarios in Table 13.

Scenario 1 showed lower correlation and CCC value (Fig. 15). The RMSE (35 cm) was similar to scenarios 2.a and 2.b. The mean error was the lowest, indicating some underestimation. The observed and predicted soil thickness showed a scattered distribution with underprediction. The soil thickness map of Scenario 1 (Fig. 16) showed high erosion in the valley bottom and some increased thickness in uplands. The areas with high thickness coincided with areas with low erosion, not taking in account the moisture accumulation. The negative residuals are concentrated northward (close to basin outlet), with some sparse points in the uplands. The positive residuals are mainly in uplands. The values of predicted thickness were in the range of 0 to 2.6 m, with an average of 0.47 m.

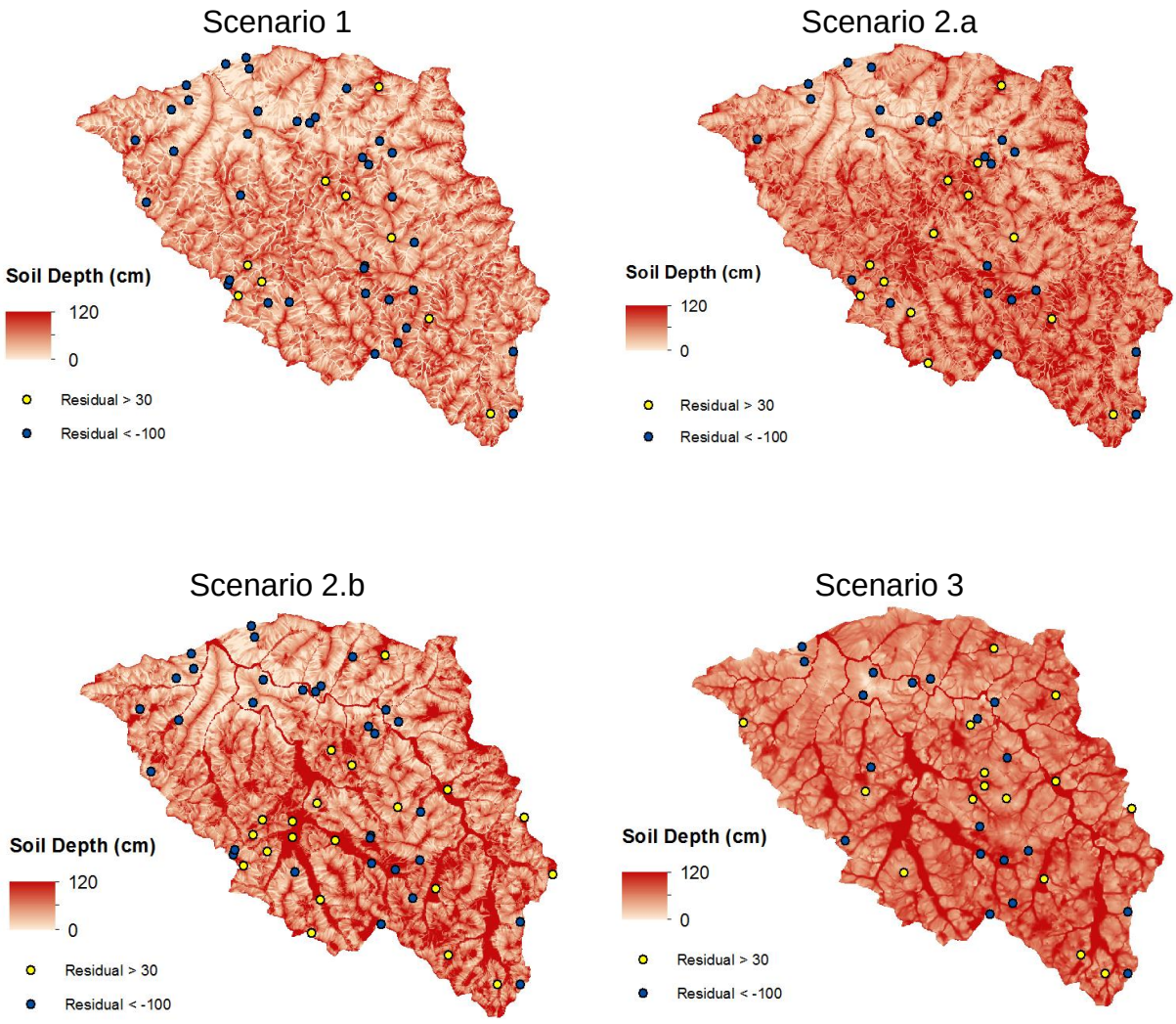


Figure 16. Maps of soil depth produced from the four different scenarios in Table 11. The blue and yellow points represent profiles with high residuals (predicted - observed), higher than 30 or lower than -100. Negative residual (blue points) indicate underestimation and positive residual (yellow points) indicate overestimation.

Scenario 2.a used a spatial function based on mSWI and showed higher correlation coefficient (0.39) and CCC (0.36) than Scenario 1. The ME was positive, suggesting an overestimation. The distribution of observed and predicted thickness (Fig. 15) was more uneven than Scenario 1 or 3, and similar to Scenario 2.b. The soil thickness map (Fig. 16) of this scenario showed the influence of moisture accumulation from mSWI. Larger predicted thickness is observed in upland areas. The negative large residuals are in valley bottoms and some in uplands. Values of predicted thickness were in the range of 0 to 3.2 m, with an average of 0.65 m.

Scenario 2.b presented the highest correlation coefficient (0.49) and CCC (0.44). The ME (6 cm) indicated a general overestimation. Comparing predicted and observed thickness, the distribution was more similar than other scenarios. The soil thickness map showed thicker soils dominating uplands, while shallow soils were dominant in the valley bottoms. The residuals had a similar pattern to the previous function, with negative values prevailing in the valley bottoms and positive values on the uplands. Soil thickness prediction varied from 0 to 4.3 m, with an average of 0.67 m.

Scenario 3 used the direct relation between SWI and soil thickness. It showed a slightly lower correlation (0.45) and CCC (0.39) than Scenario 2.b. The RMSE (28.7 cm) was the lowest and the ME (-3.11) indicated a general underestimation. The predicted and observed thickness showed a small range distribution. The soil thickness map of Scenario 3 showed less difference between valley bottoms and uplands, and also fewer profiles with high residuals. This reflects the small range distribution observed in Figure 15.

The spatial parameter SWI varied according the flow accumulation on surface. The mSWI showed values between $1x10^{-5}$ and 3.08. Figure 17 shows examples of how the soil production function varied as influenced by mSWI. The graph showed that the mSWI influence is higher when the interface bedrock-soil is closer to surface. As soil mantle increases, the influence of mSWI is reduced. The model considers that the predicted soil thickness is the depth where the soil production rates is found equivalent to erosion rates. Consequently, soils with high SWI will be thicker.

The landscape evolution model had 22% as sediment delivery ratio. The total materials eroded were 0.64 ton/ha, being 0.49 ton/ha/yr deposited and 0.15 ton/ha/yr lost. In unconcentrated flow areas, the average erosion was 4.77×10^{-2} mm/yr. In concentrated flow areas, erosion was significantly higher, around 2.86×10^{-1} mm/yr.

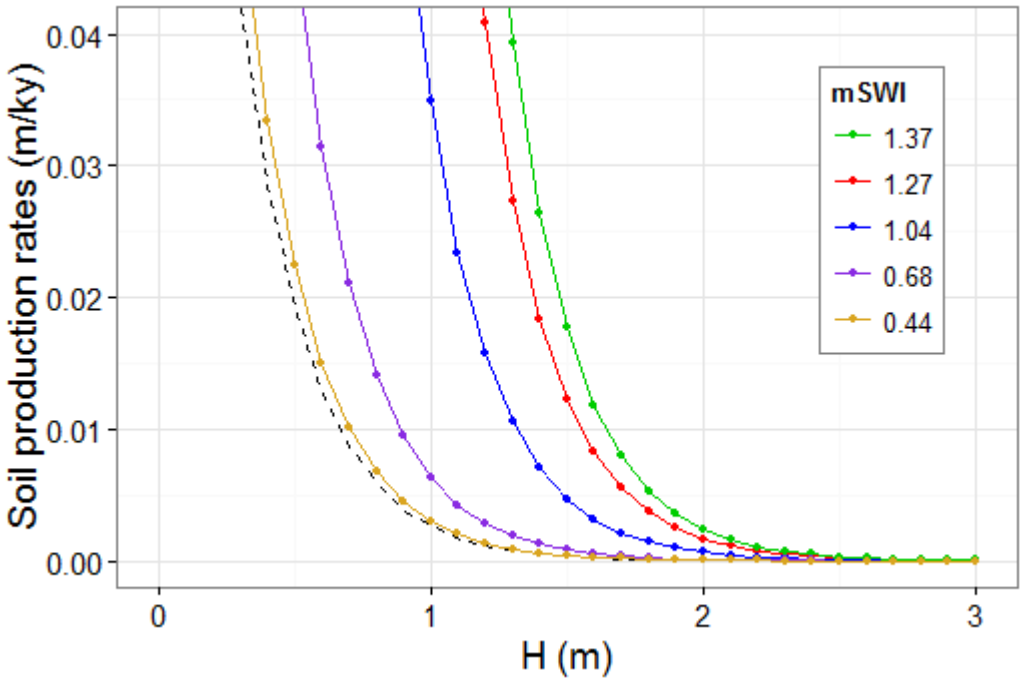


Figure 17. Examples of soil production rates as a function of depth, calculated by different SPF (Equation 11), in m/kyr. For 5 different profiles (Flores et al., 2012) the SPFs without using mSWI (Equation 8 in Table 12) yield the same SPF (dashed line), while using mSWI yield different SPFs (solid colored lines) .

4.3.2 Model 2: soil thickness changes over time

The soil thickness evolution (Figure 18) showed a reduction of soil thickness in areas close to concentrated flows areas and an increase of soil thickness in depositional areas. This trend gets stronger up to 75 kyr and not so much changes after 75 kyr. Soil thickening in valley bottom, due the SPF and depositions, is evident in the maps as well the higher erosion in channels and streams. The average soil thickness to the whole catchment changed from initial 67 cm to 92 cm (25 kyr), 98 cm (50 kyr), 101 cm (75 kyr) and 103 cm (100 kyr).

The sediment addition or erosion rates increases over time for some profiles and reduces in others (Fig. 19). Generally the soil thickness seems to reach a dynamic equilibrium (e.g. profiles 3, 35, 16,132), nevertheless in some profiles, the thickness evolution is very irregular and clearly not in a dynamic equilibrium at the end of the simulation (e.g. profile 7 and 40).

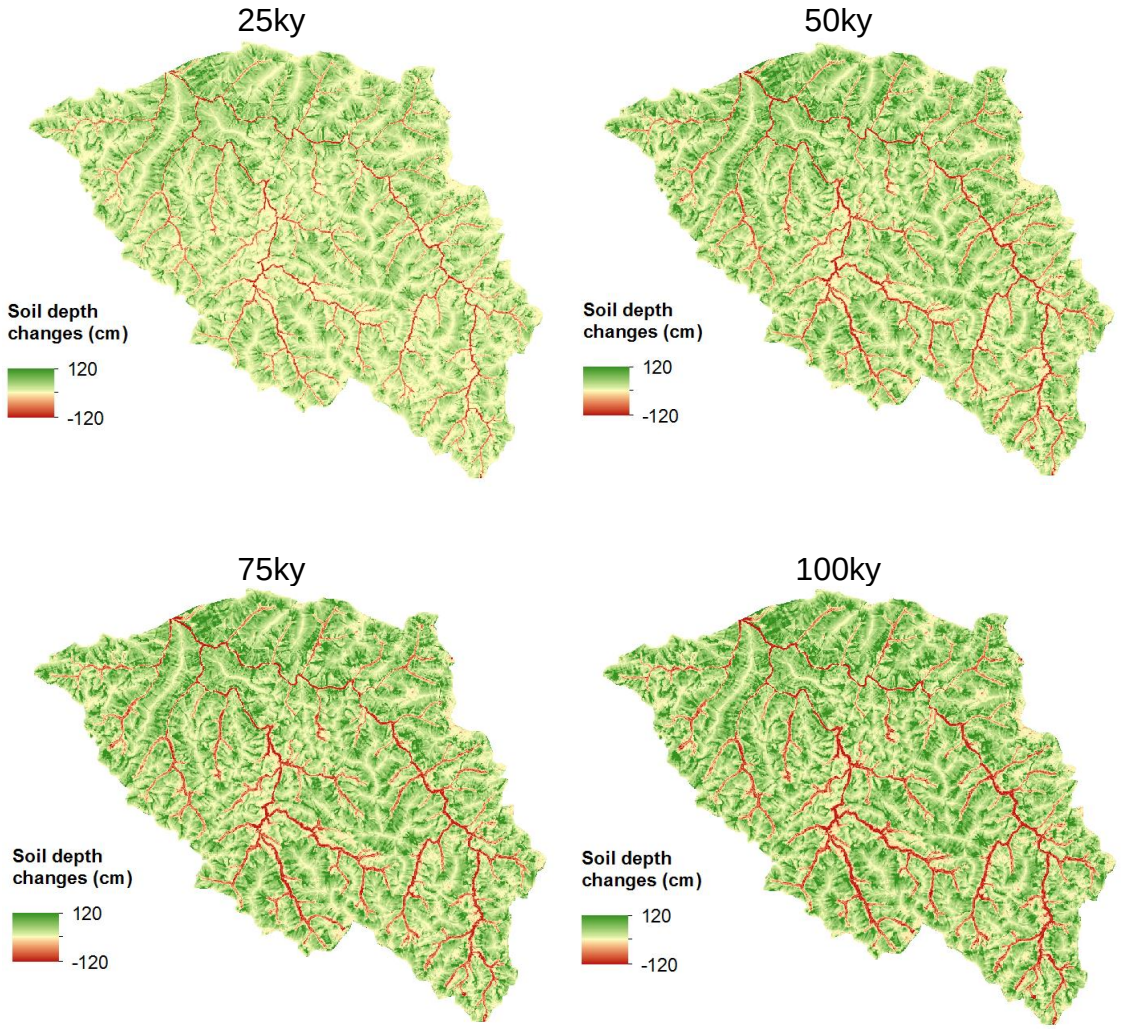


Figure 18. Difference between the soil depth predicted at different times and the initial soil depth map (Fig 17, Scenario 2.b). Green color shows increasing in soil depth and brown color shows decreasing in soil depth.

The profiles located in upper depressions, convergent back slope and foot slope showed higher variations over time. Profiles in divergent back slopes, toe slope or back slope showed a smoother and continuous variation in sediment transport and soil production rates. The profile 3 (divergent back slope) and profile 132 (back slope) revealed a tendency of remaining as shallow soils. Profile 35, in a convergent back slope, showed a tendency of significantly increasing soil thickness. The erosion rate is higher than soil production rate at time 0, but the difference reduces with time. Since the condition SPF = E was not imposed in Model 2, erosion could be higher than SPF at time 0, although it was arbitrarily limited to twice the soil production rate (section 4.2.3). The same was observed in profile 40. The profile 16, under a toe slope, was the deeper soil and presented

a tendency to keep the larger soil mantle. The sediment addition trends generally followed the sediment erosion trends. As the model employed a high time step (500y), the continuous behavior of addition or removal of sediments is not sufficiently well characterized, showing instead a leaping variation. The soil production rates reduce over time, due the thickness accumulation, and seem to reach a steady-state for some profiles. The SPF averaged over the entire study area was close to 50 mm/kyr at initial years and a steady-state seems to be reached around 40kyr, when the SPF assumes an average close to 10 mm/kyr and did not change significantly up to 100kyr.

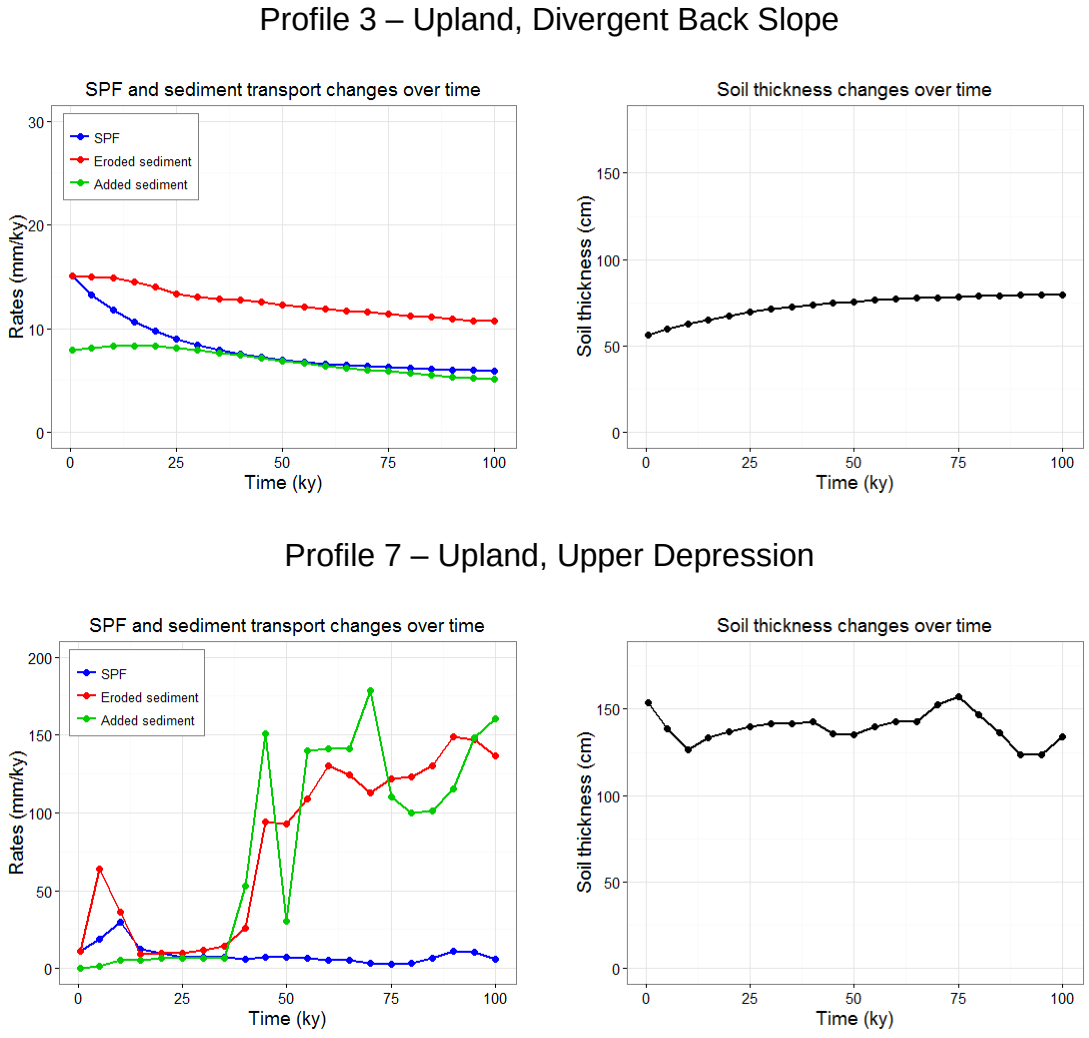
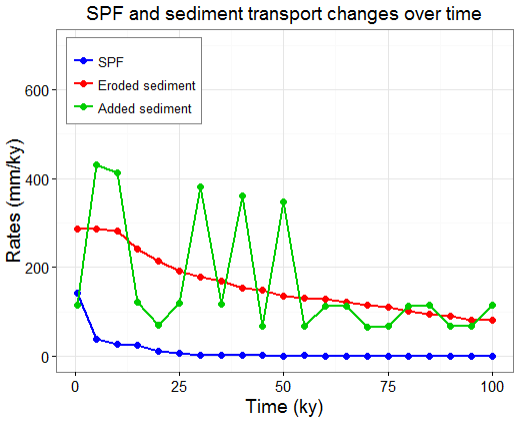
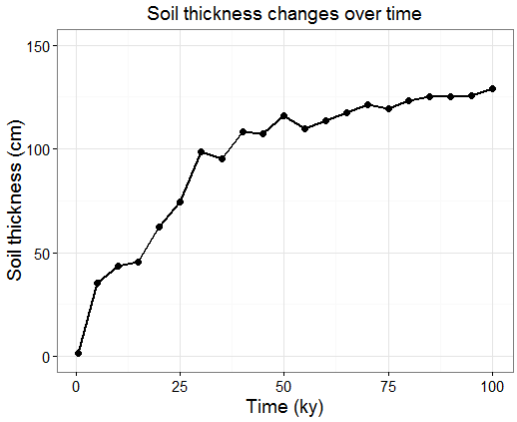


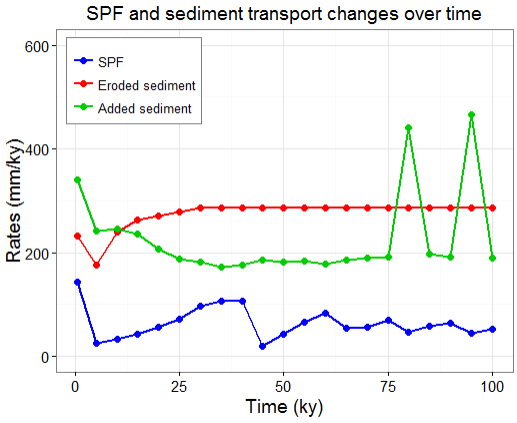
Figure 19. Estimated changes of SPF, sediment transport and soil thickness over time, using the combined SPF and landscape evolution model at representative profiles.

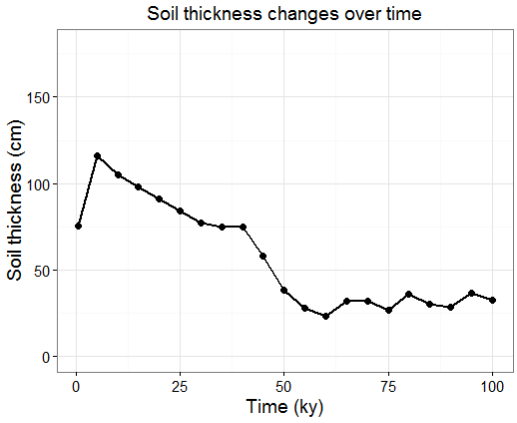
Profile 35 – Upland, Convergent Back Slope



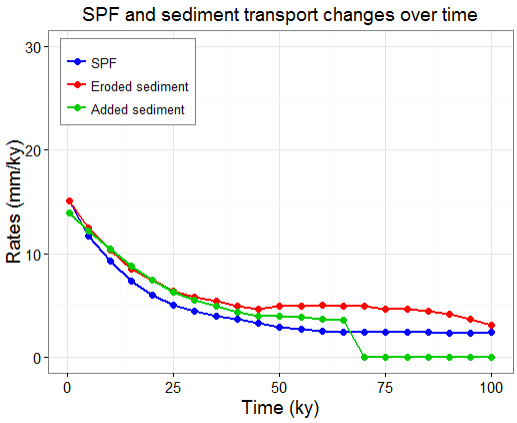


Profile 40 – Valley Bottom, Foot Slope





Profile 16 – Valley Bottom, Toe Slope



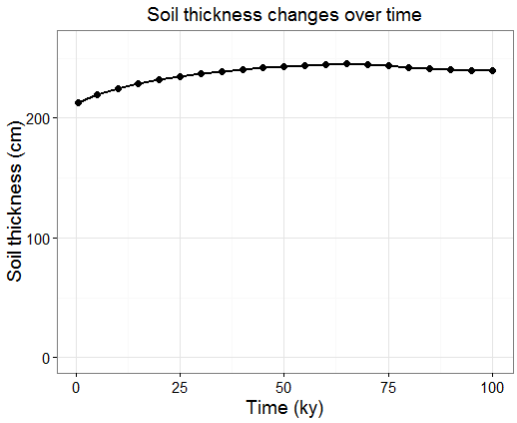
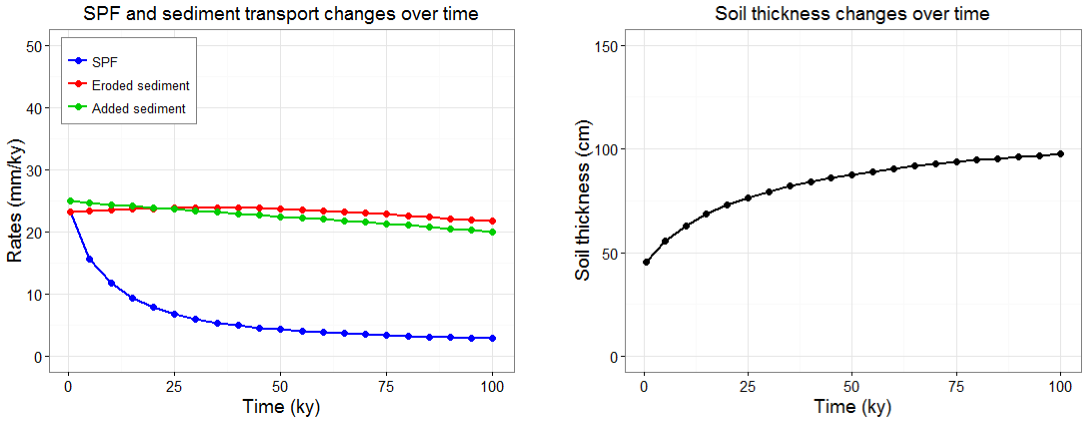


Figure 19. Continuation...

Profile 132 – Back Slope



Average Area

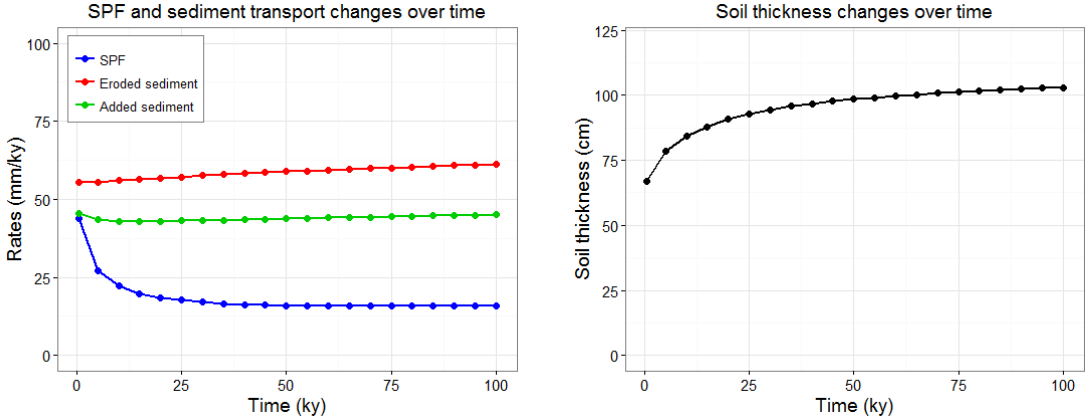


Figure 19. Continuation...

4.4 Discussion

This study used a soil production function coupled with a landscape evolution model to predict soil thickness and to analyze its evolution over 100 000 years. We tested and validated soil thickness maps produced from four different soil production functions against field observations.

4.4.1 Model1: Soil thickness prediction

The initial parameterization of the LEM for the current topography played a critical role in the quality of soil depth predictions and produced reasonable results when compared to literature.

The total erosion rate (0.64 ton/ha/yr) was similar to other studies in forest areas. Guimarães et al. (2011), using USLE, found erosion rates ranging from 0 to 1 ton/ha/yr under natural forest areas in Santa Catarina, South Brazil. In soils under agriculture, the erosion rates estimated here are similar to that found in areas under no-tillage. Cogo et al. (2003) found in Santo Angelo in Rio Grande do Sul State erosion in soils under conventional tillage about 13 ton/ha in 1.5 years, under reduced tillage about 4 ton/ha in 1.5 year and soils under no till close to 1 ton/ha in 1.5 year. Bertol et al. (2007) found in Santa Catarina State erosion rates between 0 and 1 ton/ha /yr for the most agriculture areas with no tillage, but much of the erosion was higher than 2 ton/ha/yr under conventional tillage. Martins et al. (2003) found in Espírito Santo State in Brazil an average erosion of 1.4 ton/ha/yr in planted forest of eucalyptus and 0.1 ton/ha/yr in natural forest. LEM represents the erosion dynamics satisfactorily as the published erosion rates were comparable to the results from our study; we assumed that the natural forest is the predominant vegetation during the time of soil formation.

A homogeneous land use (natural forest) was used in the models. Soils under forests have low erosion rates and currently occupy 44% of the study area whereas soils under vineyards currently occupy 31% of the area. If erosion were significantly higher in soils under vineyards than in soils under forests, this would be reflected in soil thickness. However, we found little correlation between land use and soil thickness (Table 15). Studies has shown that vineyards and cover crops can significantly reduce erosion (Botton et al., 2010; Oliveira et al., 2004; Shanks et al., 1998).

The value of sediment delivery ratio (22%) for the study area (8 118ha) is higher compared to that found in other catchments with a similar surface area (ASCE, 1975; Haan et al., 1994), where values between 10 and 15% have been reported. The model did not consider deposition in areas of concentrated flow or might indicate that the time step should be reduced. The sediment carried to streams and channels were removed from the system. Actually, sediments transported by channels may be deposited in channel sinks, river banks and floodplain areas. When considering the events of erosion only in unconcentrated flow areas, the SDR drops below 8%. In Brazil, Minella et al. (2014) found in the Arvorezinha catchment (119 ha), a SDR close to 15%. An SDR of 18% was found for the Conceição River catchment (80,000 ha), Rio Grande do Sul State by

Didoné et al. (2015). It seems that SDR decreases with increasing catchment size so we expect a SDR below 15% in our study. As pointed out by Minella et al. (2014), in making comparisons of SDR by catchment size the range of catchment characteristics, including climate, topography, drainage density, geology and soils, land use and soil management are important.

The SWI improved soil production functions. When included, the soil thickness prediction improved significantly, as observed by the differences between scenarios 2.a, 2.b and 3 when compared to scenario 1 (Fig. 15).

Scenario 1 showed a scattered distribution of observed and predicted soil thickness. Nicótina et al. (2011) applied a similar method to model soil thickness and also found a high scatter concluding that the model did not represent well the observed soil thickness patterns. In our study, however, we were able to improve model performance significantly by considering spatially varying SPF according to topographic wetness index. That was observed in Scenario 2.b, showing the best performance despite a large RMSE and ME. The ME of 10.8 cm is similar to the one found by Catani et al. (2010), of 11 cm. With higher coefficients a and d (Equation 14, Table 13), the function maximized the soil thickness variation by the SWI, which a maximization of moisture influence in soil formation compared to scenario 2.a. Scenario 3 presented similar values to scenario 2.b and the difference is noted when observing the distribution of predicted and observed values (Fig. 15). The variation in scenario 2.b is higher and the low soil thickness estimation improved, compared to scenario 3. The deepest soils presented the higher residuals, contributed to the high RMSE and ME. The scenario 3 did not show soil thickness deeper than 125 cm and the performance to predict shallow soils was poor. However, the validation was close to scenario 2.b emphasizing the importance of SWI in predictions.

The standardized SWI performed well as a spatial function. There was no fixed maximum mSWI values. The range of soil thickness was not limited, minimizing the limited range problem related by Saco et al. (2006), when attributing 1 as maximum to spatial function. The mSWI allowed to simulate the moisture effect on accelerated weathering and forming deeper soil even in pixels with high erosion. As observed in Figure 17 when erosion is higher, the soils might be shallow or deeper depending on the SWI influence.

High negative residuals where found mainly in valley bottoms (Fig. 16), indicating underestimations of soil thickness. That area coincides with high flux accumulation that would produce cumulative soils due the high deposition. This leads to some problem in predicting soil thickness, because high deposition would deviate the soil of a steady state condition.

The valley bottom of the study area has Mollisols (Chernossolos in Brazilian System of Soil Classification) as dominant soils (Flores et al., 2012). Buol et al. (2011) stated that accumulation of sediments is an important process in Mollisols occupying depositional landscape positions. Erosion occurring along hillslopes can result in overthickening of the mollic epipedon. Cumulative soils with thick A horizon are common due to deposition of organic-matter-rich material from upslope, or to organic-matter accumulation at the site while sediment is accumulating, or to a combination of both processes (Birkeland, 1999). These deposited materials undergo pedogenic processes as they are added and the A horizon may continually thicken (Buol et al., 2011; Schaetzl and Anderson, 2005). That seems to be occurring in the study area as showing by the reference profiles (Flores et al., 2012), in which Mollisols have between 2 and 6% of SOC in A horizon and a soil carbon map indicates carbon accumulation in valley bottoms (Bonfatti et al., 2016). Moreover, the high organic material contributes to form soils with high porosity, reducing the bulk density and the soil would tend to become deeper. These effects were not simulated in the SPF. The deposition could be predicted by a LEM if we knew the initial topography condition but the uncertainty about this initial condition is an important problem in landscape evolution studies (Minasny et al., 2015; Peeters et al., 2006).

The positive residuals (Fig. 16) seem to appear around a line separating uplands and valley bottoms. It seems the model was not recognizing the high erosion driving soil formation in these landscape locations. The high curvatures along this escarpment where the uplands transition into the valley bottoms, can lead to an increase in sediment transport, reducing the soil mantle and transporting the material to the valley bottoms. The hydrologic model using a DEM resolution of 15x15m may not be accurately representing this features.

The processes involving landscape evolution and soil production are complex, whereas model representation is necessarily simplified. Several aspects make an accurate prediction of soil thickness difficult. Soils are variable

over short distances and the numerous factors causing such variation are unknown. We are able to quantify the variation well but a pedological understanding is less developed. Human activities are also spatially variable and make it difficult to predict impacts, for example by changing local land use and changing erosion rate. Some areas close to concentrated flow areas are flooded at times which affects weathering and soil production. An adequate delimitation and quantification of deposition areas (not considered in SPF), is complex especially over long time spans. It would be more correct to use the bedrock DEM instead of surface DEM that can yield different results compared to a topographic wetness index used throughout a surface (Saco et al., 2006). Finally, in order to assess the validity of the assumption of the steady state condition, an evaluation of the correct evolution of uplifts and erosion with time is needed.

Despite these difficulties and uncertainties, the model predicted the tendency of soil thickness in the upland areas, with more uncertainty in lowland areas, as showed by higher residuals in lowland areas (Fig. 16).

4.4.2 Model 2: soil thickness dynamics over time

It is relevant to estimate the current soil thickness and the tendencies to change. Critical or unstable areas, with a probable decrease in soil thickness, may be identified as areas that require more attention. The soil evolution model (Fig.18) allowed identifying stable areas and areas with tendency of increasing or decreasing soil thickness over time.

As shown in the temporal evolution of representative soil profiles (Fig. 19), sediment transport did not assume a regular pattern over time. For some profiles (7, 40) sediment deposition or erosion rates increase as the landscape changes while for other profiles (3, 16, 35, 132) sediment addition or erosion rates decreased. Landscape position is explaining this trend as profile 7 and 40 are respectively in a depression and foot slope, areas with a tendency to receive upslope sediments. In profile 3, the sediment eroded was higher than sediment deposited and we could expect a decrease in soil thickness. However, the soil production rate was higher in comparison with the net sediment transport (erosion rate minus deposition rate), thus the soil thickness kept increasing.

In convergent or divergent back slopes, without considering abrupt changes on environmental conditions (climate change, tectonics, anthropic activities), the soil thickness seems to reach a dynamic equilibrium condition, then the soil mantle will not change significantly over time. The same could be observed in toe slope profiles in valley bottoms. This would reflect a continuous and balanced interaction among the sediment transport and soil production rates. Conversely, in foot slope or depressions, the representative profiles suggest a complex dynamic and then the soil thickness could present a constant variation, without reaching a steady-state. In convergent back slope or toe slope positions, the profiles suggest soil thickening. Shallow soils are predicted in non-convergent back slopes.

The changes in soil production rates over time showed a similar behavior, generally decreasing and approximating zero. Overall, the study area seems to approximate steady state towards the end of the simulation, as the produced soil thickness approaches a constant value for different locations in the landscape and for the average area.

This dynamic soil-landscape evolution model offered insight in what could happen with the soil thickness in the future. The model demonstrated the importance of varying soil production rates, interacting with sediments added or removed during soil formation. It is a useful tool to investigate the impacts of environmental.

4.5 Conclusions

The following can be concluded from this research:

- The study showed that using SPF and LEM to predict soil thickness is promising.
- In Model 1, the soil thickness predictions improved using spatially varied moisture and the accuracy was higher where the deposition process was not dominant.
- The SWI showed to be an important spatial parameter when modelling soil thickness, improving significantly the estimations.

- Model 1 predicted the tendency of soil thickness in the upland areas, with more uncertainty in lowland areas, due to be difficult to quantify and delimit the sediment deposition.

- In Model 2, areas with tendency of increasing or decreasing soil thickness were located at different positions in landscape. Soils in back slopes and toe slopes showed a continuous thickness variation over time, leading to a steady-state condition. Soils at the foot slopes and depressions showed complex behavior and did not seem to reach a steady-state condition over time.

5. CONSIDERAÇÕES GERAIS

Com base nos estudos realizados, conclui-se que tanto modelos empíricos quanto modelos mecanísticos se mostram como importantes ferramentas na predição e mapeamento de atributos de solos. O modelo empírico produz melhores resultados quando as amostras estão em um número suficiente para correlações estatísticas e são representativas do terreno. Uma mesma técnica pode ser aplicada na predição e mapeamento de diferentes atributos do solo. O modelo mecanístico é mais complexo e procura modelar o comportamento dos componentes envolvidos nos processos pedogenéticos. Tem base em estudo mais aprofundado dos fenômenos envolvidos e auxilia na explicação dos mecanismos naturais atuantes no decorrer do tempo. O modelo mecanístico também não requer alta densidade de amostragem de solo, bastando um número pequeno para calibrar os parâmetros das equações. Suas principais limitações estão relacionadas à sua complexidade física e matemática, à impossibilidade de modelar todos os inúmeros processos envolvidos na pedogênese e ao tempo computacional para a execução dos modelos. As equações e modelos são distintas para cada atributo de solo estudado. A escolha do uso de modelos empíricos ou mecanísticos pode ter base na densidade e representatividade amostral disponível e no conhecimento dos fenômenos envolvidos na formação de cada atributo.

A validação das predições para profundidade do solo, realizada tanto no Estudo 1 quanto no Estudo 2, apresentou resultados mais favoráveis ao modelo empírico, mas essas condições se deram devido à disponibilidade de dados de um levantamento detalhado de solos, com alta densidade amostral, algo não comum na maioria das regiões Brasil. O modelo mecanístico possibilitou predição com um nível de acurácia pouco menor, mas com a

possibilidade de ser calibrado com número amostral reduzido, apenas para a determinação de parâmetros para as equações.

O modelo empírico possibilitou estimar as incertezas, através de simulação estatística, o que constitui uma informação importante na avaliação de carbono orgânico no solo. A estimação da incerteza em modelos mecanísticos é mais complexa e envolve uma série de fatores que necessitam de estudo mais aprofundado.

Dentre pontos importantes identificados durante o uso de um modelo empírico estão: a incerteza na escolha de um método padrão para regressão, a indisponibilidade de dados em escala compatível e a falta de memória computacional para lidar com simulações estatísticas utilizando-se de um MDE de alta resolução. No Capítulo 1 foram apresentados estudos que utilizam de grande variedade de métodos de regressão, aplicados a diferentes propriedades ou classes de solo, sob diversas condições de clima e relevo. Não há ainda uma conclusão definitiva sobre qual método seria mais indicado para cada situação. Uma boa prática é o teste de diferentes métodos na área de estudo, identificando o que produz resultados mais próximos aos valores observados, como realizado no Capítulo 2. Durante a validação, é de fundamental importância a separação das amostras em conjunto de treinamento e conjunto de validação. Alguns modelos podem superajustar as estimativas às amostras de treinamento e ter uma performance bem diferente quando aplicados ao conjunto de amostras para validação. A escolha do método a ser utilizado deve ter como base os resultados obtidos a partir das amostras de validação.

A disponibilidade de dados em escala compatível é imprescindível para a obtenção de resultados mais acurados principalmente quando modelos empíricos são utilizados. A análise estatística ou geoestatística necessita de um número amostral significativo para produzir resultados com maior confiabilidade. Um valor geralmente utilizado como referência é o de no mínimo 100 pontos amostrais (Hengl, 2009). É importante também que os pontos sejam representativos do terreno, estando alocados em diferentes partes da superfície para representar sua variabilidade. A densidade de pontos amostrais determinará a escala de estudo. Amostragem de baixa densidade tende a reduzir a escala de trabalho, sendo necessária maior área para alcançar a quantidade mínima de pontos amostrais. A obtenção de mapas geológicos e de classificação

e de uso de solo em escalas compatíveis também é limitada. Mapas de uso de solo podem ser elaborados a partir de imagens orbitais, mas mapas geológicos e de classificação do solo ainda são escassos no Brasil para escalas mais detalhadas.

O uso de MDE de alta resolução ocasionou problemas de armazenamento em memória tanto durante sua aplicação em modelos empíricos quanto em modelos mecanísticos. Nos modelos empíricos, cada covariável gerada ocupou mesmo tamanho para armazenamento que o MDE original. Os problemas de armazenamento se tornaram mais críticos ao se implementar as 100 simulações geoestatísticas, para cada um dos 5 intervalos de profundidade do solo estudados, gerando um total de 500 imagens. A solução encontrada foi separar o processamento em diferentes lotes, com cada intervalo de profundidade correspondendo a um projeto único no software R. No modelo mecanístico, o MDE de alta resolução ampliou consideravelmente o tempo para processamento durante as iterações do modelo. O Modelo 2 executou os procedimentos para um total de 100.000 anos, utilizando um passo de 500 anos, num total de 20 iterações, levando aproximadamente 24 horas para finalizar o processamento. O modelo poderia reproduzir dados mais próximos da realidade se houvesse iterações ano a ano, num total de 100.000 iterações, o que seria impraticável devido ao tempo total para processamento e o espaço de memória necessário para armazenar os dados de cada pixel por ano. Com o avanço tecnológico, esses problemas tendem a ser minimizados.

O modelo mecanístico para predição de profundidade de solo (Modelo 1) apresentou dificuldade quanto à quantificação do depósito de sedimentos em solos cumulativos, pois a principal equação utilizada comparou taxas de erosão à taxa de produção de solos. A taxa de deposição de sedimentos poderia ser estimada caso houvesse disponível um MDE de um momento passado no qual se pudesse modelar o início de um processo deposicional e como evoluiu até o presente. Para obtenção de um MDE de um tempo passado o uso de equações dos modelos de evolução de paisagem de forma inversa seria uma possibilidade, mas ainda não há estudos conclusivos sobre a aplicabilidade dessa metodologia. Estudos nessa área são necessários para estimar quantitativamente a contribuição do depósito de sedimentos na formação da espessura do solo.

6. REFERÊNCIAS BIBLIOGRÁFICAS

ADHIKARI, K. et al. Digital Mapping of Soil Organic Carbon Contents and Stocks in Denmark. **Plos One**, San Francisco, v. 9, n. 8, p. 1-13, 2014.

ADHIKARI, K.; HARTEMINK, A. E. Digital Mapping of Topsoil Carbon Content and Changes in the Driftless Area of Wisconsin, USA. **Soil Science Society of America Journal**, Madison, v. 79, p. 155-164, 2015.

AGATONOVIC-KUSTRIN, S.; BERESFORD, R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. **Journal of Pharmaceutical and Biomedical Analysis**, Amsterdam, v. 22, n. 5, p. 717-727, 2000.

AITKENHEAD, M. J.; COULL, M. C. Mapping soil carbon stocks across Scotland using a neural network model. **Geoderma**, Amsterdam, v. 262, p.187-198, 2016.

AKPA, S. I. C. et al. Total soil organic carbon and carbon sequestration potential in Nigeria. **Geoderma**, Amsterdam, v. 271, p. 202-215, 2016.

ALMOND, R. J.; TONKIN, P. J. Pedogenesis by upbuilding in an extreme leaching and weathering environment, and slow loess accretion, south Westland, New Zealand. **Geoderma**, Amsterdam, v. 92, p. 1-36, 1999.

ANDERSON, R. S.; ANDERSON, S. P.; TUCKER, G. E. Rock damage and regolith transport by frost: An example of climate modulation of the geomorphology of the critical zone. **Earth Surface Processes and Landforms,** Chichester, v. 38, p. 299–316, 2013.

ANDRADE, S. F.; MENDONÇA-SANTOS, M. L. Predição da fertilidade do solo do polo agrícola do Rio de Janeiro por meio de modelagem solo x paisagem. **Pesquisa Agropecuária Brasileira**, Brasília, v. 51, n. 9, 2016.

ANGERS, D. A.; ERIKSEN-HAMEL, N. S. Full-Inversion Tillage and Organic Carbon Distribution in Soil Profiles: A Meta-Analysis. **Soil Science Society of America Journal**, Madison, v. 72, p. 1370–1374, 2008.

ANJOS, L. H. C. et al. Sistema Brasileiro de Classificação de Solos. In: KER, J.C.; CURI, N.; SCHAEFER, C.E.G.; VIDAL-TORRADO, P. (Ed.). **Pedologia:** Fundamentos. Viçosa: SBCS, 2012.

ARROUAYS, D. et al. GlobalSoilMap: Toward a Fine-Resolution Global Grid of Soil Properties. **Advances in Agronomy,** Amsterdam, v. 125, p. 93-134, 2014.

ASCE. **Sedimentation Engineering**. New York: ASCE, 1975.

BATJES, N. H.; DIJKSHOORN, J.A. Carbon and nitrogen stocks in the soils of the Amazon Region. **Geoderma**, Amsterdam, v. 89, p. 273–286, 1999.

BECHTEL, W. Explanation: A mechanist alternative. **Studies in History and Philosophy of Biological and Biomedical Sciences**, Oxford, v. 36, n. 2, p. 421-441, 2005.

BEHRENS, T. et al. Digital soil mapping using artificial neural networks. **Journal of Plant Nutrition and Soil Science**, Weinheim, v. 168, p. 1-13, 2005.

BENITES, V. M. et al. Pedotransfer functions for estimating soil bulk density from existing soil survey reports in Brazil. **Geoderma,** Amsterdam, v. 139, p. 90–97, 2007.

BERHONGARAY, G. et al. Land use effects on soil carbon in the Argentine Pampas. **Geoderma,** Amsterdam, v. 192, p. 97–110, 2013.

BERNOUX, M. et al. Brazil's soil carbon stocks. **Soil Science Society of America Journal,** Madison, v. 66, p. 888–896, 2002.

BERTOL, I. et al. Aspectos financeiros relacionados às perdas de nutrientes por erosão hídrica em diferentes sistemas de manejo do solo. **Revista Brasileira de Ciência do Solo,** Viçosa, v. 31, p. 133–142, 2007.

BIRKELAND, P. W. **Soils and Geomorphology**. 3rd ed. New York: Oxford University Press, 1999. 430 p.

BLISS, N. B. et al. Distribution of Soil Organic Carbon in the Conterminous United States. In: HARTEMINK, A.E.; MCSWEENEY, K. (Ed.). **Soil Carbon, Progress in Soil Science**. Dordrecht: Springer, 2014. p. 85–93.

BOCKHEIM, J. G. Solution and use of chronofunctions in studying soil development. **Geoderma**, Amsterdam, v. 24, p. 71-85, 1980.

BODDEY, R. M. et al. Carbon accumulation at depth in Ferralsols under zero-till subtropical agriculture. **Global Change Biology**, Oxford, v. 16, p. 784–795, 2010.

BÖHNER, J.; SELIGE, T. Spatial prediction of soil attributes using terrain analysis and climate regionalisation. **Göttinger Geographische Abhandlungen**, Göttingen, v. 115, p. 13–28, 2006.

BONFATTI, B. R. et al. Digital mapping of soil carbon in a viticultural region of Southern Brazil. **Geoderma**, Amsterdam, v. 261, p. 204-221, 2016.

BOTTON, M. et al. Efeito da cobertura vegetal sobre a pérola-da-terra (Hemiptera: Margarodidae) na cultura da videira. **Acta Scientiarium Agronomy**, Maringá, v. 32, p. 681–684, 2010.

BREIMAN, L. Random Forest. **Machine Learning**, Dordrecht*,* v. 45, p. 5-32, 2001.

BRETÓ, C. et al. Time Series Analysis via Mechanistic Models. **The Annals of Applied Statistics**, Bethesda, v. 3, n. 1, p. 319-348, 2009.

BROGNIEZ, D. et al. Topsoil Organic Carbon Map of Europe. In: HARTEMINK, A.E.; MCSWEENEY, K. (Ed.), **Soil Carbon, Progress in Soil Science**. Dordrecht: Springer, 2014. p. 393–405.

BUNGE, M. How does it work? The Search for Explanatory Mechanisms. **Philosophy of the Social Sciences**, Toronto, v. 34, n. 2, p.182-210, 2004.

BUOL, S. W. et al. **Soil Genesis and Classification**. 6th ed. Nova Jersey: John Wiley & Sons, 2011. 543 p.

CARSON, M. A.; KIRKBY, M. J. **Hillslope form and process**. London: Cambridge University Press, 1972. 484 p.

CATANI, F.; SEGONI, S.; FALORNI, G. An empirical geomorphology-based approach to the spatial prediction of soil thickness at catchment scale. **Water Resources Research**, Hoboken, v. 46, p. 1-15, 2010.

CERRI, C. C.; ANDREUX, F. Changes in organic carbon content in Oxisols cultivated with sugar cane and pasture, based on 13C natural abundance measurement. In: TRANSACTIONS 14th INTERNATIONAL CONGRESS OF SOIL SCIENCE, 1990, Kyoto. **Anais**… Kyoto, 1990. v. IV, p. 98–103.

CERRI, C. E. P. et al. Predicted soil organic carbon stocks and changes in the Brazilian Amazon between 2000 and 2030. **Agriculture, Ecosystems & Environment**, New York, v. 122, p. 58–72, 2007.

CHAGAS, C. D. S. et al. Spatial prediction of soil surface texture in a semiarid region using random forest and multiple linear regressions. **Catena**, Amsterdam, v. 139, p. 232-240, 2016.

CHEN, A.; DARBON, J.; MOREL, J. M. Landscape evolution models: A review of their fundamental equations. **Geomorphology**, Amsterdam, v. 219, p. 68–86, 2014.

CHENG, X. F. et al. Using GIS spatial distribution to predict soil organic carbon in subtropical China. **Pedosphere,** Nanjing, v. 14, p. 425-431, 2004.

COGO, N. P.; LEVIEN, R.; SCHWARZ, R. A. Perdas de solo e água por erosão hídrica influenciadas por métodos de preparo, classes de declive e níveis de fertilidade do solo. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 27, p. 743–753, 2003.

COHEN, S. et al. The mARM spatially distributed soil evolution model: a computationally efficient modeling framework and analysis of hillslope soil surface organization. **Journal of Geophysical Research**, Hoboken, v. 114, p. 1-15, 2009.

COHEN, S. et al. The mARM3D spatially distributed soil evolution model: three-dimensional model framework and analysis of hillslope and landform responses. **Journal of Geophysical Research**, Hoboken, v. 115, p. 1-16, 2010.

COLLARD, F. et al. Spatial prediction of soil organic carbon at different depths using digital soil mapping. In: ARROUAYS, D. et al. (Ed). **GlobalSoilMap – Basis of the global spatial information systems**. Boca Raton: CRC Press, 2014. p. 181–184.

CONANT, R.; PUASTIAN, K.; ELLIOT, E. Grassland management and conversion into grassland: effects on soil carbon. **Ecological Applications**, Washington, v. 11, p. 343-355, 2001.

CPRM – SERVIÇO GEOLÓGICO DO BRASIL. **Mapa Geológico do Estado do Rio Grande do Sul**. Brasília, 2006. Escala 1:750.000.

DIDONÉ, E. J.; MINELLA, J. P. G.; MERTEN, G. H. Quantifying soil erosion and sediment yield in a catchment in southern Brazil and implications for land conservation. **Journal of Soils and Sediments**, Heidelberg, v**.** 15, p. 2334–2346, 2015.

DIETRICH, H.; BÖHNER, J. Cold air production and flow in a low mountain range landscape in Hessia. In: BÖHNER, J.; BLASCHKE, T.; MONTANARELLA, L. (Ed.). **SAGA—seconds out**. Hamburg: University of Hamburg, 2008. p. 37- 48.

DIETRICH, W. E. et al. A process-based model for colluvial soil depth and shallow landsliding using digital elevation data. **Hydrological Processes**, Chichester, v. 9, p. 383-400, 1995.

DIETRICH, W. E. et al. Geomorphic Transport Laws for Predicting Landscape form and Dynamics. In: WILCOCK, P.R.; IVERSON, R.M. (Ed). **Prediction in Geomorphology**. Washington: American Geophysical Union, 2013. p. 103–132.

DIETRICH, W. E.; PERRON, J. T. The search for a topographic signature of life. **Nature**, London, v. 439, p. 411–418, 2006.

DON, A.; SCHUMACHER, J.; FREIBAUER, A. Impact of tropical land-use change on soil organic carbon stocks – a meta-analysis. **Global Change Biology**, Malden, v. 17, p. 1658–1670, 2011.

DORJI, T. et al. Digital soil mapping of soil organic carbon stocks under different land use and land cover types in montane ecosystems, Eastern Himalayas. **Forest Ecology and Management**, Amsterdam, v. 318, p. 91-102, 2014.

ELLERT, B. H.; BETTANY, J. R. Calculation of organic matter and nutrients stored in soils under contrasting management regimes. **Canadian Journal of Soil Science**, Ottawa, v. 75, p. 529–538, 1995.

ELLIOTT, E.T. Aggregate Structure and Carbon, Nitrogen, and Phosphorus in Native and Cultivated Soils. **Soil Science Society of America Journal,** Madison, v. 50, p. 627–633, 1986.

EMBRAPA. **Normal Climatológica:** Estação Agroclimática da Embrapa Uva e

Vinho, Bento Gonçalves, RS. Período de 1961 a 1990. [Bento Gonçalves, RS]: Embrapa, 2008.

EMBRAPA. Centro Nacional de Pesquisa de Solos. **Sistema Brasileiro de Classificação de Solos**. 3. ed. Brasília, 2013. 353 p.

EMBRAPA Uva e Vinho. **Boletim Meteorológico – Bento Gonçalves**. 2017. Disponível em: <https://www.embrapa.br/uva-e-vinho/dados-meteorologicos/bento-goncalves>. Acesso em: 26 fev. 2017.

FLORES, C. A. et al. **Os solos do Vale dos Vinhedos**. Brasília: Embrapa, 2012. 176 p.

FORGES, A. C. et al. A preliminary analysis of topsoil organic carbon contents and stocks spatial distribution in a region of France (Région Centre). In: ARROUAYS, D.et al. (Ed). **GlobalSoilMap – Basis of the global spatial information systems**. Leiden: CRC Press, 2014. p. 197–200.

GAMBILL, D. R. et al. Predicting USCS soil classification from soil property variables using Random Forest. **Journal of Terramechanics**, Amsterdam, v.65, p. 85-92, 2016.

GARCÍA, G. et al. Soil moisture estimation using multi linear regression with terraSAR-X data. **Revista de Teledetección**, Valência, n. 46, p. 9, 2016.

GE, Y. et al. Regression-kriging for characterizing soils with remote sensing data. **Frontiers of Earth Science**, Beijing, v. 5, n. 3, p. 239-244, 2011.

GIASSON, E. et al. Digital soil mapping using multiple logistic regression on terrain parameters in southern Brazil. **Scientia Agricola**, Piracicaba, v. 63, p. 262-268, 2006.

GIFFORD, R. M.; RODERICK, M. L. Soil carbon stocks and bulk density: spatial or cumulative mass coordinates as a basis of expression? **Global Change Biology**, Malden**,** v. 9, p. 1507–1514, 2003.

GOOVAERTS, P. **Geostatistics for Natural Resources Evaluation**. Oxford: Oxford University Press, 1997. 496 p.

GREGO, C. R. et al. Geostatistical analysis for soil moisture content under the no tillage cropping system. **Scientia Agricola**, Piracicaba, v. 63, p. 341-350, 2006.

GRIMM, R. et al. Soil organic carbon concentrations and stocks on Barro Colorado Island — Digital soil mapping using Random Forests analysis. **Geoderma**, Amsterdam, v. 146, p. 102-113, 2008.

GUIMARÃES, R. Z. et al. Espacialização da perda solo por erosão laminar na microbacia do Rio Campinas, Joinville SC. **RA'E GA**, Curitiba, v. 23, p. 534–554, 2011.

GÜNTHER, F.; S. FRITSCH. Neuralnet: Training of Neural Networks. **The R Journal,** Vienna, v. 2, p. 30-38, 2010.

GUO, L. B.; GIFFORD, R. M. Soil carbon stocks and land use change: a meta analysis. **Global Change Biology**, Malden**,** v. 8, p. 345–360, 2002.

HAAN, C. T.; BARFIELD, B. J.; HAYES, J. C. **Design Hydrology and Sedimentology for Small Catchments**. London: Academic Press, 1994. 588 p.

HAIR, J. et al. **Multivariate Data Analysis**. 7th ed. Harlow: Pearson Education Limited, 2014. 816 p.

HAN, J.; KAMBER, M; PEI, J. **Data Mining:** Concepts and Techniques. 3rd ed. Waltham: Elsevier, 2012. 703 p.

HARTEMINK, A. E.; MCSWEENEY, K. **Soil carbon**. Dordrecht: Springer, 2014. 506 p.

HARTEMINK, A. E. Soil chemical and physical properties as indicators of sustainable land management under sugar cane in Papua New Guinea. **Geoderma**, Amsterdam, v. 85, p. 283–306, 1998.

HARTEMINK, A. E. **Soil fertility decline in the tropics:** with case studies on plantations. Wallingford: CABI, 2003. 360 p.

HEIMSATH, A. M. et al. The Soil Production Function and Landscape Equilibrium. **Nature**, London, v. 388, p. 358–361, 1997.

HEIMSATH, A. M. et al. Cosmogenic nuclides, topography, and the spatial variation of soil depth. **Geomorphology**, Amsterdam, v. 27, p. 151–172, 1999.

HEIMSATH, A. M. et al. Soil production on a retreating escarpment in southeastern Australia. **Geology**, Boulder, v. 28, p. 787–790, 2000.

HEIMSATH, A. M. et al. Late Quaternary erosion in southeastern Australia: a field example using cosmogenic nuclides. **Quaternary International**, Amsterdam, v. 83-85, p. 169–185, 2001a.

HEIMSATH, A. M. et al. Stochastic Processes of Soil Production and Transport : Erosion Rates , Topographic Variation and Cosmogenic Nuclides in the Oregon Coast Range. **Earth Surface Processes and Landforms**, Chichester, v. 26, p. 531–552, 2001b.

HEIMSATH, A. M.; FURBISH, D. J.; DIETRICH, W. E. The illusion of diffusion: Field evidence for depth-dependent sediment transport. **Geology,** Boulder, v. 33, p. 949–952, 2005.

HENGL, T. **A Practical Guide to Geostatistical Mapping**. Amsterdam: Lulu, 2009. 291 p.

HENGL, T. et al. Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions. **Plos one**, San Francisco, v. 10, n. 6, p. 1-26, 2015.

HENLEY, S. **Nonparametric Geostatistics**. Dordrecht: Springer Netherlands, 1981. 145 p.

HOWARD, A. D., KERBY, G. Channel changes in badlands. **Geological Society of America Bulletin**, Boulder, v. 94, p. 739–752, 1983.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Folha SH.22 Porto Alegre e parte das Folhas SH.21 Uruguaiana e SI.22 Lagoa Mirim. **Levantamento de Recursos Naturais**. Rio de Janeiro, 1986.

ISAAKS, E.H.; SRIVASTAVA, R.M. **An introduction to applied geostatistics**. Oxford: Oxford University Press, 1989. 592 p.

JANTALIA, C. P. et al. Tillage effect on C stocks of a clayey Oxisol under a soybean-based crop rotation in the Brazilian Cerrado region. **Soil Tillage Research**, Amsterdam, v. 95, p. 97–109, 2007.

JENNY, H. **Factors of Soil Formation:** A System of Quantitative Pedology. New York: McGraw-Hill, 1941. 264 p.

JOBBÁGY, E. G.; JACKSON, R. B. The vertical distribution of soil organic carbon and its relation to climate and vegetation. **Ecological Applications**, Washington, v. 10, p. 423–436, 2000.

KIRKBY, M. J. Hillslope process-response models based on the continuity equation. **Institute of British Geographers Special Publications**, London, v. 3, p. 15–30, 1969.

KIRSTEN, M. et al. Stocks of soil organic carbon in forest ecosystems of the Eastern Usambara Mountains, Tanzania. **Catena**, Amsterdam, v. 137, p. 651-659, 2015.

KRÖSE, B.; VAN DER SMAGT, P. **An introduction to Neural Networks**. 8th ed. Amsterdam: University of Amsterdam, 1996. 135 p.

KUHN, M. et al. **Cubist models for regression**, 2012. Disponível em: <https://cran.r-project.org/web/packages/Cubist/vignettes/cubist.pdf>. Acesso em: 06 jan. 2017.

LACOSTE, M. et al. High resolution 3D mapping of soil organic carbon in a heterogeneous agricultural landscape. **Geoderma**, Amsterdam, v. 213, p. 296–311, 2014.

LAL, R. Forest soils and carbon sequestration. **Forest Ecology and Management**, Amsterdam, v. 220, p. 242–258, 2005.

LAL, R. Enhancing crop yields in the developing countries through restoration of the soil organic carbon pool in agricultural lands. **Land Degradation & Developement**, Malden, v. 17, p. 197–209, 2006.

LANE, P. W. Generalized linear models in soil science. **European Journal of Soil Science**, Bedford, v. 53, p. 241-251, 2002.

LARK, R. M. Towards soil geostatistics. **Spatial Statistics**, Amsterdam, v. 1, p. 92-99. 2012.

LEE, J. et al. Determining soil carbon stock changes: Simple bulk density corrections fail. **Agriculture, Ecosystems & Environment**, New York**,** v. 134, p. 251–256, 2009.

LENTZSCH, P. et al. Application of multiple regression and neural network approaches for landscape-scale assessment of soil microbial biomass. **Soil Biology and Biochemistry**, Amsterdam, v. 37, n. 9, p.1577-1580, 2005.

LIAW, A.; M. WIENER. **Classification and Regression by random Forest**. RNews, v.2/3, 2002. Disponível em: <https://cran.r-project.org/doc/Rnews/Rnews_2002-3.pdf>. Acesso em: 10 jan. 2017.

LIBOHOVA, Z. et al. Mineralizable Soil Organic Carbon Dynamics in Corn-Soybean Rotations in Glaciated Derived Landscapes of Northern Indiana. In: HARTEMINK, A.E.; MCSWEENEY, K. (Ed.). **Soil Carbon**. Dordrecht: Springer, 2014. p. 259–269.

LIEß, M. et al. Uncertainty in the spatial prediction of soil texture: Comparison of regression tree and Random Forest models. **Geoderma**, Amsterdam, v. 170, p. 70-79, 2012.

LIN, L. Concordance Correlation Coefficient to Evaluate Reproducibility. **Biometrics**, Washington, v. 45, p. 255-268, 1989.

MACMILLAN, R. A. **LandMapR Software Toolkit - C++ Version**: User's manual. Edmonton: LandMapper Environmental Solutions Inc., 2003. 110 p.

MALONE, B. **Use R for Digital Soil Mapping**. Sydney: University of Sydney, 2013. 217 p.

MALONE, B. et al. Mapping continuous depth functions of soil carbon storage and available water capacity. **Geoderma**, Amsterdam, v. 154, p. 138–152, 2009.

MALONE, B., MCBRATNEY, A. B., MINASNY, B. Empirical estimates of uncertainty for mapping continuous depth functions of soil attributes. **Geoderma**, Amsterdam, v. 160, p. 614–626, 2011.

MARTINS, S. G. et al. Perdas de solo e água por erosão hídrica em sistemas florestais na região de Aracruz (ES). **Revista Brasileira de Ciência do Solo**, Viçosa, v. 27, p. 395–403, 2003.

MCBRATNEY, A. B. et al. On digital soil mapping. **Geoderma**, Amsterdam, v. 117, p. 3-52, 2003.

MENDONÇA-SANTOS, M. L. et al. Digital Soil Mapping of Topsoil Organic Carbon Content of Rio de Janeiro State, Brazil. In: BOETTINGER, D.J.L. et al. (Ed.). **Digital Soil Mapping, Progress in Soil Science**. Dordrecht: Springer Netherlands, 2010. p. 255–266.

MENDONÇA-SANTOS, M. L.; SANTOS, H. G. The State of the art of Brazilian Soil Mapping and Prospects for Digital Soil Mapping. In: LAGACHERIE, P.; MCBRATNEY, A.B.; VOLTZ, M. (Ed.). **Digital Soil Mapping:** An Introductory Perspective. Amsterdam: Elsevier, 2007. p. 39-54.

MINASNY, B. et al. Digital Mapping of Soil Carbon. **Advances in Agronomy**, Amsterdam, v. 118, p. 1–47, 2013.

MINASNY, B. et al. Resolving the integral connection between pedogenesis and landscape evolution. **Earth-Science Reviews**, Amsterdam, v. 150, p. 102-120, 2015.

MINASNY, B.; MCBRATNEY, A. B. A rudimentary mechanistic model for soil formation and landscape development II: a two-dimensional model incorporating chemical weathering. **Geoderma**, Amsterdam, v. 103, p. 161-179, 2001.

MINASNY, B.; MCBRATNEY, A. B. Jenny, PCA and Random Forests. **Pedometron**, Sydney, v. 33, p. 10–13, 2013.

MINASNY, B.; MCBRATNEY, A. B. Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. **Chemometrics and Intelligent Laboratory Systems**, Amsterdam, v. 94, n. 1, p. 72-79, 2008.

MINELLA, J. P. G.; WALLING, D. E.; MERTEN, G. H. Establishing a sediment budget for a small agricultural catchment in southern Brazil, to support the development of effective sediment management strategies. **Journal of Hydrology**, Amsterdam, v. 519, p. 2189–2201, 2014.

MITASOVA, H. et al. GIS-Based Soil Erosion Modeling. In: SHRODER, J.F.; BISHOP; M.P. (Ed.). **Treatise on Geomorphology**: Remote Sensing and GIScience in Geomorphology. San Diego: Academic Press, 2013, v. 3, p. 228–258.

NICÓTINA, L. et al. Hydrologic controls on equilibrium soil depths. **Water Resources Research**, Hoboken, v. 47, p. 1–11, 2011.

NIEDER, R.; BENBI, D. K. **Carbon and Nitrogen in the Terrestrial Environment**. Dordrecht: Springer, 2008. 430 p.

NIKIFOROFF, C.C. Weathering and soil evolution. **Soil Science**, Filadélfia, v. 67, p. 219-230, 1949.

NUNES, J., et al. A data mining approach to improve multiple regression models of soil nitrate concentration predictions in Quercus rotundifolia montados (Portugal). **Agroforestry Systems**, Dordrecht, v. 84, n. 1, p. 89-100, 2012.

O'BRIEN, S. L.; JASTROW, J. D. Physical and chemical protection in hierarchical soil aggregates regulates soil carbon and nitrogen recovery in restored perennial grasslands. **Soil Biology and Biochemistry**, Amsterdam, v. 61, p. 1–13, 2013.

O'CALLAGHAN, J. F.; MARK, D. M. The extraction of drainage networks from digital elevation data. **Computer Vision, Graphics, and Image Processing**, Amsterdam, v. 28, p. 323-344, 1984.

OLIVEIRA, O. L. P. et al. **Manejo do solo e da cobertura verde em videiras visando sustentabilidade**. Bento Gonçalves: EMBRAPA, 2004. (Comunicado Técnico) Disponível em: <http://www.cnpuv.embrapa.br/publica/comunicado/cot055.pdf>. Acesso em: 15

jul. 2016.

PADARIAN, J.; PÉREZ-QUEZADA, J.; SEGUEL, O. Modelling the distribution of organic carbon in the soils of Chile. In: MINASNY, B.; MALONE, B.P.; MCBRATNEY, A.B. (Ed). **Digital Soil Assessments and Beyond**. Boca Raton: CRC Press, 2012. p. 329–333, 2012.

PECKHAM, S. D. Fluvial landscape models and catchment-scale sediment transport. **Global and Planetary Change**, Amsterdam, v. 39, p. 31–51, 2003.

PEETERS, I. et al. Reconstructing ancient topography through erosion modelling. **Geomorphology**, Amsterdam, v. 78, p. 250–264, 2006.

PELLETIER, J. D.; RASMUSSEN, C. Geomorphically based predictive mapping of soil thickness in upland watersheds. **Water Resources Research**, Hoboken**,** v. 45, 2009.

PENG, Y. et al. Modeling Soil Organic Carbon at Regional Scale by Combining Multi-Spectral Images with Laboratory Spectra. **Plos one**, San Francisco, v. 10, p. 1-22, 2015.

PHILLIPS, J. D. The convenient fiction of steady-state soil thickness. **Geoderma**, Amsterdam, v. 156, p. 389–398, 2010.

PHILLIPS, J. D.; MARION, D. A. Biomechanical effects of trees on soil and regolith: beyond treethrow. **Association of American Geographers**, Washington, v. 96, p. 233–247, 2006.

PRUSKI, F. F. **Conservação de solo e água:** práticas mecânicas para o controle da erosão hídrica. 2. ed. Viçosa: UFV, 2009. 279 p.

QIU, Y. et al. Spatial prediction of soil moisture content using multiple-linear regressions in a gully catchment of the Loess Plateau, China. **Journal of Arid Environments**, Amsterdam, v. 74, p. 208-220, 2010.

QUINLAN, J. R. Combining instance based and model based learning. In: TENTH INTERNATION CONFERENCE ON MACHINE LEARNING, 1993, Amherst. **Proceedings**… Amherst: University of Massachusetts, 1993. p. 236-243.

QUINLAN, J. R. Learning with continuous classes. In: AUSTRALIAN JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, 1992, Hobart. **Proceedings**… Hobart: World Scientific, 1992. p. 343-348.

ROSS, C. W.; GRUNWALD, S.; MYERS, D. B. Spatiotemporal modeling of soil organic carbon stocks across a subtropical region. **Science of the Total Environment**, Amsterdam, v. 461-462, 149–157, 2013.

RUMPEL, C.; KÖGEL-KNABNER, I. Deep soil organic matter—a key but poorly understood component of terrestrial C cycle. **Plant and Soil,** Gewerbestrasse, v. 338, p. 143–158, 2011.

SACO, P. M.; WILLGOOSE, G. R.; HANCOCK, G. R. Spatial organization of soil depths using a landform evolution model. **Journal of Geophysical Research**

Earth Surface, Washington, v. 111, p. 1-14, 2006.

SACO, P. M.; WILLGOOSE, G. R.; HANCOCK, G. R. Eco-geomorphology of banded vegetation patterns in arid and semi-arid regions. **Hydrology and Earth System Sciences**, Munique, v. 11, p. 1717–1730, 2007.

SANFORD, G. R. Perennial Grasslands Are Essential for Long Term SOC Storage in the Mollisols of the North Central USA. In: HARTEMINK, A.E.; MCSWEENEY, K. (Ed.). **Soil Carbon**. Dordrecht: Springer, 2014. p. 281–288.

SANTOS, H. G. et al. **Sistema Brasileiro de Classificação de Solos**. 2. ed. Rio de Janeiro: Embrapa Solos, 2006. 306 p.

SCHAETZL, R. J.; ANDERSON, S. **Soils:** Genesis and Geomorphology. New York: Cambridge University Press, 2005. 827 p.

SCHMIDHUBER, J. Deep learning in neural networks: An overview. **Neural Networks**, Amsterdam, v. 61, p. 85-117, 2015.

SCHRUMPF, M. et al. Storage and stability of organic carbon in soils as related to depth, occlusion within aggregates, and attachment to minerals. **Biogeosciences**, Gotinga, v. 10, p. 1675–1691, 2013.

SEIDL, M.; DIETRICH, W. The problem of channel erosion into bedrock. **Catena**, Amsterdam, v. 23, p. 101–104, 1992.

SHANKS, L.; MOORE, D.; SANDERS, C. Soil erosion. In: INGELS, C.A.et al. **Cover Cropping in Vineyards:** A Grower's Handboolk. Oakland: University of California, 1998. p. 80–85.

SHRESTHA, D. L.; SOLOMATINE, D. P. Machine learning approaches for estimation of prediction interval for the model output. **Neural Networks**, Amsterdam, v. 19, p. 225–235, 2006.

SIMÓ, I. et al. Modelling Soil Organic Carbon stocks using a detailed soil map in a Mediterranean mountainous area. In: ARROUAYS, D.ET AL. (Ed). **GlobalSoilMap – Basis of the global spatial information systems**. Leiden: CRC Press, 2014. p. 421–427.

SISTI, C. P. J. et al. Change in carbon and nitrogen stocks in soil under 13 years of conventional or zero tillage in southern Brazil. **Soil & Tillage Research**, Amsterdam, v. 76, p. 39–58, 2004.

SNEPVANGERS, J. J. J. C. et al. Soil water content interpolation using spatio-temporal kriging with external drift. **Geoderma**, Amsterdam, v.112, p. 253-271, 2003.

SOIL SURVEY STAFF. **Keys to Soil Taxonomy**. 12th ed. Washington: USDA Natural Resources Conservation Service, 2014. 360 p.

SOLOMATINE, D. P.; SHRESTHA, D. L. A novel method to estimate model uncertainty using machine learning techniques. **Water Resources Research**, Hoboken**,** v. 45, p. 1-16, 2009.

SOUZA, E. DE. et al. Comparing spatial prediction methods for soil property mapping in Brazil. In: ARROUAYS, D. et al. (Ed). **GlobalSoilMap – Basis of the global spatial information systems**. Leiden: CRC Press, 2014. p. 267–271.

SOUZA, Z. M. D. et al. Geoestatística e atributos do solo em áreas cultivadas com cana-de-açúcar. **Ciência Rural**, Santa Maria, v. 40, p. 48-56, 2010.

STOCKMANN, U.; MINASNY, B.; MCBRATNEY, A.B. Quantifying Processes of Pedogenesis, **Advances in Agronomy**, Amsterdam, v. 113, p. 1-71, 2011.

STOCKMANN, U.; MINASNY, B.; MCBRATNEY, A.B. How fast does soil grow? **Geoderma**, Amsterdam, v. 216, p. 48–61, 2014.

SUN, W., et al. Analysis and prediction of soil properties using local regression-kriging. **Geoderma**, Amstermda, v. 171–172, p. 16-23, 2012.

TARBOTON, D. G. A new method for the determination of flow directions and upslope areas in grid digital elevation models. **Water Resources Research**, Hoboken, v. 33, p. 309–319, 1997.

TEMME, A. J. A. M.; VANWALLEGHEM, T. Lorica – A new model for linking landscape and soil profile evolution: Development and sensitivity analysis. **Computers & Geosciences**, Amsterdam, v. 90, p. 131-143, 2015.

THOMPSON, J. A.; KOLKA, R. K. Soil carbon storage estimation in a forested watershed using quantitative soil-landscape modeling. **Soil Science Society of America,** Madison, v. 69, p. 1086–1093, 2005.

TIWARI, S. K., et al. Prediction Modeling and Mapping of Soil Carbon Content Using Artificial Neural Network, Hyperspectral Satellite Data and Field Spectroscopy. **Scientific Research: An Academic Publisher**, Wuhan, v. 4, p. 63-72, 2015.

TORNQUIST, C. G. et al. Soil Organic Carbon Stocks of Rio Grande do Sul, Brazil. **Soil Science Society of America,** Madison, v. 73, p. 975-982, 2009a.

TORNQUIST, C. G.; MIELNICZUK, J.; CERRI, C. E. P. Modeling soil organic carbon dynamics in Oxisols of Ibirubá (Brazil) with the Century Model. **Soil Tillage Research**, Amsterdam, v. 105, p. 33–43, 2009b.

TRANGMAR, B. B.; YOST, R. S.; UEHARA, G. Application of geostatistics to spatial studies of soil properties. **Advances in Agronomy,** San Diego, v. 38, 1985. p. 45-94.

TRANTER, G. et al. Building and testing conceptual and empirical models for predicting soil bulk density. **Soil Use and Management**, Bedford, v. 23, p. 437–443, 2007.

VANWALLEGHEM, T. et al. A quantitative model for integrating landscape evolution and soil formation. **Journal of Geophysical Research Earth Surface**, Washington, v. 118, p. 331–347, 2013.

VANWALLEGHEM, T. et al. Spatial variability of soil horizon depth in natural loess-derived soils. **Geoderma**, Amsterdam, v. 157, p. 37-45, 2010.

VASQUES, G.M. et al. Regional modelling of soil carbon at multiple depths within a subtropical watershed. **Geoderma,** Amsterdam, v. 156, p. 326–336, 2010.

VISCARRA ROSSEL, R. A. et al. Baseline estimates of soil organic carbon by proximal sensing: Comparing design-based, model-assisted and model-based inference. **Geoderma**, Amsterdam, v. 265, p. 152-163, 2016.

WARING, C. et al. Is percent “Projected Natural Vegetation Soil Carbon”a useful indicator of soil condition? In: HARTEMINK, A.; MCSWEENEY, K. (Ed.). **Soil Carbon**. Dordrecht: Springer, 2014. p. 219–227.

WASIGE, J.E. et al. Contemporary land use/land cover types determine soil organic carbon stocks in south-west Rwanda. **Nutrient Cycling in Agroecosystems**, Dordrecht, v. 100, p. 19–33, 2014.

WHIPPLE, K.X. Fluvial landscape response time: how plausible is steady-state denudation? **American Journal of Science**, Stanford, v. 301, p. 313-325, 2001.

WIESMEIER, M. et al. Estimation of total organic carbon storage and its driving factors in soils of Bavaria (southeast Germany). **Geoderma Regional**, Amsterdam, v. 1, p. 67–78, 2014.

WIESMEIER, M., et al. Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. **Plant and Soil**, Gewerbestrasse, v. 340, p. 7-24, 2011.

WILLGOOSE, G. **User Manual for SIBERIA (Version 8.30).** Scone, 2005. Disponível em: <http://nova.newcastle.edu.au/vital/access/manager/Repository/uon:12451>. Acesso em: 03 mar. 2016.

WILLGOOSE, G.; BRAS, R. L.; RODRIGUEZ-ITURBE, I. A coupled channel network growth and hillslope evolution model, 1, Theory. **Water Resources Research**, Hoboken, v. 27, p. 1671–1684, 1991.

WILLGOOSE, G. R.; RODRIGUEZ-ITURBE, I. **A physically based channel network and catchment evolution model.** Boston: MIT, 1989. 464 p.

WILLS, S. et al. Overview of the U.S. Rapid Carbon Assessment Project: Sampling Design, Initial Summary and Uncertainty Estimates. In: HARTEMINK, A.E.; MCSWEENEY, K. (Ed.). **Soil Carbon**. Dordrecht: Springer, 2014. p. 95–104.

YAALON, D.H. Soil-forming processes in time and space. In: YAALON, D.H. (Ed.). **Paleopedology:** Origin, Nature and Dating of Paleosols. Jerusalem: International Society of Soil Science/ Israel Universities Press, 1971. p. 29-39.

YU, D.S. et al. Regional patterns of soil organic carbon stocks in China. **Journal of Environmental Management,** Amsterdam, v. 85, p. 680–689, 2007.

ZELL, A. **Simulation Neuronaler Netze**. Bonn: Addison-Wesley Publishing Company, 1994. 624 p.

ZHANG, P.; SHAO, M. Spatial variability and stocks of soil organic carbon in the Gobi Desert of Northwestern China. **Plos One,** San Francisco, v. 9, p. 1-12, 2014.

ZHU, A. X. Mapping soil landscape as spatial continua: The Neural Network Approach. **Water Resources Research**, Hoboken, v. 36, p. 663-677, 2000.

ZORNOZA, R., et al. Evaluation of soil quality using multiple lineal regression based on physical, chemical and biochemical properties. **Science of The Total Environment**, Amsterdam, v. 378, p. 233-237, 2007.