SÉRGIO HENRIQUE GODINHO SILVA

TÉCNICAS E FERRAMENTAS DE MAPEAMETO DIGITAL DE SOLOS APLICADAS ÀS CONDIÇÕES BRASILEIRAS PARA AUXILIAR LEVANTAMENTOS DE SOLOS

LAVRAS – MG
2014

SÉRGIO HENRIQUE GODINHO SILVA

TÉCNICAS E FERRAMENTAS DE MAPEAMETO DIGITAL DE SOLOS APLICADAS ÀS CONDIÇÕES BRASILEIRAS PARA AUXILIAR LEVANTAMENTOS DE SOLOS

Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Programa de Pós-Graduação em Ciência do Solo, área de concentração em Recursos Ambientais e Uso da Terra, para a obtenção do título de Mestre.

Orientador

Dr. Nilton Curi

LAVRAS – MG

2014

Ficha Catalográfica Elaborada pela Coordenadoria de Produtos e Serviços da Biblioteca Universitária da UFLA

Silva, Sérgio Henrique Godinho.
Técnicas e ferramentas de mapeameto digital de solos aplicadas às condições brasileiras para auxiliar levantamentos de solos / Sérgio Henrique Godinho Silva. – Lavras : UFLA, 2014.
102 p. : il.

Dissertação (mestrado) – Universidade Federal de Lavras, 2014.
Orientador: Nilton Curi.
Bibliografia.

1. Pedologia. 2. Geoprocessamento. 3. Lógica fuzzy. I. Universidade Federal de Lavras. II. Título.

CDD – 631.470011

SÉRGIO HENRIQUE GODINHO SILVA

TÉCNICAS E FERRAMENTAS DE MAPEAMETO DIGITAL DE SOLOS APLICADAS ÀS CONDIÇÕES BRASILEIRAS PARA AUXILIAR LEVANTAMENTOS DE SOLOS

(DIGITAL SOIL MAPPING TOOLS AND TECHNIQUES APPLIED TO BRAZILIAN CONDITIONS TO AID SOIL SURVEYS)

Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Programa de Pós-Graduação em Ciência do Solo, área de concentração em Recursos Ambientais e Uso da Terra, para a obtenção do título de Mestre.

APROVADA em 29 de julho de 2014.

Dr. Nilton Curi	UFLA
Dr. Geraldo César de Oliveira	UFLA
Dr. Gilberto Coelho	UFLA

Dr. Nilton Curi
Orientador

Coorientadores
Dr. Phillip Ray Owens
M.Sc. Fausto Weimar Acerbi Júnior

LAVRAS – MG
2014

Aos meus pais, Taísa e Walter, avós, Maria José e Danilo, e Tia, Tânia.

DEDICO.

AGRADECIMENTOS

Agradeço a Deus pela proteção, saúde, paz e suporte em toda a minha vida.

À minha família, pelo amor incondicional, apoio, sempre me desejar o melhor, pela compreensão e por ser minha maior motivação.

Ao Prof. Nilton Curi, pelo apoio, amizade, confiança, ensinamentos, incentivo e por ser um exemplo durante esses anos no DCS/UFLA.

Ao coorientadores Prof. Phillip Owens e Prof. Fausto, pela confiança, grande ajuda, paciência, ensinamentos e amizade.

Ao CNPq pela concessão da bolsa de Mestrado, incentivando os estudos, e à CAPES e FAPEMIG por outros auxílios financeiros.

Ao Prof. João José, pela paciência, apoio, ensinamentos e incentivo no início do uso de programas para mapeamento de solos.

Aos Professores do DCS/UFLA, em especial Profa. Michele e Prof. Geraldo, pela amizade e ensinamentos.

Aos meus grandes amigos de toda a vida, dos quais não citarei nomes, pois não caberiam nesta página.

Aos meus amigos da turma 2008/01 de Eng. Florestal, um grande exemplo de amizade e união que continuam mesmo após o término do curso.

Aos meus amigos do DCS, Maíra, Elen, Michele, Walbert, Vladimir, Bruno, Elidiane, Leandro, Giovana, Eliete e Thaís, pela amizade e ensinamentos ao longo desses anos. E a Zélio, Samara, Bombinha, Eduardo, Érika, Anita, Soraya, Aryane, Amaral, Eduane, Dirce e Damyane.

Obrigado a todos que me ajudaram de alguma forma nesta jornada.

RESUMO

Avanços tecnológicos têm proporcionado alternativas para diminuir problemas comuns enfrentados em levantamentos de solos antes, durante e após os trabalhos de campo. Algumas ferramentas consideram o relevo, representado por Modelos Digitais de Elevação, como um adequado fator para ser associado à variabilidade das classes e propriedades dos solos ao longo da paisagem. Entretanto, ajustes regionais devem ser realizados para se atingir melhor qualidade dos mapas de solos finais devido à variabilidade espacial dos mesmos. Os objetivos desse trabalho contemplam: (I) uma revisão bibliográfica sobre algumas ferramentas de mapeamento digital de solos recentemente criadas, dando ênfase a sua importância para países em desenvolvimento; (II) criar um mapa da profundidade do *solum* de uma microbacia hidrográfica de cabeceira de Minas Gerais, Brasil, a partir da associação entre conhecimento de campo, lógica fuzzy e atributos de terreno, e fazer sua validação em campo usando o sistema de amostragem Hipercubo Latino Condicionado; e (III) comparar o mapeamento de solos convencional com o digital em relação ao ganho de informações proporcionadas pela possibilidade de criar mapas contínuos de solos e seus atributos. A revisão foi realizada através de análises de diferentes técnicas e ferramentas de mapeamento de solos recentemente publicadas em revistas científicas com o objetivo de reuni-las em um único documento, facilitar o seu uso e destacar a sua importância para tais atividades, principalmente para países em desenvolvimento. O mapa da profundidade do *solum* foi criado para uma microbacia hidrográfica de cabeceira no sul de Minas Gerais, devido à área apresentar grande importância ambiental relativa à qualidade e suprimento de água para usinas hidrelétricas, fatores que estão relacionados à profundidade do *solum*. Alguns dados de campo envolvidos na validação do mapa da profundidade do *solum* foram usados para realizar a comparação entre mapas gerados pelos métodos convencional e digital em termos de quantidade de informações apresentadas. O emprego de conhecimento de campo associado às ferramentas de mapeamento apresentaram resultados adequados, tendo o mapa da profundidade do *solum* acurácia de 80% e índice kappa de 0,161. O mapa da profundidade do *solum* criado pelo método digital apresentou maior acurácia e detalhes que o gerado pelo método convencional. Essas ferramentas e dados disponíveis têm alto potencial para serem usadas para complementar o levantamento e mapeamento de solos, principalmente em países onde são comuns dificuldades para a realização de trabalhos de campo intensivos.

Palavras-chave: Pedologia. Geoprocessamento. Lógica Fuzzy.

ABSTRACT

The advances in technology have provided alternatives to diminish common issues faced by soil surveyors prior, during and after the field work. Some tools consider the relief, obtained from Digital Elevation Models, an adequate factor to be associated with soil types and properties variability along the landscape. However, regional adjustments have to be made for a better quality of the final soil maps due to soils spatial variability. The objectives of this work contemplate: (I) a review discussing some of the recently created soil mapping tools, emphasizing their importance for developing countries; (II) to create a solum depth map for a watershed in southern Minas Gerais, Brazil, associating expert knowledge with fuzzy logic and terrain derivatives, and validate it in the field by using Conditioned Latin Hypercube sampling scheme; and (III) to compare the conventional to digital soil mapping in regard to the gain of information provided by the possibility of creating continuous maps of soil types and properties. The review was made after analyses of different soil mapping tools and techniques recently published in scientific journals in order to join them into a comprehensive document, to facilitate their use, and to highlight their importance for such activities, mainly for developing countries. The solum depth map was created for a headwater watershed in south of Minas Gerais State, Brazil, due to the mentioned area presents environmental importance for water quality and supply for hydroelectric power plants, factors that are related to solum depth. Some of the field data used to validate the solum depth map were used to make a comparison between solum depth maps created by conventional and digital mapping in terms of amount of provided information. The use of expert knowledge associated with mapping tools presented adequate results, being the solum depth map accuracy of 80% and kappa index of 0.616. Solum depth map created by digital mapping tools presented higher accuracy and details in comparison with the one created by conventional mapping method. Those available tools and data have a high potential to be used to complement soil surveys and mapping, mainly in countries where difficulties to accomplish an intensive field work are not rare.

Index terms: Pedology. Geoprocessing. Fuzzy Logic.

SUMÁRIO

PRIMEIRA PARTE 10
1 INTRODUCÃO 11
1.1 Introdução geral 11
1.2 Objetivos 12
REFERÊNCIAS 13
SEGUNDA PARTE - ARTIGOS 14
2. ARTIGO 1. Digital soil mapping approach based on fuzzy logic and expert knowledge 15
2.1 INTRODUCTION 16
2.2 EXPERT SYSTEMS 18
2.3 FUZZY LOGIC 19
2.4 SOLIM (SOIL-LAND INFERENCE MODEL) AND ARCSIE (SOIL INFERENCE ENGINE) 21
2.4.1 Ruled-based reasoning with ArcSIE (RBR) 22
2.4.2 Case-based Reasoning (CBR) with ArcSIE 28
2.5 CONDITIONED LATIN HYPERCUBE SAMPLING SCHEME 30
2.6 ACQUIRING INFORMATION FROM EXISTING SOIL MAPS FOR SOIL DATA TRANSFERABILITY 32
2.6.1 SoLIM Knowledge Miner 33
2.6.2 Boxplots 36
2.7 COMMONLY USED ACCURACY EVALUATION METHODS 37
2.8 FINAL CONSIDERATIONS 40
REFERENCES 42
3. ARTIGO 2 A technique for low cost soil mapping and validation using expert knowledge on a watershed in Minas Gerais, Brazil 48
3.1 INTRODUCTION 49
3.2 MATERIALS AND METHODS 52
3.2.1 Study area 52
3.2.2 Data used for digital soil mapping 53

3.2.3 Methods for creating inference maps 54
3.2.4 Soil Sampling for validation 59
3.3 RESULTS AND DISCUSSION 61
3.4 SUMMARY AND CONCLUSIONS 71
REFERENCES 73
4. ARTIGO 3 Solum depth spatial prediction comparing conventional with knowledge-based digital soil mapping approaches 79
4.1 INTRODUCTION 79
4.2. MATERIAL AND METHODS 81
4.2.1 Study area characterization 81
4.2.2 Conventional soil mapping approach 82
4.2.3 Knowledge-based digital soil mapping approach 83
4.2.4 Assessment of accuracy of solum depth prediction 87
4.3 RESULTS AND DISCUSSION 87
4.3.1 Conventional soil survey 87
4.3.2 Knowledge-based digital soil mapping 89
4.3.3 Solum depth predictive maps 94
4.3.4 Validation and accuracy assessment of the predicted solum depth .. 96
4.4. CONCLUSIONS 98
REFERENCES 99

PRIMEIRA PARTE

AVANÇOS TECNOLÓGICOS AUXILIARES AO MAPEAMENTO DIGITAL DE SOLOS

1 INTRODUCÃO

1.1 Introdução geral

A identificação e classificação de solos são realizadas a partir de levantamentos de solos, atividades que, segundo Resende et al. (2014), fornecem informações insubstituíveis para a estratificação de ambientes. Através do levantamento de solos, diversas classes de solos são identificadas e diferenciadas pela morfologia e características física, química, biológica e mineralógica (Motta et al., 2001). A partir dessas informações, a representação da distribuição espacial de cada classe de solo é feita através de mapas de solos.

Entretanto, pelo fato de levantamentos de solos serem procedimentos que demandam tempo devido à grande quantidade de amostras a serem coletadas e que requerem certo investimento financeiro (McBratney et al. 2003), poucos são os mapas de solos em escala detalhada na grande maioria dos países. Para o Brasil isso não é diferente. Coelho e Giasson (2010) apontam que os mapas para a maior parte do território nacional são oriundos do projeto Radam Brasil, iniciado em 1986, e são pouco detalhados (escala 1:1.000.000), o que não permite o planejamento ao nível de propriedades rurais e bacias hidrográficas. Levantamentos mais detalhados carecem de recursos financeiros, porém as vantagens que proporcionam para a tomada de decisões envolvendo uso e manejo sustentável do solo são enormes.

Iwashita et al. (2012) destacam que, durante os levantamentos de solos, as restrições financeiras e de tempo disponível tipicamente limitam a amostragem ao longo de toda a área. Por outro lado, essa limitação motiva a busca de métodos alternativos para a realização dos levantamentos (McBratney et al., 2003), uma vez que o número reduzido de amostras pode diminuir a acurácia dos mapas e, consequentemente, levar a tomadas de decisões incompatíveis com a realidade.

Diante desses fatos, avanços tecnológicos, tais como computadores cada vez mais potentes para o processamento de dados, imagens de satélite, Modelos Digitais de Elevação (MDEs) e os atributos de terreno deles derivados, e programas de computador voltados ao mapeamento digital de solos estão se tornando mais e mais utilizados e de fácil acesso no mundo todo. Esse conjunto de ferramentas proporciona alternativas para melhorar tanto a eficiência dos trabalhos de campo quanto a qualidade dos mapas de solos finais, especialmente quando essas técnicas são combinadas ao conhecimento de pedólogos que conheçam bem as relações solo-paisagem da área de interesse (conhecimento de campo), unindo, desta forma, uma fonte de conhecimento e experiência a potentes ferramentas para a extrapolação de informações e rápido processamento de dados. Essa associação pode contribuir tanto para a economia de recursos financeiros quanto para a extrapolação de informações com maior confiabilidade para locais não amostrados.

1.2 Objetivos

Os objetivos desse trabalho contemplam: (I) realizar uma revisão bibliográfica sobre algumas ferramentas de mapeamento digital de solos recentemente criadas, dando ênfase na sua importância para países em desenvolvimento; (II) criar um mapa da profundidade do *solum* de uma microbacia hidrográfica de cabeceira de Minas Gerais, Brasil, a partir da associação entre conhecimento de campo, lógica fuzzy e atributos de terreno, e fazer sua validação em campo usando o sistema de amostragem Hipercubo Latino Condicionado; e (III) comparar o mapeamento de solos convencional com o digital em relação ao ganho de informações proporcionadas pela possibilidade de criar mapas contínuos de solos e seus atributos.

REFERÊNCIAS

COELHO, F.B.; GIASSON, E. Métodos para mapeamento digital de solos com utilização de sistema de informação geográfica. **Ciência Rural**, Santa Maria, v.40, n.10, p.2099-2106, out. 2010.

IWASHITA, F.; FRIEDEL, M. J.; RIBEIRO, G. F.; FRASER, S. J. Intelligent estimation of spatially distributed soil physical properties. **Geoderma**, Amsterdã, v.170, p.1-10, jan. 2012.

MCBRATNEY, A.B.; SANTOS, M.L.M.; MINASNY, B. On digital soil mapping. **Geoderma**, Amsterdã, v.117, n.1/2, p.3-52, nov. 2003.

MOTTA, P. E. F. et al. **Levantamento pedológico detalhado, erosão dos solos, uso atual e aptidão agrícola das terras de microbacia piloto na região sob influência do reservatório de Itutinga/Camargos, MG.** Belo Horizonte: CEMIG, 2001. 51 p.

RESENDE, M. et al. **Pedologia:** base para distinção de ambientes. 6. ed. Lavras: UFLA, 2014. 404 p.

SEGUNDA PARTE - ARTIGOS

2. ARTIGO 1. Digital soil mapping approach based on fuzzy logic and expert knowledge

***Artigo nas normas da Revista Ciência e Agrotecnologia.**

ABSTRACT

In Brazil, soil surveys in more detailed scale are still scarce and necessary to more adequately support the decision makers for planning soil and environment activities in small areas. Hence, this review addresses some digital soil mapping techniques that enable faster production of soil surveys, beyond fitting continuous spatial distribution of soil properties into discrete soil categories, in accordance with the inherent complexity of soil variation, increasing the accuracy of spatial information. The technique focused here is knowledge-based in expert systems, under fuzzy logic and vector of similarity. For that, a contextualization of each tool in the soil types and properties prediction is provided, as well as some options of knowledge extraction techniques. Such tools have reduced the inconsistency and costs associated with the traditional manual processes, relying on a relatively low density of soil samples. On the other hand, knowledge-based technique is not automatic, and just as the traditional soil survey, the knowledge of soil-landscape relationships is irreplaceable.

Index terms: digital soil mapping, soil prediction, conditioned Latin hypercube sampling, knowledge miner.

RESUMO

No Brasil, levantamentos de solos em escalas maiores ainda são escassos e necessários para dar apoio mais adequado ao planejamento de atividades relacionadas a solos e ambientes em áreas menores. Em consequência, este trabalho apresenta algumas técnicas de mapeamento digital de solos que permitem a produção mais rápida de levantamentos de solos, além de ajustar a distribuição espacial contínua das propriedades do solo em categorias discretas, de acordo com a complexidade inerente da variabilidade dos mesmos, aumentando a acurácia de informações espaciais. A técnica aqui enfatizada é baseada em sistemas que empregam o conhecimento de um especialista, sob uso de lógica fuzzy e similaridade de vetores. Para isso, é proporcionada a contextualização de cada ferramenta para a predição de classes de solos e suas propriedades, assim como algumas opções de técnicas para aquisição de conhecimentos. Tais ferramentas têm reduzido a inconsistência e custos associados aos tradicionais procedimentos manuais, utilizando uma relativamente baixa densidade de amostragem. Por outro lado, a técnica baseada no conhecimento de especialistas não é automatizada, e, assim como no método tradicional de levantamentos de solos, o conhecimento das relações solo-paisagem é insubstituível.

Termos para indexação: mapeamento digital de solos, predição de solos, amostragem por Hipercubo Latino Condicionado, mineração de conhecimento.

2.1 INTRODUCTION

In Brazil, soil surveys in more detailed scale are still necessary because the lack of information or the small-scale existing maps do not adequately support planning and management of agricultural and environmental projects. Soil surveys or sampling schemes in a more detailed scale are common only in small areas, generally to attend specific projects (MENDONÇA-SANTOS; SANTOS, 2007). Since the traditional soil maps are manually produced, even on a GIS basis, and have as limitation the low speed and high production cost (ZHU et al., 2001), digital soil mapping is viewed as an opportunity to optimize soil mapping, employing more quantitative techniques for spatial prediction (MCBRATNEY et al., 2003), in which the accuracy or uncertainty has been measured and discussed, and that makes the pedologist mental model more explicit. In theory, the basis of predictive soil mapping is similar to traditional soil survey, since it is possible to use knowledge of soil-environment relations to make inferences (SCULL et al., 2003).

Various approaches have been used for fitting quantitative relationships between soil properties or types and their environment, in order to predict them (spatial inference models). The models are divided into data-driven (Pedometry approach) and quantitative soil survey approach (knowledge driven). Pedometry approaches are more quantitative and automatic, mainly based on statistics, geostatistics, machine learning and data mining techniques. A dense scheme of sampling is often required. On the other hand, the knowledge driven approach tries to fit within the conventional soil survey and mapping framework, aiming to effectively utilize the soil scientist's knowledge (SHI et al., 2009).

Soil survey is a paradigm-based science that is based on the application of conceptual soil-landscape models, in which the hypothesis is that the location and distribution of soils in the landscape is predictable (HUDSON, 1992). Such models rely on tacit pedological knowledge, generally acquired by systematic field observation of repeating relationships between soil types or properties and landform position (MACMILLAN et al., 2005). Most of the information about soils is found in soil maps and respective legend or in the mind of the soil surveyor. Hudson (1992) argued that soil survey was deficient for not expressing the scientific knowledge in a more formal and systematic way.

Thus, this review attempts to elucidate the use of expert systems under fuzzy logic and its application for predicting soil types and properties. Expert systems allow the use of existing data or expert knowledge of the pedologist in conjunction with statistical and mathematical approaches to generate soil information. Besides, they allow to fit continuous spatial distribution of soil properties into discrete soil categories, in accordance with the inherent complexity of soil variation, increasing the accuracy of spatial information (ZHU et al., 2001).

2.2 EXPERT SYSTEMS

According to Dale et al. (1989), expert systems consist of ways to harvesting and engineering knowledge, which allow exploiting the information of soil surveyor acquired through experience. Expert knowledge systems try to capture tacit knowledge and integrate it in the predictive model in order to improve it. Dale et al. (1989) delineated the components of an expert system to soil data: a source (e.g. data or environmental variables), an organizer and an information predictor, and a client to use the information. The predictor includes a knowledge-based and an inference engine which operates on the knowledge base. The computer-based knowledge can use the human expert or numerical methods. Such approach is able to exploit soil surveyor knowledge by developing rule-based systems that imitate the surveyor's conceptual model of soil variability (SCULL et al., 2003). The pioneering attempts to apply expert systems in Pedology used the Boolean logic (SKIDMORE et al.,1991), which defines a strict binary decision (true or false, 0 or 1). In terms of soil maps, the soil surveyor has to assign individual soils in the field in only one class (ZHU et al., 2001). The polygons of the maps, also referred to as crisp or Boolean, represent only the distribution of a set of prescribed soil class (central concepts of soils). The same approach is used for soil property maps, where the whole polygon assumes a property value assigned to the mapping unit.

2.3 FUZZY LOGIC

The nature of soil-landscapes are complex, whose changes in soils or properties are often more gradual and continuous, differently to the variation represented by a crisp map (polygon-based) (Figure 1a). There is uncertainty in the boundaries allocation, as well as in the values of the soil properties (LEGROS, 2006) (Figure 1b). Fuzzy logic attempts to represent the uncertainty in the predictor and predicted properties or types, as an alternative that seems more adapted to the imprecise knowledge conveyed by soil surveyors (WALTER et al., 2007), recognizing the concept of partial truth, alternatively to the subjective rigidity imposed on soils.

Instead of a crisp membership (e.g., entirely Red Latosol or Yellow Latosol, Figure 1a), the idea is that the soils in nature rarely fit exactly the classification types to which they are assigned (ZADEH, 1965). Nevertheless, there is a range of optimal values among classes. The concept of belonging to a set has been modified to include partial degrees of membership. The maximum membership is often 1 and represents the central or modal concept, whereas the 0 value expresses no membership. Values in between this range express different degrees of similarity to the central concept.

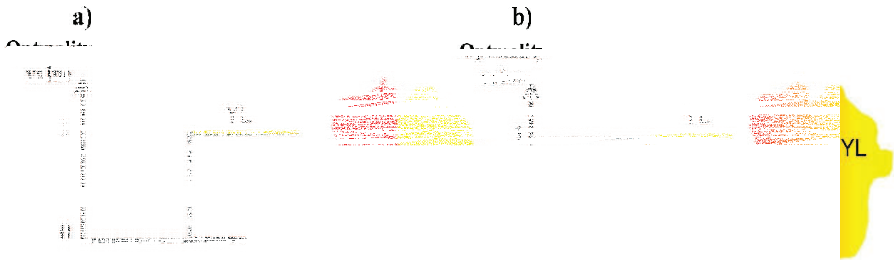

Figure 1 - Lateral distribution of an optimal value under Boolean logic (a) and fuzzy logic (b) related to distribution of Yellow Latosol (YL) and Red Latosol (RL) in the landscape.

Besides the broad application of fuzzy logic in science, Scull et al. (2003) cited two different approaches for soil prediction in a continuous way: the first is based on the fuzzy-k-means classifier, which partitions observations in multivariate space into natural classes. The second is known as the semantic import model, and is used in situations when classification schemes are pre-defined and class limits are relatively well understood. The semantic model is commonly used with expert knowledge and it refers to a data integration concerned with analysis and interpretation of a multi-source spatial data. In geographic analysis, it is frequently required the integration of spatial data with multi-sources (as raster or vector formats, crisp or continuous maps) to answer specific questions about given spatial phenomenon. In this sense, Zhu and Band (1994) presented the first approach which employs knowledge-based semantic data integration, combined with expert system techniques and fuzzy set theory for spatial data integration.

A fuzzy logic based model called similarity vector (ZHU et al., 1997) represents soils at a given location, in which the landscape is perceived as a continuum. The fuzzy logic is used to infer the membership of a soil type from environmental variables, such as parent material, canopy coverage, digital elevation model and its derivative maps. Under fuzzy logic, a soil at a given pixel (*i,j*) is represented by a *n*-element similarity vector: S_{ij}= (S_{ij}^{1} , S_{ij}^{2} ,…, S_{ij}^{k} , ..., S_{ij}^{n}), where *n* is the number of prescribed soil types over the area and S_{ij}^{k} is an index which measures the similarity between the local soil at (*i,j*) to the prescribed soil type *k*. S_{ij}^{k} is soil type *k*. The similarity value is measured according to how close the soil is to the centroid concept (between 1 and 0, as already discussed). The more similar a soil is to a prescribed soil type, the higher its similarity value (fuzzy membership).

This methodology has been successfully applied to generate soil maps (crisp maps) (ZHU; BAND, 1994; ZHU et al., 1996; MCKAY et al., 2010; MENEZES, 2011) and to predict properties in a continuous way, as depth of A horizon (ZHU et al., 1997), solum depth (QUINN et al., 2005; LIBOHOVA, 2010; SILVA, 2013), drainage classes (MCKAY et al., 2010), A horizon silt and sand contents (QI et al., 2006), soil transmissivity (ZHU et al., 1997) or aquifer recharge potential, which is a spatially distributed phenomenon and closely related to soil-landscape models (MENEZES, 2011).

2.4 SOLIM (SOIL-LAND INFERENCE MODEL) AND ARCSIE (SOIL INFERENCE ENGINE)

In order to overcome some limitations of a traditional soil survey, researches and tools have applied knowledge-based techniques and fuzzy logic concepts as a predictive approach, for instance, the softwares SoLIM (ZHU et al., 2003) and ArcSIE (SHI, 2013). They have two major components: a similarity model for representing soil spatial variation and a set of inference techniques for populating the similarity model. The improvements of the last versions also contain means of extracting rules (expert knowledge extraction).

Hereafter is provided a review about the potential of some tools to predict soil types and properties.

ArcSIE works as an extension of ArcMAP (ArcGIS - Environmental Systems Resource Institute). There are two inference methods implemented in ArcSIE for calculating fuzzy membership values: rule-base reasoning (RBR) and case-based reasoning (CBR). In other words, rule and case are two types of knowledge supported by ArcSIE. In RBR, rules are created from direct specifications of soil surveyor, while in CBR it represents the knowledge of the soil at a specific location, also called tacit points.

2.4.1 Ruled-based reasoning with ArcSIE (RBR)

Rule-based reasoning (RBR) in ArcSIE can be useful when the soil scientist knows the soil-landscape relationships and prescribes, under certain environmental conditions, where a specific soil type is more likely to occur. The premise of this technique is that one or two factors out of the five state factors (parent material, climate, organisms, time and topography, JENNY, 1941) control the distribution of soils on the landscape. For example, when climate, organisms, parent material, and time are relatively constant, the topography would be the greatest driver for soil differentiation. Continuous variation of soils are represented by continuous soil property maps derived from the similarity vectors (ZHU et al., 1997) and a lower number of sample points is required (only one typical value per soil type). The following steps are required in order to predict soil types or properties (adapted from LIBOHOVA, 2010) (Figure 2).

1) Establishing soil-landscape relationships

In order to establish the soil-landscape relationships, Zhu and Band (1994) used the knowledge drawn by a certified soil scientist in his domain expert, since he was working in the study area. Libohova (2010) used previous soil surveys and block diagrams from the county soil survey to provide visual insight into the soil-landscape model established by the field soil scientist. In Brazil, where soil series have not been established so far, Menezes (2011) used information from previous soil survey reports and scientific papers which detailed the topographic sequence of soils. For a better comprehension of spatial distribution of soils, it is required the integration of pedologic studies with other branches of science, specially Geology (stratigraphy), Geomorphology and Hydrology (VIDAL-TORRADO et al., 2005). The analysis of the phenomena studied by these disciplines and their results can help in pedologic investigations, collaborating for a better soil sampling and interpretations.

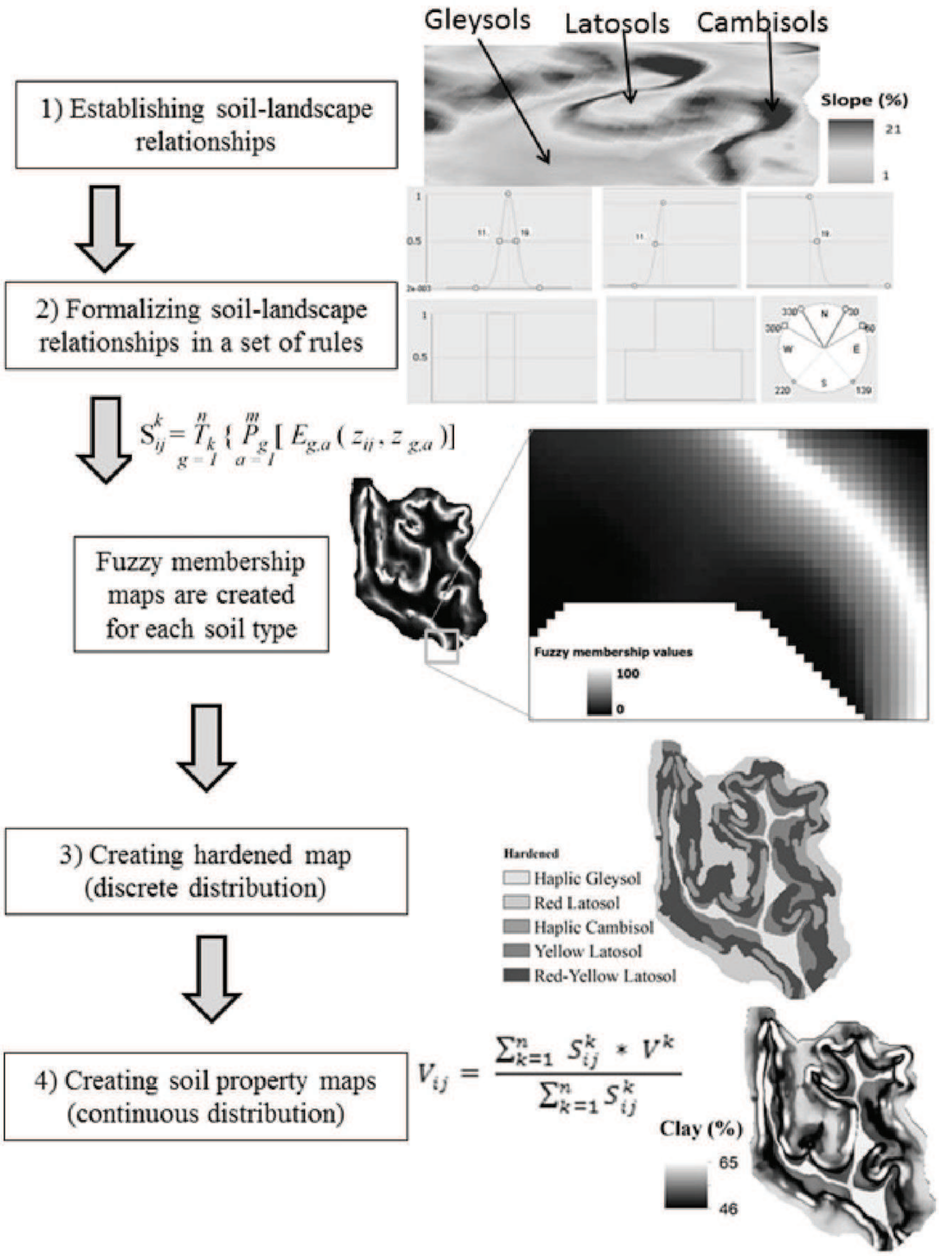

Figure 2 - Steps required for rule-based reasoning. Maps extracted from Menezes (2011).

2) Quantifying relationships between soils and terrain attributes and formalizing these relationships in a set of rules that relates to raster maps

ArcSIE provides tools for soil scientists to formalize the relationships based on pedological knowledge of the local soils. In this case, the inference is based on rules using fuzzy logic. Threshold values are identified and assigned to each soil map unit in a GIS basis. For this, data layers in a raster format that characterize environmental covariates, as terrain attributes, geology, vegetation, climate and others are prepared (SHI et al., 2009). Then, the knowledge about soil-landscape relationships in the first step is qualitatively modeled on a continuous basis in a set of created rules.

The values of the environmental covariates and ranges associated with each soil map class (rules) are used to define membership functions, which, in turn, are referred to as optimality functions as they define the relationships between the values of an environmental feature and a soil type. The rules are set within the software based on "if-then" statements, in which a central location encompasses the rules that provide 100% probability of meeting the class. As the covariates get further from meeting all the rules, the probability of the location being in that class changes and alters the soil property prediction. The number of rules is not limited and information, such as land-use derived from remotely sensed data, can be inserted as a rule and the predictions altered based on the land use type. The cutoffs are set based on knowledge from a soil scientist who understands the soil-landscape relationships (MENEZES, 2011).

The initial output from the inference is a series of fuzzy membership maps in raster format, one for each soil type under consideration (Figure 2). The fuzzy membership values represent the similarities of each place in the landscape to those soil types. The equation below describes how the knowledge of a given soil type is used for a global knowledge in RBR and CBR in order to create fuzzy membership values, represented by three functions (E, P, and T) (SHI et al., 2004):

$$S_{ij}^{k} = \underset{g=1}{\overset{n}{T_k}} \{ \underset{a=1}{\overset{m}{P_g}} [E_{g,a} (z_{ij}, z_{g,a})]$$

where S_{ij}^{k} is the fuzzy membership value at a location (i, j) for a soil k. The m is the number of environmental features used in the inference. The n is the number of instances for soil type k. $Z_{ij,a}$ is the value of the a^{th} environmental feature at location (i,j). $Z_{g,a}$ is the most optimal range given by rule or case g, defining the most favorable condition of feature a for soil k. In RBR it is directly specified by the soil scientist, while in CBR, it is derived by the computer based on the case location and the environmental data layers. E is the function for evaluating the optimality value at the environmental features level. If $Z_{ij,a}$ falls into the range of $Z_{g,a}$, E returns the maximum optimality value; otherwise, E uses a function to derive the optimality value based on the difference between $Z_{ij,a}$ and $Z_{g,a}$. Based on the nature of the environmental covariates used in the prediction, there are five choices for E: *cyclic*, *ordinal*, *nominal*, *raw values*, and *continuous* (bell-shape, z-shape and s-shape continuous curves). P integrates the optimality values from individual environmental covariates to generate an overall predicted value for soil k. T is the function for deriving the final fuzzy membership value for soil k at site (i,j) based on all the instances for soil k.

Using this toolbox, the parameters are adjusted to the curves and explicitly express the mental model of the pedologist. Accomplishing this step, fuzzy membership maps are created (Figure 2, step 2). These maps reveal more details at the spatial level than polygon maps. According to Zhu et al. (1996), the general shapes on the membership images follow the landscape better than the ones on the soil maps where inclusion or exclusion from a region is more based on restrictions derived from the scale of the map than on local conditions. The central concept of the soil type responds to local variations in the apparent soil forming environment (represented by covariables).

3) Creating hardened map

The fuzzy membership maps (Figure 2, step 3) for each soil type are aggregated in order to create a hardened or a defuzzified map, which corresponds to the traditional soil vector map (discrete distribution). For that, ArcSIE assigns at each pixel the soil type with the highest fuzzy membership value.

4) Creating soil property maps

The soil-landscape relationships are extracted and the characterized environmental conditions are linked through a set of inference techniques to populate the similarity model for a given area (ZHU; MCKAY, 2001). Thus, based on fuzzy membership values, the continuous variation of soil properties can be derived from the similarity vectors, using the following formula (ZHU et al., 1997):

$$V_{ij} = \frac{\sum_{k=1}^{n} S_{ij}^{k} \ * \ V^{k}}{\sum_{k=1}^{n} S_{ij}^{k}}$$

where V_{ij} is the estimated potential of recharge value at location *(i,j)*, V^k is a typical value of soil type k (e.g. Haplic Cambisol under native forest), and *n* is the total number of prescribed soil types for the area. If the local soil formative environment characterized by a GIS resembles the environment of a given soil category, then property values of the local soil should resemble the property values of the candidate soil type. The resemblance between the environment for local soil at (*i,j*) and the environment for soil type k is expressed by S_{ij}^{k} , which is used as an index to measure the level of resemblance between the soil property values of the local soil and those of soil category (ZHU et al., 2001). The property value S_{ij}^{k} can be any property that shows a recognizable pattern or relationship with the terrain attribute or landscape position (LIBOHOVA, 2010). The higher the membership of a local soil in a given soil type, the closer the property values (potential of recharge) at that location will be to the typical property values (ZHU et al., 2010).

2.4.2 Case-based reasoning (CBR) with ArcSIE

CBR, in general, is a method of solving problems based on similar problems solved in the past. Dutta and Bonissone (1993) defines better this type of methodology as the action of solving new problems by identifying and adapting similar problems stored in a library of past experiences.

CBR has been applied to soil science in association with fuzzy logic in order to solve problems related to soil data extrapolation. CBR emerges as an alternative to the RBR, since the formulation of rules to explain soils variability becomes laborious, even possessing the knowledge, motivating a search for alternative solutions, being one of them also provided by ArcSIE.

For instance, from a set of points (ArcSIE also works with lines, polygons and rasters as sources of information) with x, y coordinates distributed within a study area and a set of environmental covariates layers (GIS data layers), ArcSIE can extract information from each environmental covariate layer at the site where each point is located, and then associate the points classified as the same soil type with their environmental covariate values of occurrence. For example, considering two soil types (A and B), each one containing 8 and 10 sample points, respectively, and two environmental covariate layers (elevation and slope). The information obtained would be 8 slope and elevation values for soil A and 10 ones for soil B. Thus, one could predict soil properties in no sampled places according to the relationships between environmental data and soil properties. In this example of CBR use, the "former problems" would be the sampled locations, and from them, other places ("new problems") would be classified based on membership approaches characteristic of fuzzy logic.

It has been noticed that a minimum sample size covering the different combinations among environmental covariates has to be reached to allow the data extrapolation. If not, places with environmental combinations not included in the set of points would not be classified, as in the example of Figure 3. In this case, the watershed is located at a mountainous region with dense rain forest vegetation, which hampers the full access to visit and sample soil. Thus, the same property map could be successfully generated with the use of RBR, since this watershed has been intensely studied. Thus, the knowledge could make up the low density of samples.

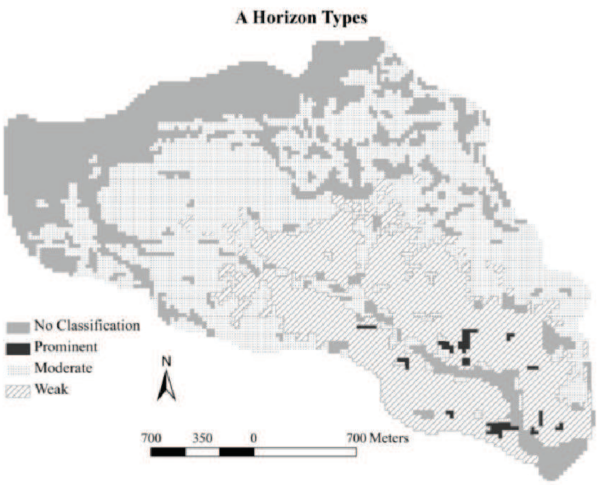

Figure 3 - An example of non-classified places due to the absence of data covering all the environmental features of the interest area.

2.5 CONDITIONED LATIN HYPERCUBE SAMPLING SCHEME

The necessity of finding out the optimal sampling method in order to adequately represent the soil variability within an area has generated many suggestions by soil scientists for years. Over the past decades, extensive work has been published on sampling schemes for soil mapping (MULDER et al., 2013). Additionally, especially in developing countries, the number of samples for a soil survey is limited not only by access difficulties, but also by time and funding restrictions, which hampers the sampling representativeness of the area and influence the final soil map quality. Also, this scenery would not allow the use of CBR for not covering all of the ranges of the environmental covariates.

In this context, Minasny and McBratney (2006) proposed the conditioned Latin Hypercube Sampling (cLHS), derived from Latin Hypercube Sampling (LHS) (McKay et al., 1979), and it has been used in soil science and environmental studies for assessing the uncertainty in a prediction model (MINASNY; MCBRATNEY, 2002). LHS is a stratified random procedure that provides an efficient way of sampling variables from their multivariate distributions (MINASNY; MCBRATNEY, 2006). It follows the idea of a Latin square where there is only one sample in each row and each column, generalizing this concept to an arbitrary number of dimensions. Also, the number of samples desired is taken into account at the time of determining the sampling locations. According to Mulder et al. (2013), if n is the desired sample size, LHS stratifies the marginal distributions of the covariates into n equally probably intervals and randomly samples the multivariate strata such that all marginal strata are included in the sample. However, it may face the issue that sometimes the sampling local may not exist in the field.

In this context, the conditioned Latin Hypercube Sampling (cLHS) adds the condition that the sample chosen must actually occur on the landscape

(BRUNGARD; BOETTINGER, 2010). Minasny and McBratney (2006) showed that cLHS closely represented the original distribution of the environmental covariates with relatively small sample sizes in a digital soil mapping project in the Hunter Valley of New South Wales, Australia.

Small sample sizes able to represent the soils variability is interesting especially for soil scientists from developing countries, where investments and time availability, area accessibility and former soil information are scarce. However, Mulder et al. (2013) highlight that, while LHS is probability sampling, conditioning the LHS on any constraints and sampling costs leads to a purposive sampling strategy since the inclusion probabilities of locations are modified by the conditioning criteria.

The cLHS may distribute the samples throughout the study area, but, sometimes, some places are very difficult or even impossible to be visited for sampling. To avoid this situation, Roudier et al. (2012) proposed a method for incorporating operational constraints into cLHS. They created a "cost" map representing the cost of reaching every place on the landscape considering terrain and landcover attributes. The mentioned work showed that a cost-constrained LHS is not as optimized as the one without cost-conditionings, but the cost of the produced sampling scheme was reduced, thus providing an alternative to implement it.

Silva (2013) used the cLHS constrained by a cost map (created according to the distance from roads, slope and vegetation cover) to indicate the sampling places for validating a rule-based Cambisol solum depth map created through fuzzy logic (RBR) and terrain derivative maps in a watershed of Minas Gerais State, Brazil. The work presented an illustration of the sampling locals disposal with and without cost-constraining the sampling scheme (Figure 4). Also, he affirmed that the cLHS indicated sampling places with different soil properties, such as solum depth, soil moisture and color, and amount of pebbles and gravels, providing an adequate idea of the soil properties distribution along with the landscape features within the study area and, mainly, this sampling scheme reduced the time and investments needed for the field work.

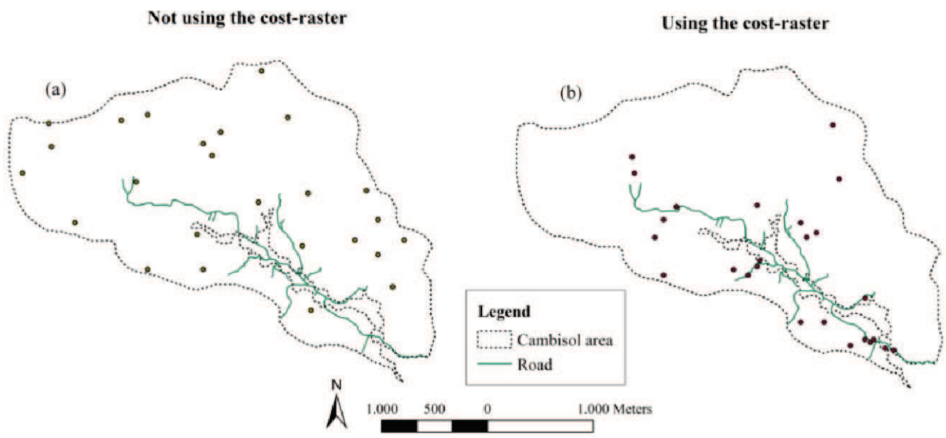

Figure 4 - Conditioned Latin Hypercube Sampling scheme without considering the cost-constrained raster (a) and considering the cost-constrained raster (b) for locating the sampling places in a Cambisol area.

2.6 ACQUIRING INFORMATION FROM EXISTING SOIL MAPS FOR SOIL DATA TRANSFERABILITY

Nowadays, there is a plenty of covariates or layers that can be used to predict soil types and properties, derived from remote sensing, digital elevation models from topographic surveys, geomorphometric variables, analogical or digital soil maps, and others. McKay et al. (2010) investigated potential data layers involved using visual assessment and comparison to known soil locations by expert scientists. Even if the soil-landscape relationships are well known, it could be a hard task to find out which covariate would be more appropriate to tell soil types apart for predictions. From an existing soil map, SoLIM and ArcSIE provide tools for a soil scientist to discover the knowledge implicitly represented by an existing soil map and revise the discovered knowledge. So, it would be possible to transfer the extracted knowledge to other areas with similar soil-landscape relationships.

Transferability of soil types or rules for predicting properties from one small area to a larger extent can be done if the digital soil mapper knows that the initial area is representative of the larger extent (MCBRATNEY et al., 1993). LAGACHERIE et al. (2001) applied this concept for extrapolating French Mediterranean soilscapes (combination of soil-forming factors in a buffer neighbor can be expressed as a vector composition of elementary landscape classes of different sizes). McKay et al. (2010) applied an accurate transferability of knowledge-based model to predict soil series and drainage classes between similar soil-landscape relationship areas. Such concept along with knowledge mining, fits with the scenery of soil surveys in Brazil, where detailed and semidetailed types are available in small areas to support local specific agricultural and environmental projects (MENDONÇA-SANTOS; SANTOS, 2007), but the necessity of more detailed soil maps in larger extensions still remains. Hereafter two ways of extracting knowledge are presented.

2.6.1 SoLIM Knowledge Miner

According to Bui (2004), soil maps represent the structured mental soil-landscape model. One way to exploit such information is provided by SoLIM software (ZHU; BAND, 1994; ZHU et al. 1996; ZHU, 1997; ZHU et al., 1997). The knowledge acquisition tool allows the users to extract pixels information from the terrain derivative maps for each polygon (mapping unit). In this context, occurrence rules for each soil type could be formulated by a soil expert in association with SoLIM knowledge acquisition tool and then transferred to a similar area to identify the places more likely to find similar soil types.

One potential application of that is in areas with limited or no soil data availability, but with some soils similarity, especially in terms of environmental factors that influence the soil formation, to another area with already existing soil maps. They could be used as a source of data for predicting soil information (MCBRATNEY et al., 2003). From an existing map, which contains the surveyor knowledge about the distribution of soils on the landscape, and employing GIS data, models could be adjusted through the analysis of terrain derivative maps, such as slope, wetness index, aspect, profile curvature and so forth, which are supposed to explain the different soil types occurrence in an area based on the catena concept (MILNE, 1935): soil profiles occurring on topographically associated landscapes will be repeated on similar landscapes. This should permit soil data transferability as a manner of assuming soil patterns in the no-data area, based on soil scientist knowledge and soil-landscape models. Zhu et al. (2001) state that the soil-landscape concept contends that if one knows the relationships between each soil and its environment within an area, then one is able to infer what soil might be at each location on the landscape by assessing the environmental conditions at that point.

For instance, it is well-known that the Gleysols are more likely to occur in low elevation and concave places, with high water accumulation (RESENDE et al., 2007), but it should be difficult to tell the values of wetness indexes or concavity in order to separate those places from the surrounding areas. Likewise, Cambisols are more likely to be found on steep relief, but how steep the topography should be in an area of interest to determine the places representative of Cambisols could be hard to tell. Thus, a tool proposed by Zhu et al. (1997)

that extracts the values of those terrain derivatives could help to understand soil types occurrence pattern and, hence, to extrapolate soil types distribution from a mapped area to a similar one that does not have soil data.

SoLIM software contains a knowledge acquisition tool which allows the users to extract pixels information from the terrain derivative maps for each polygon. Regarding a soil map, polygons should represent different mapping units. Thus, through the use of terrain derivative maps, SoLIM provides a way to acquire soil information from environment characteristics, helping to comprehend how the soil data were extrapolated to non-sampled places. This tool also generates graphics from the values of each terrain derivative map for each mapping unit. This would inform the user whether the mapping units are overlapping or not for each terrain derivative map. This latter result would allow the user to classify an area with no soil data based on environmental similarities of different areas through correlations between soil types and terrain attributes.

As an example, a watershed located in Nazareno county, in Minas Gerais State, Brazil, contains Latosols (Oxisols) in association with Cambisols (Inceptisols) on high lands, and Gleysols in low elevation areas (MOTTA et al., 2001). Using the SoLIM tool, it was possible to extract the pixel values of altitude above the channel network (AACN) map over the soil units, as shown on Figure 5. Both curves are not presenting large overlapping areas: Gleysols, as expected, present lower AACN values, basically inferior to 10, the contrary of Red Oxisols. This graphic setting those curves apart contributes to a better understanding of the soil types correlation to AACN. In this context, occurrence rules for each soil type could be formulated by a soil expert in association with SoLIM knowledge acquisition tool and then transferred to a similar area to identify the places more likely to find similar soil types.

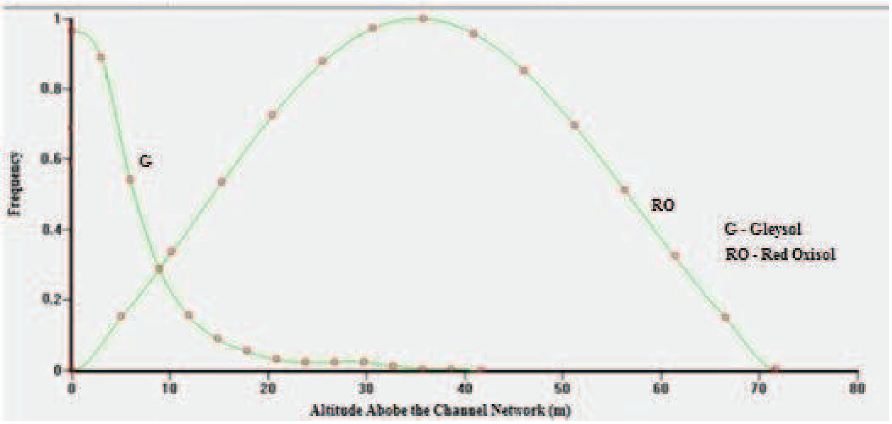

Figure 5 - Graphic showing the pixel frequency distribution (from 0 to 1) for Gleysols and Red Oxisols over altitude above the channel values.

2.6.2 Boxplots

Boxplots are another way to visualize the differences between pixel values of terrain derivative maps for different soil types and also to verify how adequate the extraction of information from existing maps was. They may show the overlapping values and present the differences or similarities of quartiles and medians according to different terrain derivatives and, thus, it makes it possible to identify the best environmental covariate for predicting soil properties.

In order to illustrate this identification tool, Figure 6 shows boxplots of four different mapping units (1, 2, 3 and 4) and four terrain derivatives (slope, profile curvature, wetness index and AACN) of a watershed in Minas Gerais State, Brazil. They were created using the R software (R DEVELOPMENT CORE TEAM, 2013).

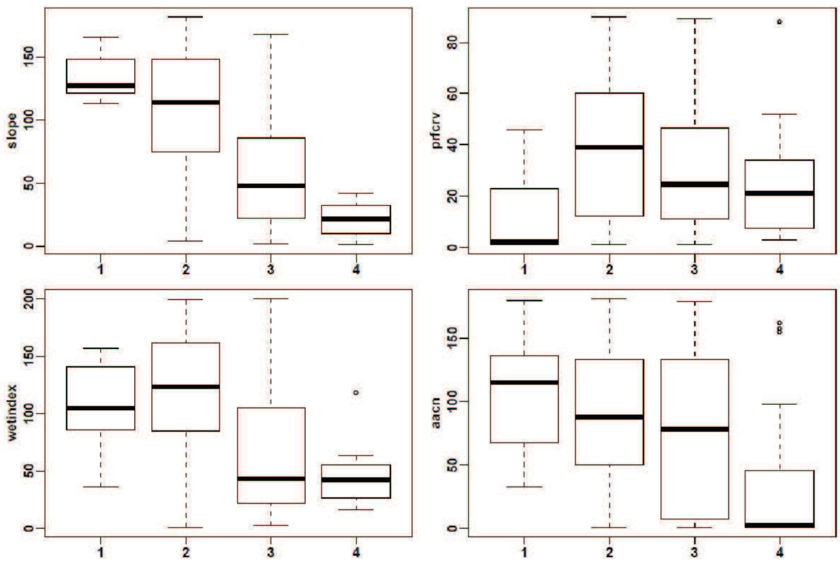

Figure 6 - Boxplots for terrain derivatives and mapping units. prfcrv - profile curvature, wetindex - wetness index, aacn - altitude above the channel network.

Analyzing the boxplots, some overlapping of ranges in values can be seen for slope data although the medians are well separated. Wetness index boxplots for mapping units 1 and 2 are entirely overlapping in values, as well as for 3 and 4 ones, indicating that this terrain attribute would not succeed in separating all the mapping units occurrence. The least overlapping of values is pursued for better understanding the mapping methodology to represent the soils distribution on the landscape.

2.7 COMMONLY USED ACCURACY EVALUATION METHODS

In general, the use of Digital Soil Mapping techniques to spatialize information requires validation of this procedure through field works, that includes soil sampling and/or prospections, to certify that maps are portraying the reality and to measure their accuracy. Many validation methods have been adopted by Digital Soil Mapping community, although some works have not used any validation procedure (Grunwald, 2009).

In spatialization of information in soil mapping, estimate data are assigned to places that have not been sampled. However, those estimate data should be verified by acquiring information in some of those places. In this sense, a simple way of verifying the similarity between the predicted and real data in a specific place is possible through a 1:1 ratio graphic. In this graphic, the closer the set of data is from the main diagonal, the more similar are the real data in relation to the estimate data. From this set of data it is also possible to calculate factors of measurement, such as coefficient of determination (R^2), in which values closer to 1 represent higher similarity between observed and estimate data. As an example, Silva et al. (2014) used 1:1 ratio graphics to verify the accuracy of real soil moisture data measured in the field and estimate soil moisture obtained from Digital Soil Mapping techniques. An example of 1:1 graphic ratio is shown on Figure 7 comparing clay percentage values obtained from estimates by data spatialization (estimate clay percentage) and from laboratory analysis of soil samples (real clay percentage).

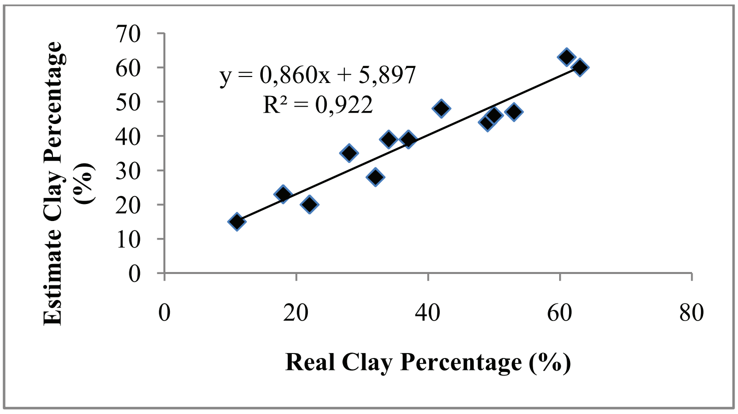

Figure 7 - Example of 1:1 ratio graphic to compare real with estimate clay percentage in soils.

Another commonly employed comparison factor is the root mean square error (RMSE). It measures the difference between values predicted by a model and values observed in the field. Values closer to zero indicate less differences of the real in comparison with the estimate set of data. RMSE can be calculated through the following equation:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(ei - mi)^2}$$

where: *n* is the number of observations, *ei* is the estimated value and *mi* is the measured value. Santos et al. (2013) used RMSE and other factors to evaluate the performance of pedotransfer functions to estimate soil water content at -33 and -1500 KPa for different soil classes of Rio Grande do Sul State, Brazil. Omran (2012) used RMSE as one of the comparison measurements to determine the best interpolation method for predicting diverse soil properties in Egypt.

In addition to the formerly presented accuracy evaluation methods, Kappa index (or Kappa coefficient) is a statistical measure for categorical (qualitative) data, such as soil classes, proposed by Cohen (1960). It is employed when two raters or methods are employed to classify items into categories. Then, it represents the degree of agreement between the two classifiers. Kappa index is calculated according to the following equation:

$$\mathrm{k} = \frac{\Pr(a) - Pr(\mathrm{e})}{1 - \Pr(e)}$$

where Pr(a) is the relative observed agreement between classifiers and Pr(e) is the hypothetical probability of chance agreement.

This index is represented by values ranging from 0 to 1, this latter representing higher accuracy. According to Landis and Koch (1977), Kappa index may be interpreted in five levels: 0-0.2 (slight agreement), 0.2-0.4 (fair agreement), 0.4-0.6 (moderate agreement), 0.6-0.8 (substantial agreement), 0.8-1.0 (almost perfect agreement). As examples of using Kappa index on digital soil mapping, Abdel-Kader (2013) employed Kappa index to compare a reproduction of an original soil map to one created according to regression methods, while Giasson et al. (2011) used kappa index to evaluate the correspondence between original and predicted soil maps created through decision tree methods in Rio Grande do Sul State, Brazil.

2.8 FINAL CONSIDERATIONS

The tools presented in this review have a potential for faster production of soil surveys, since the techniques reduce the inconsistency and costs associated with the traditional manual processes (ZHU et al., 2001). Also, when compared with pedometric approaches, a low density of soil samples is necessary. On the other hand, knowledge-based technique is not automatic, and just as the traditional soil survey, the knowledge of soil-landscape relationships is necessary, and its use has been considered both efficient and economical (HUDSON, 1992; MCMILLAN et al., 2007).

As raised by Hudson (1992), the soil survey has so far failed in not expliciting the mental model of the soil surveyor. Expliciting the rules in functions to get optimality values, as well as the use of knowledge miner techniques in order to utilize the legacy data, can contour this limitation of traditional soil survey. Once the knowledge is explicit, extracted or established in reference areas, the transferability to larger areas within same soil-landscape relationships should be tested (MCKAY et al., 2010), as an opportunity to raise the geographic expression of surveyed areas, very much needed in Brazil.

Since fuzzy membership maps represent soil types and can be viewed as a non-linear transformation of environmental variables based on expert knowledge of a soil-landscape model (ZHU et al., 2010), its use as an auxiliary in soil property prediction should be more explored. One example of such application is related to pedometric prediction methods. Those that do not incorporate the use of auxiliary variables (interpolation relying only on point observations) have been outperformed by hybrid methods (interpolation relying on point observations combined with interpolation based on regression of the target variable on spatially exhaustive auxiliary information). Hybrid methods explore the extra information when there is auxiliary information (maps of covariates related to terrain, land use, and others) able to explain part of variation (HENGL et al., 2007). In this sense, Zhu and Lin (2010) compared maps generated from linear regression and environmental variables with regression using fuzzy membership maps as auxiliary. The non-linearity and complexity inherent to the steeper terrain with more variable soil types were well captured by a set of soil membership maps, which can be used to describe model and non-linear variation of soil property values. The linear regression using environmental variables would be more appropriate to be used on gently rolling landscapes, where soil-environment model is simple and stable over space.

Finally, the mapping tools presented in this work show the advantages of associating them to the field expert-knowledge in order to enhance the final results quality. Along with that, however, it is worthy to remind that these tools should be used on soil surveys and mapping to assist the field work (and never in order to replace it), mainly because the soil variability is not completely predictable, which makes this field activity irreplaceable for soil mapping.

REFERENCES

ABDEL-KADER, F.H. Digital soil mapping using spectral and terrain parameters and statistical modelling integrated into GIS-Northwestern Coastal Region of Egypt. In: SHAHID, S.A.; TAHA, F.K.; ABDELFATTAH, M.A. (eds.), **Developments in Soil Classification, Land Use Planning and Policy Implications.** New York: Springer, 2013, p.353-372.

BRUNGARD, C.W.; BOETTINGER, J.L. Conditioned Latin Hypercube Sampling: Optimal Sample Size for Digital Soil Mapping of Arid Rangelands in Utah, USA. In BOETTINGER, J.L.; HOWELL, D.W.; MOORE, A.C.; HARTEMINK, A.E.; KIENAST-BROWN, S. (eds.), **Digital Soil Mapping**: Bridging Research, Environmental Application, and Operation (Progress in Soil Science 2). New York: Springer Science+Business Media B.V., 2010, v.2, p. 67-75.

BUI, E.N. **Soil survey as a knowledge system.** Geoderma, Amsterdam, v.120, n.1–2, p. 17–26, 2004.

COHEN, J. A coefficient of agreement for nominal scales. **Educational and Psychological Measurement**, v.20, n.1, p.37–46, 1960.
DALE, M.B.; MCBRATNEY, A.B.; RUSSELL, J.S. On the role of expert systems and numerical taxonomy in soil classification. **Journal of Soil Science**, HOBOKEN, v.40, n.2, p.223–234, 1989.

DUTTA, S.; BONISSONE, P.P. Integrating Case- and Rule-based Reasoning. **International Journal of Approximate Reasoning**, New York, v.8, p.163-203, 1993.

GIASSON, E.; SARMENTO, E.C.; WEBER, E.; FLORES, C.A.; HASENACK, H. Decision trees for digital soil mapping on subtropical basaltic steeplands. **Scientia Agricola**, Piracicaba, v.68, n.2, p.167-174, 2011.

GRUNWALD, S. Multi-criteria characterization of recent digital soil mapping approaches. **Geoderma**, Amsterdam, v.152, p.195-207, 2009.

HENGL, T.; HEUVELINK, G.; ROSSITER, D.G. About regression-kriging: from equations to case studies. **Computer and Geosciences**, London, v.33, n.10, p.1301-1315, Oct. 2007.

HUDSON, B. D. The soil survey as a paradigm-based science. **Soil Science Society of America Journal**, Madison, v.56, p.836–841, 1992.

JENNY, H. **Factors of soil formation**. New York: McGraw-Hill, 1941. 281p.

LAGACHERIE, P.; ROBBEZ-MASSON, N.; NGUYEN-THE, N.; BARTHÈS, J.P. Mapping of reference area representativity using a mathematical soilscape distance. **Geoderma**, Amsterdam, v.101, p.105-118, 2001.

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

LEGROS, J.P. **Mapping of the Soil**. Enfield: Science Publishers, 2006. 411p.

LIBOHOVA, Z. **Terrain attribute soil mapping for predictive continuous soil property maps**. Ph.D. Thesis. West Lafayette: Purdue University, 2010. 122 p.

MCBRATNEY, A.B.; MINASNY, B.; MACMILLAN, R.A.; CARRÉ, F. Digital soil mapping. In: HUANG, P.M.; YUNCONG, L.; SUMNER, M.E. (eds). **Handbook of soil sciences**: properties and processes. 2^{nd}ed. Boca Raton: CRC Press, 1993, 1442p.

MCBRATNEY, A.B.; SANTOS, M.L.M.; MINASNY, B. On digital soil mapping. **Geoderma**, Amsterdam, v.117, n.4, p.3-52, Jun. 2003.

MCKAY, M.D.; BECKMAN, R.J.; CONOVER, W.J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. **Technometrics,** Alexandria, v.21, p.239–245, 1979.

MCKAY, J.; GRUNWALD, S.; SHI, X; LONG, R.F. Evaluation of the transferability of a knowledge-based soil-landscape model. In: BOETTINGER, J.L.; HOWELL, D.W.; MOORE, A.C.; HARTEMINK, A.E.; KIENAST-BROWN, S. (eds.). **Digital soil mapping:** bridging research, environmental application, and operation. London: Springer, 2010, p. 165-177.

MACMILLAN, R.A.; MOON, D.E.; COUPE, R.A. Automated predictive ecological mapping in a Forest Region of B.C., Canada, 2001–2005. **Geoderma**, Amsterdam, v.140, p.353–373. 2007.

MACMILLAN, R.A.; PETTAPIECE, W.W.; BRIERLEY, J.A. An expert system for allocating soils to landforms through the application of soil survey tacit knowledge. **Canadian Journal of Soil Science**, Ottawa, v.85, p.103-112, 2005.

MENDONÇA-SANTOS, M.L.; SANTOS, H. The state of the art of Brazilian soil mapping and prospects for digital soil mapping., In: LAGACHERIE, P.; MCBRATNEY, A.B.; VOLTZ, M. (eds). **Developments in Soil Science**. New York: Elsevier, 2007. v.31, p.39-54

MENEZES, M. D. **Levantamento pedológico de hortos florestais e mapeamento digital de atributos físicos do solo para estudos hidrológicos**. – . Ph.D. Thesis. Lavras: UFLA, 2011. 225p.

MILNE, G. Some suggested units of classification and mapping particularly for East African soils. **Soil Research**, Vitoria, v.4, p.183-198. 1935.

MINASNY, B; MCBRATNEY, A.B, Uncertainty analysis for pedotransfer functions. **European Journal of Soil Science**, Hoboken, v.53, 417–430, 2002.

MINASNY, B.; MCBRATNEY, A.B. A conditioned Latin hypercube method for sampling in the presence of ancillary information. **Computers and Geosciences**, Oxford, v.32, n.9, p.1378–1388, 2006.

MOTTA, P.E.F.; CURI, N.; SILVA, M.L.N.; MARQUES, J.J.G.S.M.; PRADO, N.J.S.; FONSECA, E.M.B. **Levantamento pedológico detalhado, erosão dos solos, uso atual e aptidão agrícola das terras de microbacia piloto na região sob influência do reservatório de Itutinga/Camargos, MG**. Belo Horizonte: CEMIG, 2001. 51 p.

MULDER, V.L.; BRUIN, S.; SCHAEOMAN, M.E. Representing major soil variability at regional scale by contrained Latin Hypercube Sampling of remote sensing data. International **Journal of Applied Earth Observation and Geoinformation**, Amsterdam, v.21, p.301-310. 2013.

OMRAN, E.E. Improving the prediction accuracy of soil mapping through geoestatistics. **International Journal of Geosciences**, v.3, p.574-590, 2012.

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

QUINN, T.; ZHU, A.X., BURT, J.E. Effects of detailed soil spatial information on watershed modeling across different model scales. **International Journal of Applied Earth Observation and Geoinformation,** Amsterdam, v.7, p.324–338, 2005.

R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing.** Vienna: R Foundation for Statistical Computing. Available at: <http://www.r-project.org> Accesses in: 24 Jun. 2013.

RESENDE, M.; CURI, N.; REZENDE, S.B.; CORRÊA, G.F. **Pedologia**: base para distinção de ambientes. 5.ed. Lavras: UFLA, 2007. 322 p.

ROUDIER P.; BEAUDETTE, D.E.; HEWITT, A.E. A conditioned Latin hypercube sampling algorithm incorporating operational constraints. Proceedings: 5th Global Workshop on Digital Soil Mapping 2012: **Digital Soil Assessments and Beyond**, Sydney, p.10-13, 2012.

SANTOS, W.J.R.; CURI, N.; SILVA, S.H.G.; ARAÚJO, E.F.; MARQUES, J.J. Pedotransfer functions for water retention in different soil classes from the center-shouthern Rio Grande do Sul State. **Ciência e Agrotecnologia**, Lavras, v.37, n.1, p.49-60, 2013.

SCULL, P.; FRANKLIN, J.; CHADWICK, O.A.; MCARTHUR, D. Predictive soil mapping: a review. **Progress in Physical Geography**, London, v.27, n.2, p.171-197. 2003.

SHI, X. **ArcSIE user's guide**. Available in: <http://www.arcsie.com/index.htm> Acessed on: Jul 4, 2013.

SHI, X.; LONG, R.; DEKETT, R.; PHILIPPE, J. Integrating Different Types of Knowledge for Digital Soil Mapping **Soil Science Society of America Journal**, Madison, v.73, n.5, Sep/Oct. 2009.

SHI, X.; ZHU, A.X.; BURT, J.E.; QI, F., SIMONSON, D. A case-based reasoning approach to fuzzy soil mapping. **Soil Science Society of America Journal**, Madison, v.68, p.885–894, 2004.

SILVA, B.M.; SILVA, S.H.G.; OLIVEIRA, G.C.; PETERS, P.H.C.R.; SANTOS, W.J.R.; CURI, N. Soil moisture assessed by digital mapping techniques and its field validation. **Ciência e Agrotecnologia**, Lavras, v.38, n.2, p.10-148, 2014.

SILVA, S.H.G. **Cambisol (Inceptisol) solum thickness mapping based on expert knowledge with limited data from a watershed in Minas Gerais, Brazil**. Monograph. Lavras: UFLA, 2013. 41p.

SKIDMORE, A.K.; RYAN, P.J.; DAWES, W.; SHORT, D.; O'LOUGHLIN, E. Use of an expert system to map forest soil from a geographical information system. **International Journal Geographical Information Science,** Wageningen, v.5, p.431–445. 1991.

VIDAL-TORRADO, P.; LEPSCH, I.F.; CASTRO, S.S. Conceitos e aplicações das relações pedologia-geomorfologia em regiões tropicais úmidas. In: VIDAL-TORRADO, P.; ALLEONI, L.R.F.; COOPER, M.; SILVA, A.P. da; CARDOSO, E.J. (eds.) **Tópicos em Ciência do Solo**. Viçosa: Sociedade Brasileira de Ciência do Solo, 2005, v.4, p.145-192.

WALTER, C.; LAGACHERIE, P.; FOLLAIN, S. Integrating Pedological Knowledge into Digital Soil Mapping. In: LAGACHERIE, P.; MCBRATNEY, A.B.; VOLTS, M. (eds). **Digital Soil Mapping -** an Introductory Perspective Developments in Soil Science, New York: Elsevier, 2007, v.31. p.281-300.

ZADEH, L.A. Fuzzy sets. **Information and Control***,* v.8, p.338–353. 1965. Available in: < http://www-bisc.cs.berkeley.edu/Zadeh-1965.pdf>. Accessed on Jun 29, 2013.

ZHU, A.X. A similarity model for representing soil spatial information. **Geoderma**, Amsterdam, v.77, p.217–242. 1997.

ZHU, A.X., BAND, L.E. A knowledge-based approach to data integration for soil mapping. **Canadian Journal of Remote Sensing**, Kanatan, v.20, p.408–418, 1994.

ZHU, A.X.; BAND, L.E.; DUTTON, B.; NIMLOS, T.J. Automated soil inference under fuzzy logic. **Ecological modeling**, Amsterdam, v.90, n.2, p.123-145, 1996.

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

ZHU, A.X.; BURT, J.E.; MOORE, A.C.; SMITH, M.P.; LIU, J.; FENG, Q. **SoLIM**: A New Technology For Soil Mapping Using GIS, Expert Knowledge and Fuzzy Logic. 2003. Available in: <http://solim.geography.wisc.edu/pubs/Overview2007-02-16.pdf>. Accessed on: Jun 29, 2013.

ZHU, A.X.; HUDSON, B.; BURT, J.; LUBICH, K.; SIMONSON, D. Soil mapping using GIS, expert knowledge, and fuzzy logic. **Soil Science Society of American Journal**, Madison, v.65, n.5, p.1463-1472, Apr./May 2001.

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

ZHU, A.X.; MCKAY, D.S. Effects of spatial detail of soil information on watershed modeling. **Journal of Hydrology**, Amsterdam, v. 248, n. 4, p. 54-77, July 2001.

ZHU, A.X.; QI, F.; MOORE, A.; BURT, J.E. Prediction of soil properties using fuzzy membership values. **Geoderma**, Amsterdam, v.158, n. 3/4, p.199-206, Sept. 2010.

3. ARTIGO 2. A technique for low cost soil mapping and validation using expert knowledge on a watershed in Minas Gerais, Brazil

***Artigo nas normas da Soil Science Society of America Journal.**

ABSTRACT

Understanding the soil attributes and types occurring within a region is critical for providing the best land-use decisions. Soils vary in their ability to clean and store water, provide water for plant growth, and many other ecosystem services. Soil variability is dependent on climate, parent material, organisms, time, and topography. When only topography varies within an area, the topography and redistribution of water should be the main drivers for soils differentiation. Digital soil mapping (DSM) has advantages due to computational tools and easily accessible digital elevation models (DEMs) at multiple resolutions. Terrain attributes (e.g., slope, wetness index, and profile curvature) are derived from the DEM and, in association with a soil expert, knowledge-based models can be applied to predict soil variability. The objective of this study was to create and validate a predicted Cambisol (Inceptisol) solum depth map for Lavrinha Creek Watershed (LCW) in Minas Gerais, Brazil, by applying DSM techniques for the Brazilian soil landscapes. The best available 30-m DEM was used to derive the terrain derivatives. A set of rules were formulated according to the terrain attributes, limited data, and expert knowledge to predict the solum depth behavior throughout the watershed. Conditioned Latin hypercube sampling scheme was used for allocating the validation points. In this study, 20 out of the 25 validating samples were correctly classified yielding a Kappa index of 0.616. Soil expert knowledge and Digital Soil Mapping techniques can be employed for mapping areas, especially

in countries where there is limited data available, which will provide a useful soil map for planning while saving time and investments.

Abbreviations: AACN, altitude above the channel network; cLHS, conditioned Latin hypercube sampling; DEM, digital elevation model; DSM, digital soil mapping; GIS, geographic information systems; LCW, Lavrinha Creek Watershed; LHS, Latin hypercube sampling; SWI, SAGA wetness index.

3.1 INTRODUCTION

Understanding the role soil plays in ecosystem functions occurring over large regions is critical for informing the best use and aid management decisions of natural resources. In Brazil, soils information is most often obtained from soil surveys, which are commonly coarse resolution, outdated, or based on taxonomic divisions. Due to funding limitations, soil surveys for most of Brazil are available only at small scale (1:750,000), and just a small portion of the Brazilian territory has semidetailed or detailed soil surveys (Giasson et al., 2006). Furthermore, there are many financial and time restrictions that typically limit sampling at temporal and spatial scales, which also slow the progress towards fine resolution soil maps (Iwashita et al., 2012). Due to these limitations, the availability of soils information has not kept up with the demand for soil information, and this demand motivates the development of alternative soil mapping methods based on limited data (McBratney et al., 2003).

Introduction of digital technologies, such as remote sensing and DSM techniques, has provided new opportunities to predict soil properties and processes (Grunwald, 2009). Among them, knowledge-based DSM methods have advantages for being a quick and economic alternative (Mendonça-Santos et al., 2008), not only for using local soil scientist knowledge of soil–landscape relationships but also for capturing the pedologists' mental model. Lagacherie and Voltz (2000) highlight that the regional soil pattern knowledge from local soil scientists can be an important tool due to the possibility of predicting the soil types occurring in non-mapped areas, using previous information generated in reference areas. Thus, associating expert knowledge with new soil mapping methods may improve soil mapping final products, especially in countries where information and field activities are limited.

Since the factors of soil formation (climate, organisms, topography, parent material, and time) have been described for soil development (Jenny, 1941), this mental model may be used to develop predictive soil maps. Also, using Milne's catena concept that soils occur on topographically associated landscapes that are repeated on similar landscapes (Milne, 1935), soils should be predictable where only topography is varying across the landscape (Menezes et al., 2013). Gessler et al. (2000) found high correlation among several soil properties and topographic indices (e.g., slope, profile curvature, and topographic wetness index) derived from DEMs. Like no other time in history, pedologists have access to DEMs available at different resolutions, which can aid in relating soils differentiation based on the relief factor, soil expert knowledge, and mapping techniques, such as fuzzy logic.

Fuzzy logic has been increasingly employed in soil science due to its ability to capture and represent the continuous nature of soil spatial variation (Zhu and Band, 1994), and, along with knowledge-based digital soil mapping, it has been extensively used to predict soil types and soil physical-chemical properties (McKay et al., 2010; Zhu and Band, 1994; Zhu et al., 1997, 2010), since these properties are inherent to soils and landscapes (Menezes et al., 2013). In this sense, soil properties in non-sampled places can be predicted according to their membership to places wherein the relation between soil properties and environmental variables, such as topographic factors, is known. The membership values range from 0 to 1 and the higher the value, the higher the similarity between different places. Zhu et al. (1997) successfully used this methodology to predict the depth of A horizon, while Quinn et al. (2005) employed this technique to predict solum depth. However, this knowledge-based mapping method requires a non-biased validation method, and validation of predictive maps is one of the most crucial steps in creating the product.

Validation involves sampling schemes that contemplate soil property variability over the entire area and ancillary data, as terrain attributes, can be used for this purpose. In this context, conditioned Latin hypercube sampling (cLHS) can be used to validate the knowledge-based inference maps. The conditioned Latin hypercube has been increasingly adopted by the DSM community (Minasny and McBratney, 2010). It was proposed by Minasny and McBratney (2006) as a derivation of Latin hypercube sampling (LHS) (McKay et al., 1979). Contrary to LHS, cLHS adds the condition that the chosen sampling place must actually occur on the landscape (Brungard and Boettinger, 2010) and may be conditioned to other requirements such as location to access. Roudier et al. (2012) proposed the association of the cLHS with a cost raster, which represented the cost (difficulty) of reaching every place within a study area, taking into account the terrain attribute maps and other variables, such as distance from roads and landcover. Thus, the cLHS scheme would give preference to sample in the low-cost areas. Although Roudier et al. (2012) found that this cost-constrained cLHS is not as representative of the variability as the standard cLHS because the former under-sampled high cost places, this sampling scheme might also be used for choosing validation locales in predictive maps, especially in areas of difficult access.

In this context, this study was performed to accomplish two objectives: (i) to create low cost, reliable soil maps with limited data based on soil expert knowledge inference predictive soil maps, using DSM tools in Brazil; and (ii) to test the cost-sensitive validation method using conditioned Latin hypercube sampling techniques. In Brazil and much of the world, a low cost alternative is needed to create soil maps at reasonable scales for adequate use and management. This method will combine soil expert knowledge with DSM techniques (relationship between terrain attributes and soil properties) creating a predicted solum thickness map for Cambisols (Inceptisols) in LCW, with limited data availability.

3.2 MATERIALS AND METHODS

3.2.1 Study area

This study was conducted in LCW, which is located in Bocaina de Minas county, in Minas Gerais State, Brazil, within the Serra da Mantiqueira physiographical region (between the longitudes 44°26′21″ and 44°28′39″ W, and latitudes 22°06′53″ and 22°08′28″ S). This study site is a component of the Alto Rio Grande Basin and is one of the headwater streams, which flow into the Grande River, which is very important for the water supply for Camargos Hydroelectric Power Station.

This area is characterized as having semitropical of high altitudes climate, with an average temperature in the hottest month lower than 22°C, dry winters, and rainy summers. The annual average precipitation is 2000 mm, with an annual water deficit ranging from 50 to 100 mm and an annual water surplus greater than 800 mm. Lavrinha Creek Watershed covers an area of 676 ha, the altitude ranges from 1160 to 1729 m, it is included in the Andrelândia Plateau, and the soil parent material is derived from gneiss. The native vegetation is Atlantic forest, and the land uses include native Atlantic forest reserve in the steepest slopes and degraded pasture and *Eucalyptus* plantation in the gentle slopes.

3.2.2 Data used for digital soil mapping

The available site information was obtained from Menezes et al. (2009), which include the LCW soil taxonomic map (Fig. 1), at a scale of 1:20,000 based on samples collected throughout the area and profile descriptions. Cambisols (Inceptisols) on the sloping landscape cover 92% of the watershed, followed by 7% and 1% of Fluvic Neosols (Fluvents) and Haplic Gleysol (Aquents), respectively, on floodplains. A 30-m pixel Aster DEM was obtained from the website www.gdem.aster.ersdac.or.jp (accessed on May, 14th 2012) to create raster maps of terrain attributes derived from the DEM. Slope, profile curvature and SAGA wetness index (SWI) were used in the solum depth map prediction, whereas altitude above the channel network, elevation, slope, and SWI were used for selecting the lowest elevation areas of the landscape, which were not classified as Cambisols (Inceptisols) and are not a focus of this study.

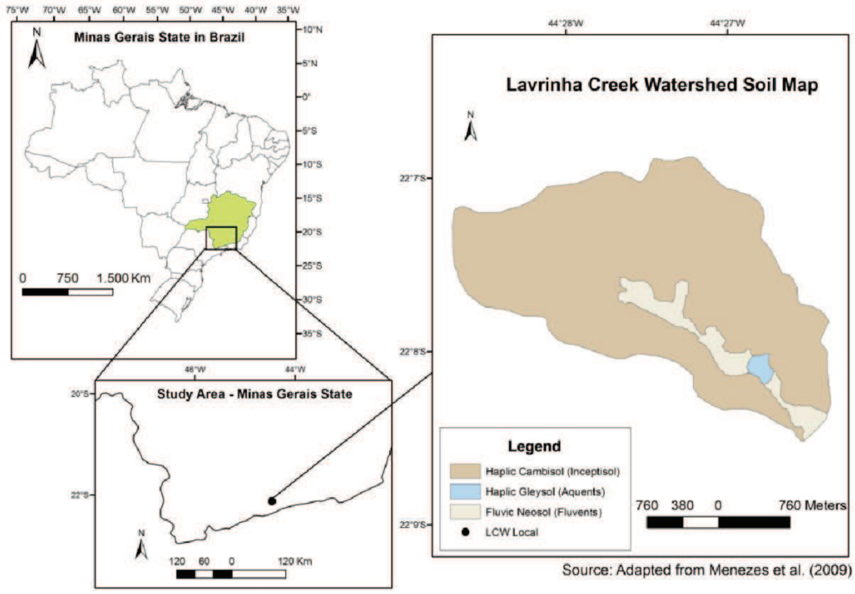

Fig. 1. Lavrinha Creek Watershed soil map and its location within Minas Gerais State, Brazil.

3.2.3 Methods for creating inference maps

Terrain attribute maps of SWI, profile curvature, altitude above the channel network (AACN), and slope were created in SAGA geographic information systems (GIS) (Böhner and Conrad, 2009) and ArcGIS 9.3 (ESRI) software to characterize topographic attributes. Those terrain attributes were selected based on methodologies used by Kuriakose et al. (2009), Boer et al. (1996), and Gessler et al. (2000), and were adapted for the study area. From them (environmental conditions) and their relationships to solum thickness, rules of occurrence for each solum depth class were created, according to soil expert knowledge, for inference mapping. Solum thickness (A+B horizons) was aimed to be predicted instead of soil depth to the bedrock because, in this region, C horizon is very thick, sometimes reaching more than 26 m of depth (Rezende, 1980).

Based on this methodology and on the developed rules, different combinations of the terrain attributes slope, profile curvature, and SWI were related to each expected solum depth class for each pixel on the map as described in the subsequent sections. The depth classes are in agreement with the Brazilian Soil Classification System (Embrapa, 2013): Shallow (≤50 cm), Moderate Deep (>50 and ≤100 cm), Deep (>100 and ≤200 cm), and Very Deep (>200 cm).

The environmental conditions in which each solum depth class was expected (rules) were based on a local soil expert's knowledge, with experience on soil survey, genesis, and classification, and the following conditions: it is expected that places with higher wetness index (>10.5) have more water and sediments accumulation (deep soil) and in places where the wetness index is lower (≤10.5), water movement is faster, and, hence, due to low water infiltration, erosion is more likely to occur (shallow soil); profile curvature values lower than −0.0008 represent convex landscapes and the values between −0.0008 and +0.0008 define linear landscapes. Both intervals indicate regions more likely to lose soil, while the values greater than +0.0008 represent concave places more likely to receive soil eroded from higher altitude and, therefore, to present thicker solum; deeper soils are also expected to be found under gentle slopes whereas shallower soils are expected to be found on steep slopes (Schaetzl and Anderson, 2005).

The rules presented in Table 1, established by a soil expert, were inserted in an ArcGIS 9.3 (ESRI) extension named ArcSIE (Soil Inference Engine) (Shi, 2013). This tool uses expert knowledge and fuzzy logic to show on maps the areas that better fit each rule by presenting high membership values, which corresponds to more similarities to the given rules (Table 1). The greater the similarity of a pixel to a certain rule, according to its environmental conditions, the higher its membership to that rule and, hence, to the solum depth, which that rule represents. Table 2 presents the parameters used in ArcSIE to insert each rule presented in Table 1.

Table 1 – Rules to predict the Cambisol solum depth based on soil expert knowledge.

Rule	Slope (%)	Profile Curvature	Wetness Index	Solum Depth (cm)
1	≤ 20	≥ -0.0008 ≤ +0.0008	≤ 10.5	> 50 and ≤ 100
2	≤ 20	≥ -0.0008 ≤ +0.0008	>10.5	> 100
3	≤ 20	< -0.0008	> 10.5	> 100
4	≤ 20	< -0.0008	≤ 10.5	> 100
5	≤ 20	>0.0008	> 10.5	> 100
6	≤ 20	>0.0008	≤ 10.5	> 100
7	> 20 and ≤ 45	≥ -0.0008 ≤ +0.0008	≤ 10.5	> 50 and ≤ 100
8	> 20 and ≤ 45	≥ -0.0008 ≤ +0.0008	> 10.5	> 50 and ≤ 100
9	> 20 and ≤ 45	< -0.0008	> 10.5	> 100
10	> 20 and ≤ 45	< -0.0008	≤ 10.5	> 50 and ≤ 100
11	> 20 and ≤ 45	>0.0008	> 10.5	> 100
12	> 20 and ≤ 45	>0.0008	≤ 10.5	> 50 and ≤ 100
13	> 45	≥ -0.0008 ≤ +0.0008	≤ 10.5	≤ 50
14	> 45	≥ -0.0008 ≤ +0.0008	> 10.5	> 50 and ≤ 100
15	> 45	< -0.0008	> 10.5	> 50 and ≤ 100
16	> 45	< -0.0008	≤ 10.5	≤ 50
17	> 45	>0.0008	> 10.5	> 50 and ≤ 100
18	> 45	>0.0008	≤ 10.5	≤ 50

Table 2 – Parameters and classes used in ArcSIE to predict the Cambisol solum depth based on soil expert knowledge.

Terrain Attribute	Class of Values	v1 and v2	w1	w2	Curve Shape
Slope (%)	≤ 20	0.20	--	0.02	z-shaped
	> 20 and ≤ 45	0.32	0.12	0.13	bell-shaped
	> 45	0.45	0.02	--	s-shaped
Profile Curvature	< -0.0008	-0.0008	--	0.0004	z-shaped
	≥ +0.0008 and ≤ -0.0008	0	0.0008	0.0008	bell-shaped
	> +0.0008	0.0008	0.0004	--	s-shaped
SWI	≤ 10.5	10.5	--	1	z-shaped
	> 10.5	10.5	1	--	s-shaped

SWI: Saga Wetness Index; v1 and v2: threshold values that determine high membership below or above that value according to the curve shape; w1 and w2: deviations from the v1 and v2 that represent 50% membership.

The expected Cambisol solum depth map was generated using the mentioned ArcGIS extension. In Table 2, “v1” and “v2” are the thresholds, “w1” and “w2” are deviations, and curve shapes indicate the values that should have high membership values. The “s-shaped” curve is used to determine that raster values that are greater than a pre-established value (v1) are assigned high membership; while a “z-shaped” curve determines that raster values lower than the pre-established one, v2 in this case, are assigned high membership; and the bell-shaped curve assigns higher membership to the raster values in between two other pre-established values (w1 and w2).

For instance, the first line on Table 2 indicates an expert knowledge-based rule for identifying places where the slope gradient is lower than 20%. Thus, it uses a z-shaped curve (Fig. 2), being the threshold 0.20 (v2) and a deviation of 0.02 (w2), which means that slope values lower than 20% will be assigned 100% membership, while slope values between 20 and 22% (v2+w2 or 0.20+0.02) are assigned membership decreasing from 100% as pixel values get farther from the threshold (v2 or 20%). On the other hand, a s-shaped curve (Fig. 2) was employed to determine places where slope is greater than 45%, where v1 is 0.45 and the w1 deviation is 0.02. Thus, slope gradients greater than 45% are assigned 100% membership, and the ones between 43 (0.45−0.02) and 45% are assigned decreasing membership according to the rule (slope ≥ 45%). To identify places where slope ranges from 20 to 45%, a bell-shaped curve was used (Fig. 2). v1 was 0.32, w1 was 0.12, and w2 was 0.13. This means that raster values ranging from 20 (0.32 − 0.12) to 45% slope (0.32 + 0.13) are assigned higher membership than those which are not included in this interval. In this example, only slope was discussed, but other terrain attributes such as wetness index and profile curvature were also employed to define the solum depth class variability in the study area.

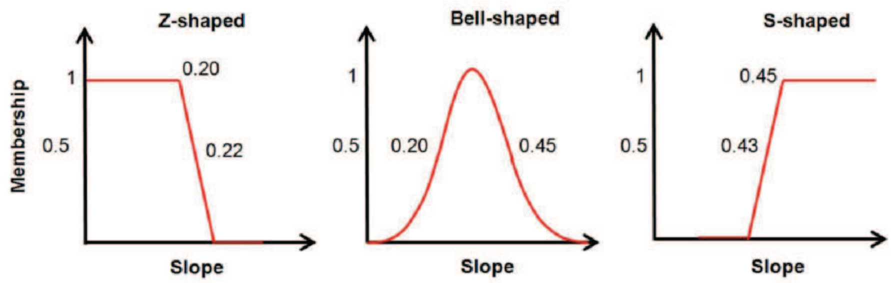

Fig. 2. Curve shapes used to define values of terrain attributes (i.e., slope) to which high membership is assigned according to different rules to predict solum depth in Lavrinha Creek Watershed: lower than 20% slope in z-shaped curve, between 20 and 45% slope in bell-shaped curve, and greater than 45% in s-shaped curve.

It was necessary to delineate the lowest areas in the landscape near the drainage channels so that the rules would not apply to these non-Cambisol areas. The geomorphic and pedogenic processes are different within these areas and must be separated. ArcSIE was used again to select and then exclude from the LCW map the lowest areas on the landscape, where the lowland soils are located, so that only Cambisols would be represented on the map. For this purpose, elevation, slope, AACN, and SWI were used. Table 3 presents the instances, also based on soil expert knowledge, for each of those parameters used in ArcSIE to select the lowest areas on the landscape. The validation of this procedure was made by sampling two random places within this area to certify that area was not occupied by Cambisols.

Table 3 – Parameters applied in ArcSIE for pointing out the lowest areas on the landscape based on soil expert knowledge.

Attributes	Curve Shape	v	w
Elevation	Z-shaped	1220	12
Slope	Z-shaped	0.06	0.02
AACN	Z-shaped	12	3
Wetness Index	S-shaped	15	0.43

AACN: Altitude Above the Channel Network; v: threshold value that determine high membership below or above that value according to the curve shape; w: deviations from the v that represent 50% membership.

3.2.4 Soil sampling for validation

The cLHS, proposed by Minasny and McBratney (2006), was used for choosing the locations for validating the predicted map by verifying the actual solum depth in the field. This method may place each validation sample according to environmental variables (slope, profile curvature, and SWI), considering that their variability should explain the soil properties of interest, in this case, the solum depth. The cLHS could be considered valid when the assumed major varying factor driving pedogenesis within the studied area is the topography.

Keeping this in mind, the cLHS will determine the combinations needed for statistical validity but restricts choosing a place within the watershed that is difficult to access. Therefore, the sample locations were selected in places with similar environmental characteristics located in easy-to-reach areas. Following the steps proposed by Roudier et al. (2012) to create a cost raster, in this study, distance from the road, slope, and vegetation cover were considered the constraining attributes that would possibly hamper the sampling process.

Each of those terrain attribute rasters was divided into cost classes (Table 4), which present the difficulty of reaching a place with certain characteristics. For example, considering the slope raster for the value class greater than 45%, it was given a cost of 9, whereas for the class 0–3%, the assigned cost is 1, which means it is easier to reach places under gentle slopes than steep slopes. This procedure does not mean that slopes greater than 45% will not be sampled, they will only have a cost associated with the sample. Then, the three rasters reclassified according to costs were added to one another, resulting in a final cost raster that represents the difficulty of reaching each place (pixel) in the field.

Table 4 – Values assigned per class for each attribute raster in order to create a cost raster showing the difficulty to reach every part of the study area.

Slope Class (%)	Slope Value	Distance Class (m)	Distance Value	Vegetation	Vegetation Value
0-3	1	0-50	1	Pasture	1
3-8	3	50-100	5		

To be continued...

Table 4 - Conclusion.

8-20	5	100-200	13	Native Forest	35
20-45	7	200-300	25		
>45	9	>300	50		

With this final cost raster as constraining factor and considering the three other terrain attribute rasters, the analysis was conducted using R software (R Development Core Team, 2009). The cLHS scheme for collecting 25 validating samples was accomplished through the use of the following R packets: clhs (Roudier, 2012), raster (Hijmans and van Etten, 2012), proj4 (Urbanek, 2011), rgdal (Keitt et al., 2012), and shapefiles (Stabler, 2006). Those 25 validating places were chosen based only on the environmental variables and the cost raster. The predicted solum depth map was not taken into account on that procedure because its accuracy would be known with this validation. A global index, Kappa index, commission, and omission errors were calculated to evaluate how many samples out of the 25 ones for field validation correctly match the predicted solum depth class on the map.

3.3 RESULTS AND DISCUSSION

The predicted Cambisol (Inceptisols) solum depth map was created according to knowledge-based inferences about the landscape conditions more likely to characterize solum depth variability (Fig. 3). Visually, there is an obvious relationship between the shallow soils and the steepest slopes (steeper than 45%) on the landscape. The shallow sola on steep slopes are common due to the high ratio between erosion and pedogenesis. Steep slopes reduce the water infiltration, increase overland flow, and decrease water available for weathering-leaching, preventing the solum from becoming thicker. All these processes would lead to the shallow soils found on steep slopes (Schaetzl and Anderson, 2005). Dietrich et al. (1995) proposed a model based on topographic measurements to predict soil depth and also noted the tendency of thin soils forming on steep landscapes.

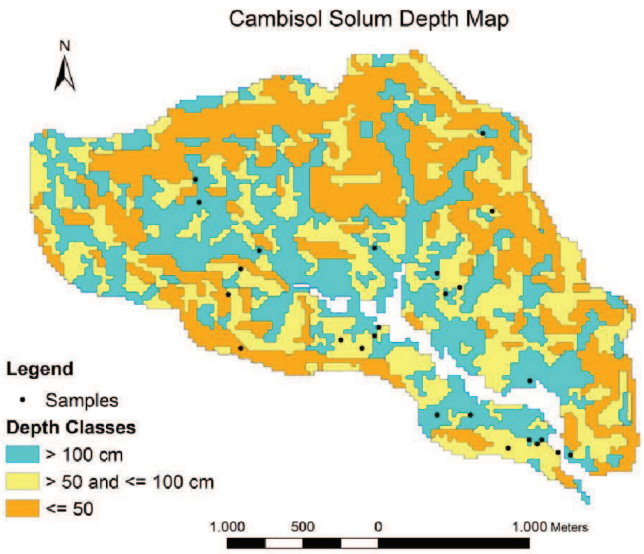

Fig. 3. Predicted solum depth map for Lavrinha Creek Watershed created using soil expert knowledge associated with fuzzy logic and environmental covariates.

The 30-m DEM was used to derive three terrain attribute rasters of slope, profile curvature, and SWI and were used for the proposed solum depth map creation (Fig. 4). These data in Fig. 4 illustrate that the lowest areas on the landscape, excluding the floodplains, are represented by high wetness indexes and gentle topography, favoring the sediment accumulation transported from higher elevation areas, which should increase the Cambisol solum thickness in those areas (Dietrich et al., 1995). Although the highest wetness indexes occur on floodplains, those areas were not included on the predicted Cambisol solum depth map for being mostly occupied by Fluvic Neosols and Haplic Gleysols (Fig. 1). The highest areas are found under the lowest wetness index values and predominantly steep slopes. Those steep conditions are generally related to shallow soils because small amounts of water are expected to reach the bedrock, which reduces the weathering rates and, hence, the soil thickening (Schaetzl and Anderson, 2005). Also, on steep side slopes, rates of erosion outpace soil development (Aquino et al., 2013).

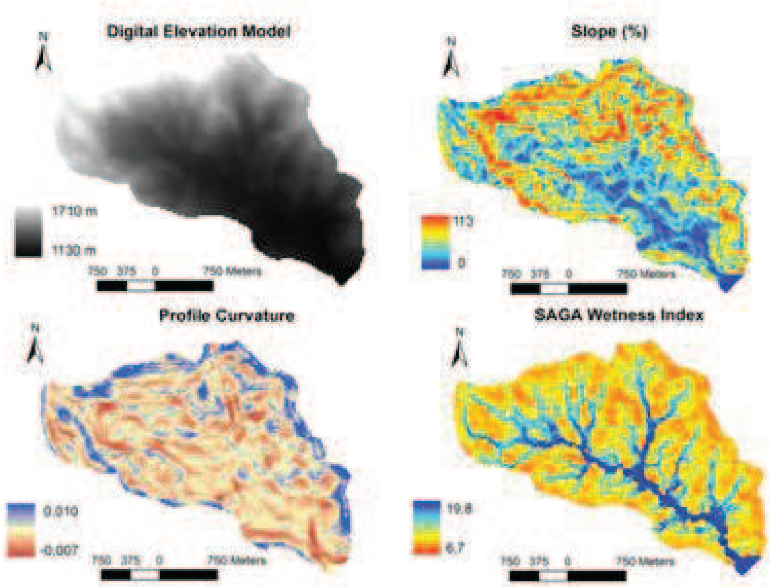

Fig. 4. Digital Elevation Model and the terrain derivative maps of slope, profile curvature, and SAGA wetness index used along with soil expert knowledge for the predicted solum depth map creation.

According to the predicted solum depth map, a solum shallower than 50 cm (Shallow class) occupies 34.9% of the area, which is the dominant class, and it is located on steep slopes (Fig. 3). This depth class is followed by Moderate Deep class solum covering 32.8%. Although the steep relief is the dominant topography in this watershed, the Deep solum class encompasses 32.3% of the area, which is the solum thicker than 100 cm. In this class, the BC horizon is generally very thick and can be as much as 40 cm. As reported by Mello and Curi (2012), the Cambisols of this watershed are deeper than the Cambisols found in similar landscapes in the rest of the Alto do Rio Grande Basin and the deeper soils provide higher recharge potential. This deeper solum at LCW has been attributed to the higher permeability of the gneiss (parent rock in LCW) when compared with the lower permeability of mica schist, parent material on major part of Alto do Rio Grande Basin, favoring the weathering rates and, hence, the solum depth.

In general, the relief at LCW is composed of narrow hills, and it makes it difficult for the 30-m DEM to capture all the relief variability in the study area. Soil and landscape vary in distances less than 30 m, so some transitions or soil variability were not captured by the DEM in this work. Thus, the coarse resolution DEM may have contributed to some errors in the predicted solum depth map. The influence of the DEM resolution (grid size) on soil landscape modeling, as well as on its applications, has been widely discussed (e.g., Vázquez et al., 2002; Claessens et al., 2005). The free higher resolution DEMs with pixels smaller than 30 m by 30 m, when available in Brazil, will possibly improve the final results of DSM products related to prediction of continuous soil properties, especially when working in places where the soil management requires large scale maps.

The low elevation and gentle-slope areas were not considered under the same validation sampling scheme and were removed from the LCW soil map (Fluvic Neosols and Haplic Gleysols) due to different geomorphic and pedogenic processes occurring in these positions when compared to the Cambisols area. To validate this removing procedure, during the field work, two locations were sampled within this area (557520, 7551502 WGr; 556194, 7552491 WGr, 23K). Both places were identified as containing lowland soils that differed from the Cambisols. Also, two soil profiles (557321, 7551784 WGr; 556825, 7552219 WGr, 23K) described by Menezes et al. (2009) containing the local soil property and soil type information (lowland soils) were added to the map in ArcMap (ESRI) to verify whether they were correctly included in that non-Cambisol area. Both pair of coordinates were located within the non-Cambisol area, confirming that the mentioned area is composed of lowland soils that should not be considered in the predicted solum map.

Figure 5 presents the cost raster map and the variables used for its creation only on the Cambisol area. The increased blue in the pixel color on the cost raster relates to higher difficulty to reach the pixel location. It is possible to see that the places surrounding the road and under pasture have lower cost of sampling than locations farther from the road and under native forest. Gentle topography places were also considered easier to reach compared to steep terrain. Thus, those low-cost sampling areas were preferred by cLHS.

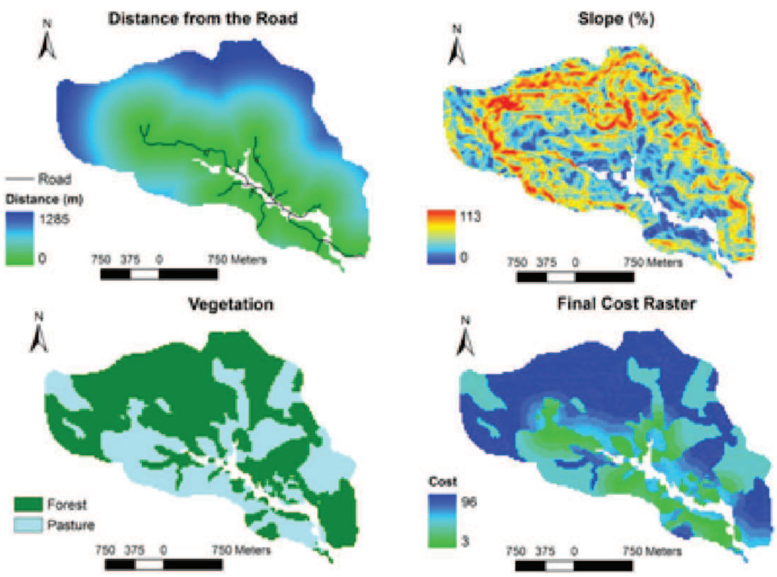

Fig. 5. Variables used for the creation of the map of the cost (difficulty) of reaching every place in the study area.

Two sampling schemes were created to demonstrate how sampling locations change with and without constraining the cLHS to low-cost places according to the cost raster (Fig. 6) and to verify whether the cost-constrained samples adequately represent soil variability according to terrain attributes. In the field, it was noticed that places where the cost-constrained validation samples were selected presented different soil attributes such as color, slope gradient, amount of gravel, moisture, and solum depth, the feature to be validated (validation results in subsequent sections). Furthermore, some samples were located in high-cost sampling locations even when conditioning them to low-cost sampling areas. The high-cost locations presented unique soil properties compared to other validating locations probably because there was no possibility for the cost-constrained cLHS to transfer them to low-cost sampling areas without losing the sampling representativeness. However, the number of relocated samples demonstrates that the cost-constrained cLHS is an efficient way to change the sampling locations to easy-to-reach areas, reducing the time and investments needed during field evaluation for validating the predicted solum map. Also, this sampling scheme can help researchers working in areas with limited access to capture soil variability, especially where the major topographic and land-use conditions are closely related.

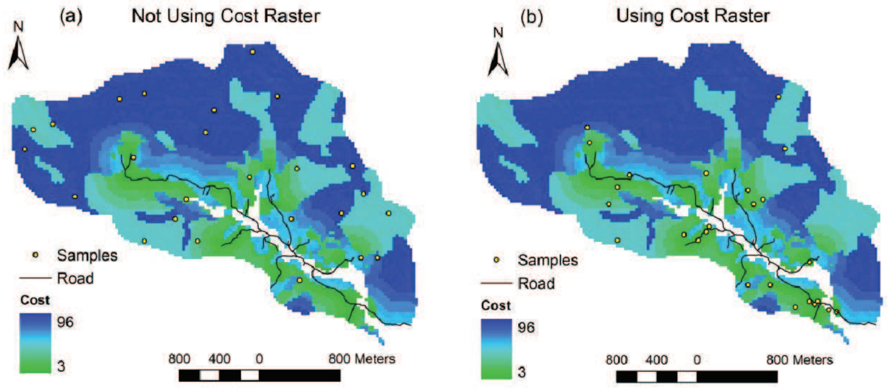

Fig. 6. (a) The standard conditioned Latin hypercube sampling and (b) cost-constrained conditioned Latin hypercube sampling locations showing how the sampling points move to easy-to-reach (low cost) areas when the sampling scheme is cost-constrained by a cost raster.

Despite this cost reduction during the field activities, from the 25 validating places selected by the cLHS, only two samples were located within the predicted Shallow class for its validation. The Shallow class locations were selected based only on the terrain attribute maps and the cost raster, not taking into account the predicted solum classes, which is needed for validation. The other two depth classes were well represented on the validation scheme, where 15 samples were located for the Deep class and 8 for the Moderate Deep class (Table 5). Roudier et al. (2012), comparing sampling schemes made by both standard cLHS and the cost-constrained cLHS, concluded that the latter did not provide as optimized sampling representativeness as the scheme using standard cLHS. The cLHS penalizes and under-represents some of the features that occur in difficult terrain. On the other hand, the same authors state that the cost of the produced sampling scheme was reduced. In future works, a comparison between different sample sizes and the cost of performing those sampling schemes in the field could be tested to determine the optimal relation between sample sizes and cost of sampling, maintaining the quality of the final maps.

Table 5 - Confusion matrix showing the number of correctly predicted places in order to assess the predicted solum depth map accuracy according to the real solum depth at each validation point.

Class	Shallow	Moderate Deep	Deep	Row Total
Shallow	1	1	0	2
Moderate Deep	0	6	2	8
Deep	0	2	13	15
Column Total	1	9	15	25

Using the data points collected according to the cLHS to verify the real solum depth at those places, the predicted solum depth map was validated through global index, Kappa index, and commission and omission error assessments. According to Story and Congalton (1986), the global index is the most common way to express the precision of both images and maps, reporting the percentage of the mapped area correctly classified in comparison with the reference data, verified in the field. Table 5 shows the confusion matrix in which the sum of the main diagonal values (1+6+13) presents the number of validating samples that correctly match the predicted solum depth class. Likewise, 20 out of the 25 samples match the predicted solum depth class, which results in a global index of 80%. This value is lower than the 85% globally accepted, according to Jensen (1986), indicating that there were some errors, most likely due to DEM pixel size (30 m) and solum depth variability in the study area. This landscape has relief with narrow hills, which are not always correctly represented on the terrain derivative maps at the 30-m resolution used for prediction. Also, the solum depth variability tends to be higher on Cambisols (this study) than on more pedogenetically developed soils (Oliveira et al., 2013).

Kappa coefficient considers the proportion of the correctly classified samples corresponding to the ratio between the sum of the numbers on the confusion matrix main diagonal (correctly classified samples) and the sum of all of the matrix components (number of total samples), taking as reference the total number of classes (Congalton and Green, 2009). The Kappa index, calculated according to the confusion matrix shown on Table 5 resulted in 0.616, a value which corresponds to a substantial classification according to Landis and Koch (1977). Kappa coefficient value is lower than the global index because the former takes into account all of the confusion matrix values and not only the values on the main diagonal as does the global index.

Omission and commission errors were assessed to evaluate the predicted solum map quality per depth class. Omission error is defined as the exclusion of a sample from the class that it actually belongs, while the commission error is to include a sample in a class in which it should not have been included. Omission and commission errors for the solum depth classes are shown on Table 6, and the lower the values, the more accurate the predicted solum depth class. In this study, the commission error for the Shallow class was the lowest, which means the sample presented the same depth class predicted on the map, whereas the omission error was the greatest (50%) because the shallow class contains two validation samples and the one collected where the solum is shallow in the field occurs where the map indicates the Moderate Deep class. However, due to this class containing one sample indicated by the cLHS for validation, the errors can easily be either increased or decreased. On the other hand, Deep class has low error values, showing that the classification of this solum depth class was acceptable, being only two samples misclassified out of the 15 belonging to this class. Also, most validating samples were located on this depth class (15 out of 25). The Moderate Deep class presented intermediate values and also adequate classification, containing only two errors out of eight samples.

Table 6 – Omission and commission errors calculated according to the number of validation points with correctly predicted solum depth per solum depth class.

Class	Omission error	Commission error
Shallow	[1-1/1)]*100 = 0	[1-(1/2)]*100 = 50
Moderate Deep	[1-(6/9)]*100 = 33.33	[1-(6/8)]*100 = 25
Deep	[1-(13/15)]*100 = 13.33	[1-(13/15)]*100 = 13.33

The actual solum depth measured in the field and the predicted solum class for each validation point are presented in Table 7. The results show that five out of the 25 samples did not match the real depth and the predicted depth class. This result is significant considering that the predicted solum depth map was totally based on soil expert knowledge that defined the topographical conditions for each solum depth class for LCW. This result supports the advantages of considering a soil expert who knows the area of interest as a strong, reliable, low cost, and helpful mapping tool to improve results, especially in countries such as Brazil where some of the fundamental elements of a soil mapping project are difficult to obtain (i.e., pedological information, access due to rugged terrain, high resolution DEMs, and funding). Bui (2004) states that soil maps and their legends are representations of structured knowledge, namely the soil surveyor's mental soil–landscape model. Thus, one who understands soil distribution in an area of interest could apply the acquired knowledge to aid soil mapping. Also, the produced solum depth map may aid planning soil use and management at LCW, drawing attention for areas where the solum is shallow and, thus, more susceptible to erosion.

Table 7 – Predicted solum depth classes confronted with the real solum depth verified in the field at 25 validating places selected by conditioned Latin Hypercube Sampling scheme.

Validation Point	Predicted Solum Depth (cm)	Real Solum Depth (cm)	Validation Point	Predicted Solum Depth (cm)	Real Solum Depth (cm)
1	> 100	> 100	14	> 50 and ≤ 100	57
2	> 50 and ≤ 100	90	15	> 50 and ≤ 100	> 100
3	> 100	> 100	16	≤ 50	46
4	> 50 and ≤ 100	74	17	> 50 and ≤ 100	96
5	> 100	> 100	18	> 100	79
6	≤ 50	65	19	> 100	> 100
7	> 50 and ≤ 100	63	20	> 100	> 100
8	> 100	> 100	21	> 50 and ≤ 100	> 100
9	> 100	> 100	22	> 100	> 100
10	> 100	> 100	23	> 100	> 100
11	> 50 and ≤ 100	92	24	> 100	> 100
12	> 100	86	25	> 100	> 100
13	> 100	> 100	--	--	--

3.4 SUMMARY AND CONCLUSIONS

The new advances in technology provide tools for soil scientists, which can be beneficial for soil survey and mapping in areas with limited access and data. The experts can provide an idea about patterns that may be found in the study area. This is very important in Brazil because it is difficult to gather enough information about a specific region before the mapping process begins. It is worthy to highlight that even employing the most advanced technologies and local pedologist knowledge for predicting soil properties, predictions may not always be totally correct due to soils variability among different regions, primarily in a large country such as Brazil. This statement supports the importance of field validation of predictive soil maps, a procedure that makes the soil map more usable and reliable. The predicted solum thickness map presented adequate validation results for an area with very limited spatial information about this soil property, supporting and confirming the importance of soil expert knowledge application as a soil mapping tool in association with the new computational assessment tools. Also, the terrain attributes applied were efficient to predict solum depth, corroborating the findings of Gessler et al. (2000). These types of functional maps are crucial for addressing many land-use questions and in particular the issue of soil-water storage for recharge into streams that supply hydroelectric power plants. Land-use planners will increasingly need more refined maps, and these technologies will be useful for future developments.

The new mapping techniques and DEM availability for soil scientists tend to somehow compensate the lack of previous pedologic information aiding the creation of predictive maps for the interest area, especially in Brazil, where more detailed soil maps are only available for small areas. Also, the cost-constrained cLHS, although providing a less optimized sampling scheme in comparison to standard cLHS (Roudier et al., 2012), would be very useful for Brazilian conditions, for limiting the sampling in low-cost areas, since access difficulties due to absence of roads and dense vegetation are common. Thus, it presents a practical way to choose the soil sampling places to save time and expenses in the field work.

The soil expert knowledge in association with fuzzy logic is a good way to support soil map creation, especially when it is difficult to appropriately sample throughout the study area, which is common. This method of soil mapping can use the soil–landscape relationships to extrapolate information based on some environmental support and does not require as many samples as geostatistical approaches. The weakness with maps using expert knowledge is that maps may be created differently by different experts. However, the maps are explicit and the information needed to create them is recorded so that future updates can be easily performed and the knowledge of the expert can be captured. In this context, the association of fuzzy logic and a soil expert is even more important not only because fuzzy logic provides a way to classify soils as a continuum, capturing their spatial variability (Zhu et al., 1997) but also because a soil expert can provide more detailed information than the existing maps. Furthermore, a pedologist can inform, with higher reliability, the soils more likely to occur in places very hard to reach, based on soil–landscape relationships, which would diminish the uncertainties about soils in those locations. Non-pedologists should consider maps to be dynamic rather than static maps. We can produce the best maps possible with the available data with the expectation that the maps will be useful for current land management, and also the maps can be constantly improved with additional data to make improvements of property predictions with associated uncertainties.

REFERENCES

Aquino, R.F., M.L.N. Silva, D.A.F. Freitas, N. Curi, and J.C. Avanzi. 2013. Soil Losses from typic Cambisols and Red Latosol as related to three erosive rainfall patterns. R. Bras. Ci. Solo. 37:213-220.

Boer, M., G. Del Barrio, J. Puigdefabres. 1996. Mapping soil depth classes in dry mediterranean areas using terrain attributes derived from a digital elevation model. Geoderma 72(1–2):99–118. doi:10.1016/0016-7061(96)00024-9

Böhner, J., and O. Conrad. 2009. System for Automated Geoscientific Analyses (SAGA) 2.0.5. http://sourceforge.net/projects/saga-gis/files/. Accessed 20 Feb 2011.

Brungard, C.W. and J.L. Boettinger. 2010. Conditioned Latin Hypercube Sampling: Optimal Sample Size for Digital Soil Mapping of Arid Rangelands in Utah, USA. In: J.L. Boettinger et al. (eds.), Digital Soil Mapping, Progress in Soil Science 2, Springer Neatherlands, Dordrecht, Neatherlands. p.67-75. doi:10.1007/978-90-481-8863-5_6

Bui, E.N. 2004.Soil survey as a knowledge system. Geoderma. 120(1–2):17–26. doi: 10.1016/j.geoderma.2003.07.006

Claessens, L., G.B. Heuvelink, J.M. Schoorl, and A. Veldkamp. 2005. DEM resolution effects on shallow landslide hazard and soil redistribution modeling. Earth Surf. Proc. Land. 30:461–477. doi: 10.1002/esp.1155

Congalton, R.G. and K. Green. 2009. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. 2nd Ed. CRC/Taylor and Francis, Boca Raton, 183p.

Dietrich, W.E., R. Reiss, M.-L. Hsu, and D.R., Montgomery. 1995. A process-based model for colluvial soil depth and shallow landsliding using digital elevation data. Hydrol. Process. 9:383–400. doi: 10.1002/hyp.3360090311

Embrapa – Empresa Brasileira de Pesquisa Agropecuária. 2013. Sistema Brasileiro de Classificação de Solos. 3rd ed. Rio de Janeiro, 353 p.

Gessler, P.E., O.A. Chadwick, F. Chamran, L. Althouse, and K. Holmes. 2000. Modeling soil landscape and ecosystem properties using terrain attributes. Soil Sci. Soc. Am. J. 64:2046–2056.

Giasson, E., R.T. Clarke, A.V. Inda Junior, G. H. Merten, and C.G. Tornquist. 2006. Digital soil mapping using multiple logistic regression on terrain parameters in Southern Brazil. Sci. Agr. 63(3):262-268. doi: 10.1590/S0103-90162006000300008

Grunwald, S. 2009. Multi-criteria characterization of recent digital soil mapping and modeling approaches. Geoderma. 15:195–207. doi: 10.1016/j.geoderma.2009.06.003

Hijmans, R.J., and J. van Etten. 2012. raster: Geographic analysis and modeling with raster data. R package version 2.0-04. http://CRAN.R-project.org/package=raster. Accessed 20 June 2012.

Iwashita, F., M.J. Friedel,; G.F. Ribeiro, and S.J. Fraser. 2012. Intelligent estimation of spatially distributed soil physical properties. Geoderma. 170:1–10. doi: 10.1016/j.geoderma.2011.11.002

Jenny, H. 1941. Factors of Soil Formation: A System of Quantitative Pedology. McGraw-Hill, New York, 281p.

Jensen, J.R. 1986. Introductory digital image processing. Englewood Cliffs: Prentice - Hall, 51 p.

Keitt, T.H., R. Bivand, R. Pebesma, and B. Rowlingson. 2012. rgdal: Bindings for the Geospatial Data Abstraction Library. R package version 0.7-11. http://CRAN.R-project.org/package=rgdal. Accessed 20 June 2012.

Kuriakose, S.L., S. Devkota, D.G. Rossiter, and V.G. Jetten. 2009. Prediction of soil depth using environmental variables in an anthropogenic landscape, a case study in the Western Ghats of Kerala, India.Catena. 79:27-38. doi: 10.1016/j.catena.2009.05.005

Lagacherie, P. and Voltz, M. 2000. Predicting soil properties over a region using sample information from a mapped reference area and digital elevation data: a conditional probability approach.Geoderma. 97:187-208. doi: 10.1016/S0016-7061(00)00038-0

Landis, J.R., and G.G. Koch.1977. The measurement of observer agreement for categorical data. Biometrics. 33(1):159–174.

McKay, J., S. Grunwald, X. Shi, and R.F. Long. 2010. Evaluation of the transferability of a knowledge-based soil-landscape model. In: Boettinger, J. L.

et al. (Ed.). Digital soil mapping: bridging research, environmental application, and operation. London: Springer, p.165-177. doi: 10.1007/978-90-481-8863-5_14

McKay, M.D., R.J. Beckman, and W.J. Conover. 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics. 21:239–245.

Mendonça-Santos, M.L., H.G. Santos, R.O. Dart, and J.G. Pares. 2008. Digital mapping of soil classes in Rio de Janeiro State, Brazil: data, modeling and prediction. In: Hartemink, A.E., A. McBratney, and M.L. Mendonça-Santos, (Org.). Digital soil mapping with limited data. Elsevier, Amsterdam. p.381-398. doi: 10.1007/978-1-4020-8592-5_34

McBratney, A.B., M.L. Mendonça-Santos, and Minasny, B. 2003. On digital soil mapping. Geoderma. 117:3-52. doi: 10.1016/S0016-7061(03)00223-4

Mello, C.R., and Curi, N. 2012. Hydropedology. Cienc. Agrotec. 36(2):137-146. doi: 10.1590/S1413-70542012000200001

Menezes, M.D. , J.A. Junqueira Junior, C.R. Mello, A.M. Silva, N. Curi, and J.J. Marques. 2009. Dinâmica hidrológica de duas nascentes, associada ao uso do solo, características pedológicas e atributos físico-hídricos na sub-bacia hidrográfica do Ribeirão Lavrinha, Serra da Mantiqueira, MG. Sci. For., Piracicaba, 37(1):175-184.

Menezes, M.D., S.H.G. Silva, P.R. Owens, N. Curi. 2013. Digital Soil Mapping Approach Based on Fuzzy Logic and Field Expert Knowledge. Cienc.Agrotec. 37(4):287-298. doi: 10.1590/S1413-70542013000400001.

Milne, G. 1935. Some suggested units of classification and mapping particularly for East African soils. Soil Res. 4:183-198.

Minasny, B., and A.B. McBratney. 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. Comput. Geosci. 32:1378–1388. doi: 10.1016/j.cageo.2005.12.009

Minasny, B. and A.B. McBratney. 2010. Methodologies for global soil mapping. In: Boettinger, J., D. Howell, A. Moore, A. Hartemink, and S. Kiesnast-Brown (eds.), Digital Soil Mapping: Bridging Research, Environmental Application, and Operation, Progress in Soil Science. New York: Springer. p.429-436. doi: 10.1007/978-90-481-8863-5_34

Oliveira, A.H., M.L.N. Silva, N. Curi, J.C. Avanzi, G.K Neto, and E.F. Araújo. 2013. Water erosion in soils under eucalyptus forest as affected by development stages and management systems. Cien. Agrotec. 37(2):159-169. doi: 10.1590/S1413-70542013000200007

Quinn, T., A.X. Zhu, and J.E. Burt. 2005. Effects of detailed soil spatial information on watershed modeling across different model scales. Int. J. Appl. Earth Obs. Geoinf. 7:324–338. doi: 0.1016/j.jag.2005.06.009

R Development Core Team. 2009. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. http://www.r-project.org. Accessed 20 June 2012.

Rezende, S.B.R. 1980. Geomorphology, Mineralogy and Genesis of four soils on gneiss in Southeastern Brazil. Purdue University (Ph.D. Thesis). 143p.

Roudier, P. 2012. clhs: a R package for conditioned Latin hypercube sampling. http://cran.r-project.org/web/packages/clhs/clhs.pdf. Accessed 20 June 2012.

Roudier, P., D.E. Beaudette, and A.E. Hewitt. 2012. A conditioned Latin hypercube sampling algorithm incorporating operational constraints. In: Minasny, B, B.P. Malone, A.B. McBratney (eds.), Digital Soil Assessments and Beyond: Proceedings of the 5th Global Workshop on Digital Soil Mapping 2012, Sydney, Australia, 10-13 April 2012. CRC Press, Boca Raton, FL. 482p.

Schaetzl, R. J. and S. N. Anderson. 2005. Soils: Genesis and Geomorphology. Cambridge University Press, Cambridge, UK. 827p.

Shi, X. 2013.ArcSIE user's guide. http://www.arcsie.com/index.htm. Accessed 4 July 2013.

Stabler, B. 2006.shapefiles: Read and Write ESRI Shapefiles. R package version 0.6. http://cran.r-project.org/web/packages/shapefiles/index.html. Accessed 20 June 2012.

Story, M., and R.G. Congalton. 1986. Accuracy assessment: a user's perspective. Photogram. Eng. Remote Sens., 52(3):397-399.

Urbanek, S. 2011. proj4: A simple interface to the PROJ.4 cartographic projections library. R package version 1.0-7. http://CRAN.R-project.org/package=proj4. Accessed 20 June 2012.

Vázquez, R.F., L. Feyen, J. Feyen, and J.C. Refsgaard. 2002. Effect of grid size on effective parameters and model performance of the MIKE-SHE code. Hydrol.Process., 16:355–372. doi: 10.1002/hyp.334

Zhu, A.X., and L.E. Band. 1994. A knowledge-based approach to data integration for soil mapping. Can. J. Remote Sens.. 20(4):408-418.

Zhu, A.X., L. Band, R. Vertessy, and B. Dutton. 1997. Derivation of soil properties using a soil land inference model (SoLIM). Soil Sci. Soc. Am. J. 61(2):523-533.

Zhu, A.X., A. Moore, and J.E. Burt. 2010. Prediction of soil properties using fuzzy membership values. Geoderma. 158(3/4):199-206. doi: 10.1016/j.geoderma.2010.05.001

4. ARTIGO 3. Solum depth spatial prediction comparing conventional with knowledge-based digital soil mapping approaches

***Artigo nas normas da Scientia Agricola.**

ABSTRACT

Solum depth and its spatial distribution play an important role on different types of environmental studies. Several approaches have been used for fitting quantitative relationships between soil properties and their environment in order to predict them spatially. This work aimed to present the steps required for solum depth spatial prediction from knowledge-based digital soil mapping, comparing its prediction to the conventional soil mapping approach through field validation, in a watershed located at Mantiqueira Range region, Minas Gerais State. Conventional soil mapping had aerial photo-interpretation as a basis. The knowledge-based digital soil mapping applied fuzzy logic and similarity vectors in an expert system. The knowledge based digital soil mapping approach showed the advantages over the conventional soil mapping approach by applying the field expert-knowledge in order to enhance the quality of final results, predicting solum depth with suited accuracy in a continuous way, making the soil-landscape relationship explicit.

Keywords: Soil Survey, Fuzzy Logic, Similarity Vectors.

4.1 INTRODUCTION

The solum depth (A+B horizon) has been applied in distributed hydro-ecological models to simulate watershed processes as net photosynthesis and stream flow (Quinn et al., 2005; Zhu and McKay, 2001), affecting the soil storage capacity (Follain et al., 2007) or soil drainage condition (Odeh et al., 1995). Solum depth is strongly linked to landscape characteristics and it is important for soil mapping (Chartin et al., 2011), land use planning and management.

Several approaches have been used for fitting quantitative relationships between soil types and/or properties and their environment in order to predict their spatial distribution and variability (spatial inference models) (McBratney et al., 2003). Such models are divided into data-driven (Pedometric approach) and knowledge-driven (Shi et al., 2009). From pedometric approach (statistic and geostatistic), the accuracy of prediction is generally related to a dense sampling scheme, which is not always feasible due to cost and time constraints (Zhu and Lin, 2010).

Zhu and Band (1994) and Zhu (1997) presented an alternative approach based on limited observations per soil class, using fuzzy logic and similarity vectors, in an expert system. Possessing the maps that represent soil forming factors (environmental variables), the knowledge of pedologists can be incorporated into spatial prediction, where the qualitative soil-landscape model is converted to quantitative predictions using relationships between soils and, more frequently, terrain attributes, such as slope, topographic wetness index, and profile curvature. It overcomes a limitation of conventional soil mapping approach, as raised by Hudson (1992), which failures in not expliciting the soil surveyor mental model. Because this approach requires an understanding from a soil scientist perspective on the repeating soil patterns on the landscape, as conventional mapping approach, it is considered a knowledge-driven digital soil mapping approach and it has been considered efficient and economical (Hudson, 1992; MacMillan et al., 2007).

This work aimed to present the steps required for solum depth spatial prediction from knowledge-based digital soil mapping, comparing it to the conventional soil mapping approach through field validation, in a watershed located at Mantiqueira Range region, Minas Gerais State.

4.2. MATERIAL AND METHODS

4.2.1 Study area characterization

This study was carried out at Lavrinha Creek Watershed located at Mantiqueira Range, Southern Minas Gerais state (Figure 1). It is a typical headwater watershed, representative of the Alto Rio Grande Basin, an important hydrological region due to its potential to generate electric energy on the basis of hydraulic energy. There is predominance of dense rain forest, with high slope gradients and few roads, hindering the access and the traffic in the area. The main characteristics of the study site are presented in Table 1.

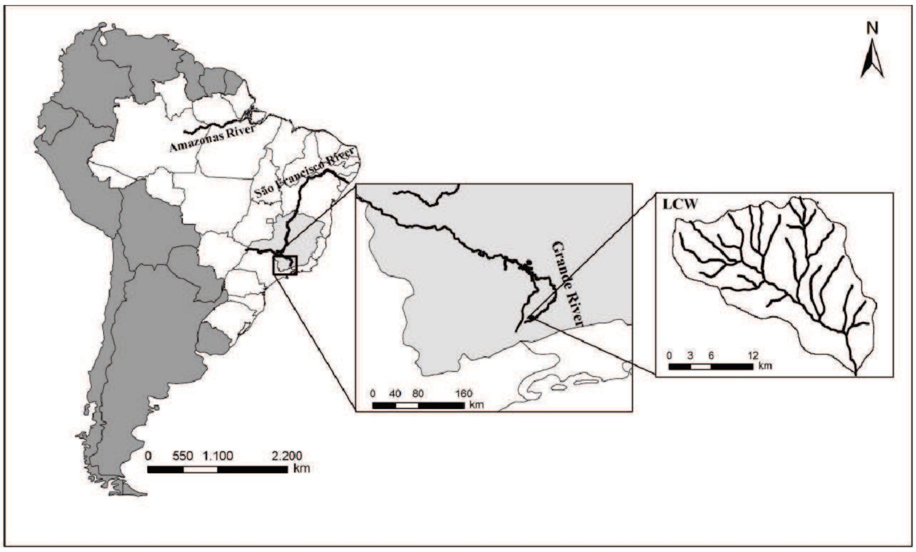

Figure 1. Geographical location of Lavrinha Creek Watershed.

Table 1. Basic characteristics of Lavrinha Creek Watershed.

	Basic characteristics
Location	Between latitudes 22º6’53.7” and 22º8’28.1” S and longitudes 44º26’21.1” and 44º28’39.2” W
Area	676 ha
Elevation	From 1151 to 1687 m
Mean annual temperature	15 °C
Annual Precipitation	2000 mm
Land agricultural suitability	Fauna and flora reserve
Parent material	Alluvial material transported by water (floodplains) and the massive rock gneiss (high lands)

4.2.2 Conventional soil mapping approach

The photo-interpretation of the watershed was performed using a stereoscope, with vertical pancromatic aerial photograph at a scale of approximately 1:35,000. Physiographycally homogeneous areas were separated, which constituted the preliminary mapping units. This map was further tuned in the field. The correlation between map units and landform features were verified and boundaries were redrawn when necessary. This work aided to select representative sites for describing soil profiles and making prospections. After the photo interpretation procedure, the landform map was digitalized and displayed in Geographic Information Systems (GIS) environment (ESRI, 2010), in which places easily identified in the photo and in the Statistics and Geography Brazilian Institute (IBGE) maps were used for georeferencing the landform maps. With respect to solum depth, each soil map unit assumes a unique value based on the soil profile described, which represents the central or modal concept for that soil map unit.

4.2.3 Knowledge-based digital soil mapping approach

The steps required to predict solum depth were mostly accomplished in ArcSIE (Soil Inference Engine) version 9.2.402, a toolbox that functions as an ArcMap extension (Shi et al., 2009). ArcSIE is designed for creating soil maps using fuzzy logic and supports the knowledge based approach to establishing the relationships between soil and its environment, providing tools for soil scientists to formalize the relationship based on pedological knowledge of the local soils. A knowledge-based digital soil mapping is performed according to existing relationships between soil attributes and landforms. The landforms can be obtained from DEM derivatives that create the Digital Terrain Models (DTMs) in a GIS environment. From DTMs and pedological information, soil-landscape relationships can be employed for extrapolating information to non-sampled places through mapping techniques (fuzzy logic and similarity vectors) (Zhu, 1997). In order to predict the solum depth, the following steps were conducted:

a) Establishing soil-landscape relationships to predict soil classes.

This step is the basis for setting rules and was based on soil scientist's knowledge, maps from previous soil survey and other types of soil research developed in the study site. Considering the soil-landscape relationships at LCW, the alteration of gneiss resulted in predominance of Udepts (moderately developed and well-drained soils) (US Soil Taxonomy - Soil Survey Staff, 1999). The relief is steep with concave-convex slopes, predominated by linear pedoforms and narrow floodplain. Hydromorphic soils occupy the toeslope position, where the water table is near to the surface in most part of the year.

b) Quantifying relationships between soils and terrain attributes and formalizing these relationships in a set of rules

Analogous to a Digital Elevation Model (DEM), DTMs are represented in an ordered array of numbers that represent the spatial distribution of terrain attributes across a landscape, in a raster-based format. Terrain models were based on a 30 m resolution DEM, generated from the Brazilian source of contour lines at 1:50,000 scale (IBGE, 1973). The sinks were filled and a hydrologically consistent DEM was created using ArcGIS version 10.0 (ESRI, 2010). In order to calculate the terrain attributes from DEM, the System for Automated Geoscientific Analysis (SAGA) (Böhner et al., 2006), version 2.0.8, ArcMap spatial analyst and ArcMap extension Soil Inference Engine (ArcSIE), version 9.2.402 were used. The following primary (calculated directly from DEM) and secondary (calculated from the combination of two or more primary terrain attributes) terrain attributes were derived from DEM:

- Primary: slope is the gradient of elevation. Profile curvature is the slope shape in the direction of the maximum slope and is, therefore, important for water flow. Plan curvature is the slope shape perpendicular to the slope direction, which measures the convergence or divergence and, hence, the concentration of water in a landscape (Moore et al., 2003);

- Secondary: SAGA wetness index (WI) was used instead of well-known topographic wetness index ($ln(a/tan\beta)$), where a - ratio of upslope contributing area per unit contour length and β - local slope). Both wetness indexes are similar, however, in SAGA it is possible to adjust the width and convergence of the WI multidirectional flow to single directional flow. Large WI values indicate an increase likelihood of saturated conditions and are usually found in lower parts and convergent hollow areas and soils with small hydraulic conductivity or areas of gentle slope (Beven and Wood, 1983). These indices have been used to identify water flow characteristics in landscape (Sumfleth and Duttmann, 2008).

Soil-landscape relationships were qualitatively modeled using DTMs, based on the terrain attributes that represent soil and hydrologic processes. Next, a qualitative soil landscape model from step a) was used to quantify soil-landscape relationships on a continuous basis, based on different terrain attributes and their histogram distribution values. For the model development, a set of rules for the entire watershed was created for each soil map unit (step a) and applied in ArcSIE in order to create a soil map for the entire watershed. ArcSIE provides different types of knowledge integration. In this work, the rule-based reasoning was applied allowing for the covering of the entire mapping area (Shi et al., 2009).

The soil-landscape relationships were extracted and the characterized environmental conditions were linked through a set of inference techniques to populate the similarity model for the area (Zhu and MacKay, 2001). The terrain attribute values and ranges associated with each soil map class were used to define membership or optimality functions (curves), which define the relationship between the values of an environmental feature and soil type. The initial output from the inference is a series of fuzzy membership maps in raster format, one for each soil type under consideration (Shi et al., 2009). The fuzzy membership values represent the similarities of each pixel in the landscape to the soil types. Then, these fuzzy membership maps are combined into one final soil class map, in which only the soils with highest membership are assigned to that pixel.

c) Creating soil property map (solum depth)

After creating the soil class map, the soil property map (solum depth) can be created. This technique allows the prediction in a continuous way of any soil property that shows a recognizable relationship with the terrain attribute or landscaping position. Based on fuzzy membership values, the continuous variation of soils can be represented by continuous solum depth derived from the similarity vectors, using the following formula (Zhu et al., 1997):

$$V_{ij} = \frac{\sum_{k=1}^{n} S_{ij}^{k} * V^{k}}{\sum_{k=1}^{n} S_{ij}^{k}}$$

where V_{ij} is the estimated solum depth at location *(i,j)*, V^{k} is a typical value of soil type *k* (e.g. Udepts), and *n* is the total number of prescribed soil classes for the area. The typical value consists of the central concept of the soil type, and corresponds to these same soil profiles used in the conventional soil mapping approach. If the local soil formative environment characterized by a GIS resembles the environment of a given soil category (solum depth), then property values of the local soil should resemble the property values of the candidate soil type. The resemblance between the environment for soil at (*i,j*) and the environment for soil type *k* is expressed by S_{ij}^{k} , which is used as an index to measure the level of resemblance between the soil property values of the local soil and soil category (Zhu et al., 2001). The property value S_{ij}^{k} can be any property that shows a recognizable pattern or relationship with the terrain attribute or landscape position. The higher the membership of a local soil in a given soil type, the closer the property values (solum depth) at that location will be to the typical property values (Zhu et al., 2010).

Based on the 5 soil class map units established from step a) and the resulting fuzzy membership map from step b) measured solum depth values from the five soil profiles were assigned to their respective fuzzy soil membership maps.

4.2.4 Assessment of accuracy of solum depth prediction

A set of data containing the solum depth information was obtained for assessing the accuracy of solum depth maps (conventional and digital). In order to compare the solum depth information (real vs. estimated) contained on the soil map created as the conventional manner to the one on the rule-based reasoning map, the R^2 and R^2_{adj}, the mean error (ME) and root mean square error (RMSE) calculated through the formulas below, using R software, (R Development Core Team) were adopted for comparison purposes:

$$ME = \frac{1}{n}\sum_{i=1}^{n}(ei - mi)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(ei - mi)^2}$$

where: *n* is the number of observations, *ei* is the estimated value of the solum depth and the *mi* is the measured value of the solum depth.

4.3 RESULTS AND DISCUSSION

4.3.1 Conventional soil survey

The map of landforms (Figure 2A) was the basis for creating the soil class map (Figure 2B). The following landforms were identified: convex hills, ravined hills, steep slopes, alluvial plains, and embedded valley. The relief played an important role on soil distribution, since it is the only varying factor in the study area out of the five soil forming factors (Jenny, 1941). The removal of soil through geologic erosion from the steepest portions of relief and material accumulation by alluvial addition in floodplains explain the spatial variability of Udepts in the first case, and Fluvents and Acquents in the second case. Figure 2B shows the soil profiles used for assigning solum depth and the validation points for comparing conventional and digital knowledge-based solum depth maps.

The conventional soil map has only one solum depth assigned to each soil polygon map unit from soil profile, and does not necessarily reflect the variability and continuous nature of solum depth within and between soil polygon map units. The polygon model assumes a discrete distribution with definite boundaries, in which spatial generalization occurs due to scale limitations. Delineations smaller than the minimum mappable area, according to the soil survey scale, are included in a larger polygons and their actual spatial locations are lost (Zhu, 1997). The polygon represents only the distribution of a set of prescribed soil classes (central concepts of the soil), and other minor soil classes/minor components are not spatially represented. Pedologist knows that there are local soils that differ from the central concepts of the assigned class, but this expert knowledge cannot be conveyed using polygon-based soil mapping (Zhu et al., 2001). This procedure results in a simplification of a solum depth mapping and loss of information.

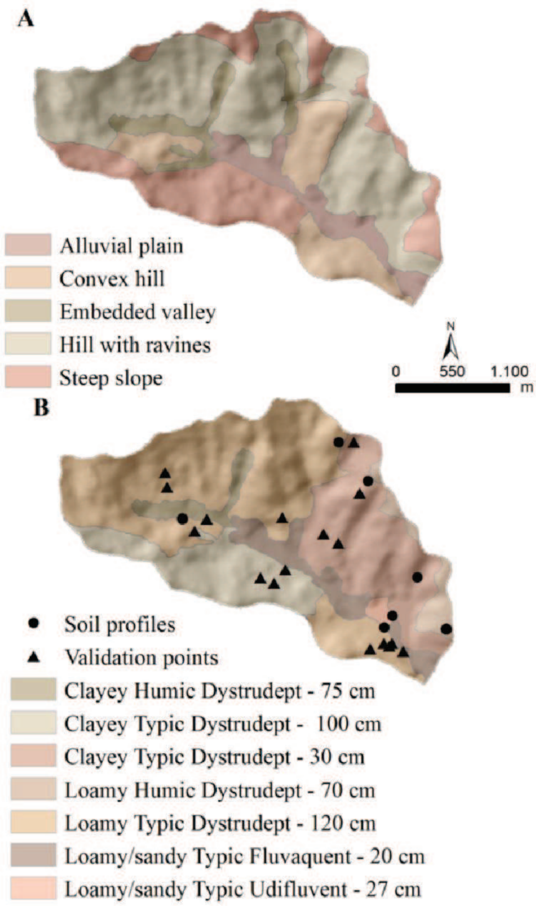

Figure 2. A) Map of landforms from aerial photointerpretation; B) conventional soil map, validation points, soil profiles used for assigning solum depth and the respective solum depth prediction for each mapping unit at Lavrinha Creek Watershed – MG.

4.3.2 Knowledge-based digital soil mapping

The DTMs used in the prediction are presented in the Figure 3. These models numerically describe the surface form as a continuum, which is more appropriate to represent geographic features than the discrete polygon model. According to Zhu (1997), pixel-based maps minimize the discrepancies between the spatial resolution of soil spatial information and environmental data (relief in this case).

The ranges and curve shapes (Table 2) that define the modal soil types were adjusted using DTMs. This table represents information of optimality curves that describe quantitatively the relationships between soil type and a particular DTM (Zhu et al., 1997). It overcomes a limitation of conventional soil mapping approach, as described by Hudson (1992), which is the failure to not represent the soil surveyor mental model. Figure 4 shows two examples of curve shapes used in this study. For the bell-shaped (a), the optimality value decreases as the difference between the environmental feature value and the central values ($v1$ and $v2$) increases. For example, in Table 2, for classifying any place in the landscape as clayey Typic Dystrudept, the optimal (central) slope value to receive 100% membership is 15 and the curve shape is bell, which indicates that as slope values decrease from 15 to 10 or increase to 20 the pixels will receive membership values decreasing from 100% to 50%, those latter being, therefore, less characteristic for clayey Typic Dystrudept to occur (Figure 4A). On the other hand, the Z-shaped or the-lower-the-better shape curve defines that all the values inferior to the central one will correspond to 100% membership. In Table 2, this curve type is used to define the typical conditions for loamy/sandy Typic Udifluvent occurrence. The full membership altitude value was defined as 1156 m, and all of the altitude values smaller than 1156 m will also receive 100% membership due to the Z-shaped curve. However, as the altitude increases to 1200 m, the membership decreases until it reaches 50% (Figure 4B), indicating environmental conditions less characteristics of loamy/sandy Typic Udifluvent.

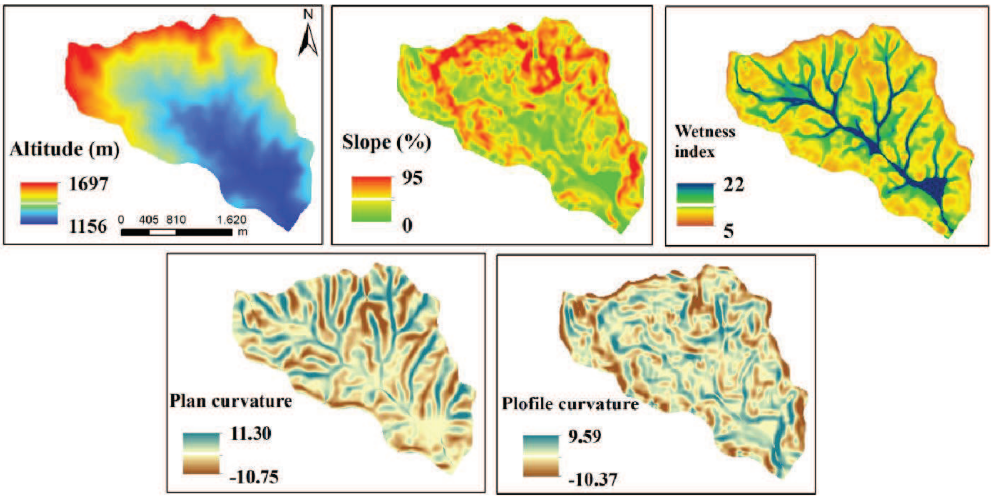

Figure 3. Digital terrain models for Lavrinha Creek Watershed.

Table 2. Environmental control variables of soil classes at LCW.

Soil type	Full membership				
	Altitude	Slope	WI	Plan curvature	Profile curvature
Fluvents	1156	1	15; 22	-	-
Udepts1	-	32.5	7	1	2.3
Udepts2	-	15	7	-1	0
Udepts3	-	32.5	7	-1	0
Udepts4	-	51	7	-1	0
	50 % membership				
Fluvents	1.200	10	14; 22	-	-
Udepts1	-	19.5; 45.5	0; 14	0.11; 3	1.56; 9.5
Udepts2	-	10; 20	0; 14	-11; 0	-1.5; 1.5
Udepts3	-	19.5; 45.5	0; 14	-11; 0	-1.5; 1.5
Udepts4	-	45; 95	0; 14	-11; 0	-1.5; 1.5
	Curve shape				

To be continued...

Table 2 - Conclusion.

Fluvents	Z	Z	Bell	-	-
Udepts1	-	Bell	Bell	Bell	Bell
Udepts2	-	Bell	Bell	Bell	Bell
Udepts3	-	Bell	Bell	Bell	Bell
Udepts4	-	Bell	Bell	Bell	Bell

Figure 4. Bell-shape (A) and Z-shape (B) optimality curves adjusted in ArcSIE interface.

According to Table 2, higher values of WI and low slopes were used for mapping hydromorphic soils in flatter alluvial areas (footslope). Udepts occupy the well-drained portions of the landscape with lower values of WI (summit, shoulder and backslope) formed by different combinations and ranges of slope, plan and profile curvatures that represents different landforms. This procedure reduced the inconsistency and costs associated with the conventional manual processes (Zhu et al., 2001).

A fuzzy logic based on the model called similarity vector (Zhu, 1997) represents soils at a given location perceiving the landscape as a continuum. The fuzzy logic is used to infer the membership of a soil type from environmental variables, such as digital elevation model and its derivative maps. A soil at a given pixel (*i,j*) is represented by a *n*-element of similarity vector: $S_{ij} = (S_{ij}^1$, S_{ij}^2, ...,S_{ij}^k,...$S_{ij}^n)$, where *n* is the number of prescribed soil types over the area, S_{ij}^k is an index which measures the similarity between the local soil at (*i,j*) to the prescribed soil type *k*. The similarity value is measured according to how close the soil is to centroid concept (between 1 and 0). The more similar a soil is to a prescribed soil type, the higher its similarity value (fuzzy membership). The soil class, as well as the continuous spatial prediction is done under fuzzy assignment, which a soil object can be labeled as more than one soil type with different degrees of assignment depending on the similarities between the soil and a set of prescribed soil classes. The more similar a soil is to a prescribed soil type, the higher its similarity value, and from a fuzzy perspective, such values are the same as fuzzy memberships of the local soil to a soil type (Zhu et al., 2010).

Figure 5 shows the fuzzy membership maps created according to the instances for the five soil types (Table 2). They are the first product generated by the inference process. Every pixel is classified assuming a value ranging from 0 to 100, being high or low according to its similarity to the soil class which is being classified. These maps reveal more details about soil types than polygon maps because they are made at pixel size spatial resolution. According to Zhu et al. (1996), the general shapes on the membership maps follow the landscape better than the ones on the soil polygon maps where inclusion or exclusion from a region is based more on restrictions derived from the scale of the map than on local conditions. The central concept of the soil type responds to local variations in the apparent soil forming environment (represented by DTMs or terrain attributes). Fuzzy membership maps can be viewed as a non-linear transformation of the environmental variables (DTMs) (Zhu et al., 2010) and can be used to portray the uncertainty associated with the hardened or polygon map (McKay et al., 2010).

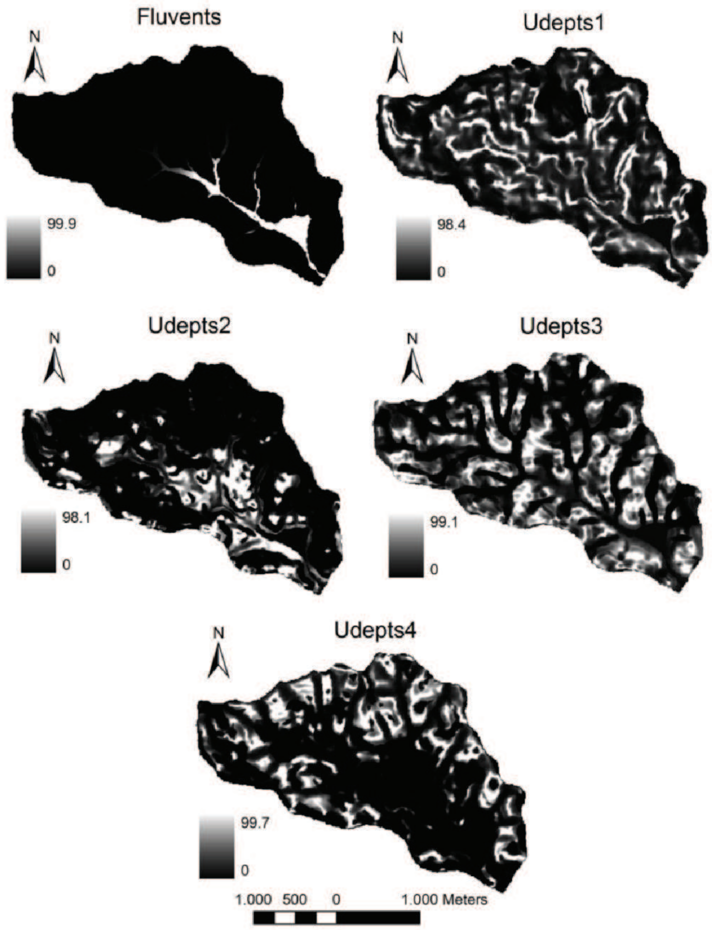

Figure 5. Fuzzy membership maps for each soil type described at Lavrinha Creek Watershed.

4.3.3 Solum depth predictive maps

Figure 6 shows the solum depth prediction map from knowledge-based digital soil mapping. The shallowest sola display hydromorphic features, occurring under low elevation regions, with gentle slopes, higher wetness index and concave landforms (Figure 3), where the water table is closer to surface in most part of the year. Also, those soils do not contain B horizon because of the frequent sediments deposition due to floods, which prevent the soil development, limiting the solum depth to the A horizon thickness only at those places. On the other hand, Udepts are formed under different landforms, slopes, and are not subjected to floods, which, in turn, allow the development of a B horizon and, hence, the solum depth. The moderately deep solum areas (yellow and light blue on the map), related to Udepts, corresponds to places on steep slopes (Figure 3) and they are thicker than the ones from lowland that are poorly developed. The deepest Udepts are related to places with gentle slopes and intermediate wetness index (high lands). Such conditions allow the soil development with current characteristics that may reduce erosion rates and provide higher water infiltration, thus enhancing the pedogenesis development rates. Also, those areas tend to receive soil eroded (colluvium) from upper lands which further contributes to their increased thickness. As reported by Menezes et al. (2009), the only detailed soil survey report at the Mantiqueira Mountain region, the Udepts around the Lavrinha Creek Watershed, under the same climate and parent material, are deeper than the ones found in the rest of the Alto Rio Grande Basin, that are influenced by the faster weathering of gneiss and the intense precipitation regime.

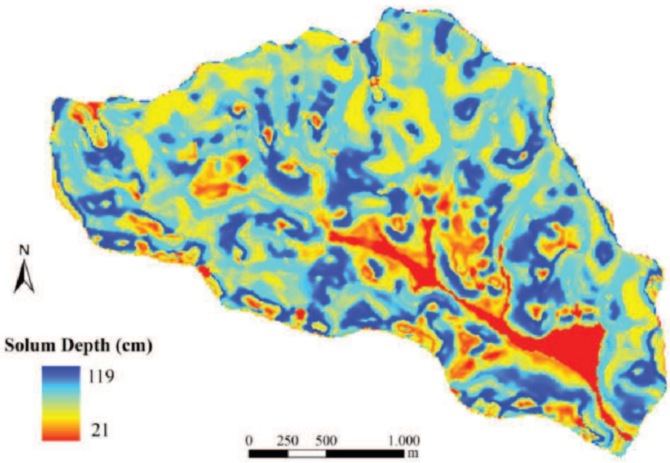

Figure 6. Solum depth prediction from knowledge-based digital soil mapping and the field validation points.

4.3.4 Validation and accuracy assessment of the predicted solum depth

The scatterplot graphics to compare the accuracy of conventional and knowledge-based digital soil mapping approaches are shown in Figure 7, and the results of the comparison parameters RMSE, ME, R^2 and R^2_{adj} for the knowledge-based and conventional solum depth map are presented in Table 3. The scatterplot graphic for conventional solum depth map (7A), shows a greater spread of data points compared to the knowledge-based graphic (7B), which means a greater discrepancy between predicted and real solum depths. It is also apparent through the low R^2 and R^2_{adj}. RMSE and ME values, which indicate that the knowledge-based solum depth map is more accurate compared to the solum depth map derived from the conventional soil polygon map (Table 3).

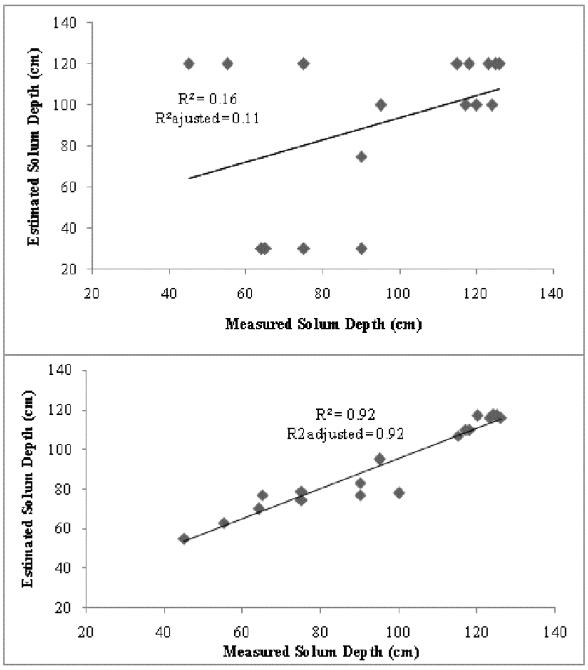

Figure 7. Scatterplot of measured vs. predicted solum depth, R^2 and R^2 adjusted from conventional (A) and knowledge based digital soil mapping approach (B).

Table 3. Comparison parameters between conventional and knowledge-based mapping to predict solum depth.

Parameters	Conventional	Knowledge-based
RMSE	35.56	9.12
ME	-3.94	-3.29
R^2	0.16	0.92
R^2_{adj}	0.11	0.92

RMSE: root mean square error; ME: mean error.

The knowledge based digital soil mapping showed the gradual changing of solum depth through the landscape, which is more realistic and resulted in greater spatial detail and accuracy when compared to the conventional map.

Also, the knowledge-based solum depth map provides information about the smaller but potentially important environmental niches that may be described by higher resolution DTMs (Zhu, 1997).

The knowledge-based digital soil mapping approach has been successfully applied in the prediction of A horizon depth (Zhu et al., 1997), drainage classes (McKay et al., 2010), A horizon silt and sand contents (Qi et al., 2006), soil transmissivity (Zhu et al., 1997), hydraulic conductivity (Zhu and McKay, 2001), and solum depth (Quinn et al., 2005; Zhu and McKay, 2001). While the information about surface topography can nowadays be derived from easily accessible DEMs in different spatial resolutions and accuracies (Hengl and MacMillan, 2009), aerial photography interpretation is becoming harder to be used due to the limited number of Pedologists trained with this methodology and the difficulty of acquiring aerial photographs at adequate scales in comparison to widely available high resolution satellite images. Furthermore, the use of digital soil mapping approach employed in this study provided adequate estimates of the solum thickness distribution at LCW. It reinforces the need of associating knowledge of soil experts and soil-landscape relationships to predict soil properties along the landscape, especially in areas with limited data availability (Menezes et al., 2013).

4.4. CONCLUSIONS

The knowledge-based digital soil mapping approach showed the advantages over the conventional soil mapping approach by applying the field expert-knowledge in order to enhance the quality of final results, predicting solum depth with suited accuracy in a continuous way, making the soil-landscape relationship explicit. A low density of samples was used, which is suited to the low financial resources for soil survey programs in Brazil.

The Mantiqueira Range region, where LCW is located, plays an important role on water production. For being a headwater watershed, knowing the solum depth may aid the decision makers about the most adequate soil management for each segment of the landscape. Also, since the soil acts as a filter of particles to which it bounds, the thicker the soil, the greater the travel pathway of contaminants to the water table.

The use of digital elevation models to derive terrain attributes and the possibility to use them to predict soil attributes using fuzzy logic provide adequate results for study areas with various soil types and difficult to access.

REFERENCES

Beven, K.; Wood, E. F. 1983. Catchment geomorphology and the dynamics of runoff contributing areas. Journal of Hydrology 65: 139-158.

Böhner, J.; McCloy,K.R.; Strobl, J. 2006. SAGA - Analysis and Modeling Applications. Göttinger Geographische Abhandlungen, v.115, 130p.

Chartin, C.; Bourennane, H.; Salvador-Blanes, S.; Hinschberger, F.; Macaire, J.J. 2011. Classification and mapping of anthropogenic landforms on cultivated hillslopes using DEMs and soil thickness data - example from the SW Parisian Basin, France. Geomorphology 135: 8–20.

ESRI 2010. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute.

Follain, S.; Walter, C.; Legout, A.; Lemercier, B.; Dutin, G. 2007. Induced effect of hedgerow networks on soil organic carbon storage within an agricultural landscape. Geoderma, 142: 80–95.

Hengl, T.; Macmillan, R. A. 2009. Geomorphometry: a key to landscape mapping and modelling. p. 433-460. *In*: Hengl, T.; Reuter, H. I., eds. Geomorphometry: concepts, software, applications. Developments in Soil Science 33, Elsevier, Amsterdam, Holand.

Hudson, B. D. The soil survey as a paradigm-based science. 1992. Soil Science Society of America Journal 56: 836–841.

IBGE - Instituto Brasileiro de Geografia e Estatística. Carta do Brasil. Rio de Janeiro, 1973. Escala: 1:50000. = Brazilian Institute of Geography and Statistics. Letter of Brazil. Rio de Janeiro. Scale: 1:50000.

Jenny, H. 1941. Factors of soil formation. McGraw-Hill, New York, USA.

MacMillan, R.A.; Moon, D.E.; Coupe, R.A. 2007. Automated predictive ecological mapping in a Forest Region of B.C., Canada, 2001–2005. Geoderma 140: 353–373.

McBratney, A.B.; Santos, M.L.M.; Minasny, B. 2003. On digital soil mapping. Geoderma 117: 3-52.

McKay, J.; Grunwald, S.; Shi, X.; Long, R.F. 2010. Evaluation of the transferability of a knowledge-based soil-landscape model. p. 165-177. *In*: Boettinger, J. L.; Howell, D.W.; Moore, A.C.; Hartemink, A.E.; Kienast-Brown, S., eds., Digital soil mapping: bridging research, environmental application, and operation. Springer, London, United Kingdom.

Menezes, M. D.; Junqueira Junior, J.A.; Mello, C.R.; Silva, A.M.; Curi, N.; Marques, J.J. Hydrological dynamics of two springs, associated to land use, soil characteristics and physical-hydrological attributes at Lavrinha creek watershed

– Mantiqueira Mountains (MG). 2009. Scientia Forestalis 37: 175-184. (in Portuguese, with abstract in English).

Menezes, M.D.; Silva, S.H.G.; Owens, P.R.; Curi, N. 2013. Digital soil mapping approach based on fuzzy logic and field expert knowledge. Ciência e Agrotecnologia, 37: 287-29.

Moore, I. D.; Gessler, P. E.; Nielsen, G. A.; Peterson, G. A. 1993. Soil attribute prediction using terrain analysis. Soil Science Society of American Journal 57: 443-452.

Odeh, I. O. A.; Mcbratney, A. B.; Chittleborough, D. 1995. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. Geoderma 67: 215-226.

Qi, F.; Zhu, A.X.; Harrower, M.; Burt, J.E. 2006. Fuzzy soil mapping based on prototype category theory. Geoderma 136: 774–787.

Quinn, T.; Zhu, A.X.; Burt, J.E. Effects of detailed soil spatial information on watershed modeling across different model scales. 2005. International Journal of Applied Earth Observation and Geoinformation 7: 324–338.

Shi, R.; Long, R.; Dekett, R.; Phillip, R. 2009. Integrating different types of knowledge for digital soil mapping. Soil Science Society of America Journal 73: 1682-1692.

Soil Survey Staff. 1999. Soil Taxonomy. A basic system of soil classification for making and interpreting soil surveys. 2 ed. Handbook 436. USDA-SCS, Washington, DC, USA.

Sumfleth, K.; Duttmann, R. 2008. Prediction of soil property distribution in paddy soil landscape using terrain data and satellite information as indicators. Ecological Indicators 8: 485-501.

Zhu, A.X. A similarity model for representing soil spatial information. 1997. Geoderma 77: 217–242.

Zhu, A. X.; Band, L. E. 1994. A knowledge-based approach to data integration for soil mapping. Canadian Journal of Remote Sensing 20: 408-418.

Zhu, Q.; Lin, H. S. 2010. Comparing ordinary kriging and regression kriging for soil properties in contrasting landscapes. Pedosphere, 20: 594-606.

Zhu, A. X.; McKay, D. S. 2001. Effects of spatial detail of soil information on watershed modeling. Journal of Hydrology 248: 54-77.

Zhu, A.X., Band, L.E., Dutton, B., Nimlos, T.J., 1996. Automated soil inference under fuzzy logic. Ecological Modelling 90: 123-145.

Zhu, A. X.; Band, L.; Vertessy, R.; Dutton, B. 1997. Derivation of soil properties using a soil land inference model (SoLIM). Soil Science Society of American Journal 61: 523-533.

Zhu, A.X., Hudson, B., Burt, J., Lubich, K., Simonson, D. 2001. Soil mapping using GIS, expert knowledge, and fuzzy logic. Soil Science Society of American Journal 65: 1463–1472.

Zhu, A.X.; Qi, F.; Moore, A. Burt, J.E. 2010. Prediction of soil properties using fuzzy membership values. Geoderma 158: 199–206.