DEVELOPMENT OF PREDICTIVE MAPPING TECHNIQUES FOR SOIL SURVEY AND SALINITY MAPPING

ABDELHAMID A. ELNAGGAR

Ph.D.

2007

AN ABSTACT OF THE DISSERTATION OF

Abdelhamid A. Elnaggar for the degree of Doctor of Philosophy in Soil Science
Presented on June 5, 2007
Title: Development of Predictive Mapping Techniques for Soil Survey and Salinity Mapping

Abstract approved:

Jay S. Noller

Conventional soil maps represent a valuable source of information about soil characteristics, however they are subjective, very expensive, and time-consuming to prepare. Also, they do not include explicit information about the conceptual mental model used in developing them nor information about their accuracy, in addition to the error associated with them.

Decision tree analysis (DTA) was successfully used in retrieving the expert knowledge embedded in old soil survey data. This knowledge was efficiently used in developing predictive soil maps for the study areas in Benton and Malheur Counties, Oregon and accessing their consistency. A retrieved soil-landscape model from a reference area in Harney County was extrapolated to develop a preliminary soil map for the neighboring unmapped part of Malheur County. The developed map had a low prediction accuracy and only a few soil map units (SMUs) were predicted with significant accuracy, mostly those shallow SMUs that have either a lithic contact with the bedrock or developed on a duripan. On the other hand, the developed soil map based on field data was predicted with very high accuracy (overall was about 97%).

Salt-affected areas of the Malheur County study area are indicated by their high spectral reflectance and they are easily discriminated from the remote sensing data. However, remote sensing data fails to distinguish between the different classes of soil salinity. Using the DTA method, five classes of soil salinity were successfully predicted with an overall accuracy of about 99%. Moreover, the calculated area of salt-affected soil was overestimated when mapped using remote sensing data compared to that predicted by using DTA. Hence, DTA could be a very helpful approach in developing soil survey and soil salinity maps in more objective, effective, less-expensive and quicker ways based on field data.

Development of Predictive Mapping Techniques for Soil Survey and Salinity Mapping

by

Abdelhamid A. Elnaggar

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented June 5, 2007

Commencement June 2008

Doctor of Philosophy dissertation of Abdelhamid A. Elnaggar presented on June 5, 2007

APPROVED:

Major Professor, representing Soil Science

Head of the Department of Crop and Soil Science

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Abdelhamid A. Elnaggar, Author

ACKNOWLEDGEMENTS

First, great thanks to my almighty God for helping and guiding me to accomplish this work. I would like to take this opportunity to thank many people for their supported and contributed toward the completion of this dissertation. First, I would like to express my sincere appreciation to my major advisor, Dr. Jay Noller, for his continuous guidance and support throughout my program of study and this research work. My sincere gratitude goes to my committee members Dr. Herbert Huddleston, Dr. John Baham, Dr. Jon Kimerling and Dr. James Thompson for their precious time and technical support during my graduate program. I would like to thank Dr. Anne Nolin for her time and her help with the analysis of remote sensing data. My sincere appreciation is to the NRCS and BLM folks, Mark Keller, Alina Rice, and Charlie Tackman, for their technical and logistical support with the field work in Malheur County. Thanks to Matthew Fillmore for his support with Benton County data.

My gratitude goes to my colleagues Sheila Slevin and Zomenia Zomeni for their comments and their help with the data analysis. My appreciation goes to Cameron Bergen, Sara Hash, Elizabeth Cervelli and Nathan Goodson for their help with collecting field data. Special appreciation goes to Dr. Fred Kizito and his kind family for their nice friendship and support. Many thanks to the department administrative staff, Peggy Mullet, Tracy Mitzel, and Jayne Smith for their great support. My appreciation goes to Joan Sandeno for proof-editing this work. I have lost three members of my family while I am here in the USA, my father, mother, and my sister. I dedicate this thesis to their souls for their great sacrifice. I would like to thank my dear wife, Shayma, and my kids, Ahmed, Wafa, and Amr for their sacrifice, patience, and great love.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

Figure Page

LIST OF TABLES

LIST OF TABLES (Continued)

Table Page

LIST OF APPENDIX TABLES

Development of Predictive Mapping Techniques for Soil Survey and Salinity Mapping

CHAPTER 1

General Introduction

Conventional soil maps represent a valuable source of information about soil characteristics and are the most popular form of soil inventory. Soil survey maps are developed in the conventional way in three steps according to Wilding (1985) and Cook et al. (1996). First, available data (aerial photographs, geology, vegetation, etc.) is studied and soil-profile properties are described at visited locations. Second, a conceptual soil-landscape model is developed based on the interpretation of the collected field data to infer soil spatial variations. Third, the conceptual model is applied to the survey area to predict the spatial distribution of soils at unvisited locations. This process is labor-intensive, expensive, time-consuming, and sometimes impractical in inaccessible areas. Almost less than 0.001% of a typical soil survey area is actually observed due to the high cost of a typical soil process (Burrough et al., 1971). Producing a soil map in the conventional way takes several years to be compiled and published, and most efforts required teams of individuals 5 to 15 years to complete. As a result, many areas world-wide still do not have a soil survey map which represents the core foundation for landscape management and sustainability.

Conventional soil mapping (CSM) techniques have been criticized in the scientific literature for being subjective and qualitative in character, where soil maps are

developed based on a mental model developed by the soil surveyors. Unfortunately, most of the knowledge involved in creating that model is undocumented and unavailable (Hudson, 1992). Also, soil survey maps do not have information about their accuracy (Burrough et al., 1971; Hudson, 1992). Spatial patterns among soil map units have been captured and displayed as a dasymetric or area-class maps with discrete boundary lines between them. Two problems follow from this approach (Burrough, 1986): First, the lines drawn on the soil survey maps may not accurately depict the boundaries between map units, where the boundaries between soil map units are often diffused, not sharp (Mark and Csillag, 1989) leading to locational errors. Second, the inferred homogeneities do not exist for many physical and chemical attributes that affect environmental modeling and soil management. In response to these criticisms, new mapping techniques have been developed called predictive or digital soil mapping.

Predictive or digital soil mapping (PSM or DSM) techniques can be defined as the development of a numerical or statistical model of the relationship among environmental variables and soil properties, which is then applied to a geographic data base to create a predictive map (Franklin, 1995; McBratney et al., 2000). Also, a new field has been added to soil science called Pedometrics, (Webster, 1994) defined as the application of mathematical and statistical methods for the study of soils. The main goal of PSM techniques is to utilize the wealth of available data, which have been collected for decades, in studying the spatial distribution of soils in more objective, effective, less-expensive, and faster ways. Much of the driving force behind this is to support resource managers and decision makers with the critical information required for soil management and sustainability.

Technological advances of the last few decades have created a tremendous potential for improvement in the way that soil maps are produced (McKenzie et al., 2000). PSM techniques have been developed as a result of the remarkable development in computational power, data acquisition technology (e.g., remote sensing, spectroradiometers, etc.), digital elevation models (DEMs), spatial statistics (e.g., geostatistics, neural network, fuzzy logic, classification and regression trees), and geographic information system (GIS). All of these factors result in a significant development in the methods used in producing soil and natural resource maps.

Remote-sensing data represent an integral part of digital soil mapping. They provide valuable information about the soil physical (e.g., particle size distribution and surface roughness), chemical (e.g., salt content, clay content, organic matter content, iron-oxide content, and surface mineralogy), and biological (e.g., vegetation types, density, and strength) properties in a spatially continuous manner across the landscape. Remote-sensing data have been intensively used in environmental research in the last few decades. Recently, great improvements have been made in remote sensors and their spatial and spectral resolutions. Satellite images are now available at a spatial resolution of less than one meter pixel size (i.e., IKONOS and QuickBird), which provides fine-scale digital representation of the earth's surface. Hyperspectral remote sensors (i.e., AVIRIS) acquire detailed information about vegetation and soil properties, which has the potential of significantly improving data input in predictive soil and natural resource mapping models. Furthermore, many recent studies also have been carried out in Australia on the possibility of using certain remote sensors (e.g., gamma radiometrics and electromagnetic sensors) in studying subsurface soil properties (IAEA, 1991).

Remote sensing data have been successfully used in mapping vegetation cover. However, their use in mapping soils and their properties was limited to areas that have low or sparse vegetation, such as in arid and semi-arid environments. The complex illumination structure caused by terrain, cloud interference and atmospheric attenuation, and/or reflectance from vegetation made it difficult to directly use images in the visible and infrared parts of spectrum to map soils in all parts of the study area (Dobos et al., 2006). The first and longest application of remote sensing in mapping soil units was through aerial image-interpretation and image processing (ca.1930 onwards). Remote sensing data also have successfully been used in mapping soil salinity for decades (Singh et al., 1977; Manchanda, 1984; Sharma and Bhargawa, 1988; Csillag et al., 1993; Joshi and Sahai, 1993; Moreau, 1996; Khan et al., 2001; Spies and Woodgate, 2005; Sethi et al., 2006).

Active remote sensors that operate in the microwave portion of electromagnetic spectrum also represent a valuable source of information in PSM techniques. They work under all conditions, such as poor visibility due to cloud covers or dust storms. Also, they work at any time of day or night, providing a significant advantage over the passive remote sensors that operate in solar-illuminated areas using the visible and near-infrared portions of spectrum. Synthetic aperture radar (SAR) is an example of active remote sensor. It has been used in producing high-resolution DEMs and mapping soil salinity and soil moisture content (Metternicht, 1998; Narayanan and Hirsave, 2001). Light detection and ranging (LiDAR) is another system similar to active remote sensing systems, but it uses pulses of laser light rather than microwave energy to illuminate the surface. LiDAR has revolutionized the survey and mapping world (Roy et al., 1993).

LiDAR data have been used in producing a highly accurate, very fine resolution DEM. Also, LiDAR data have been used in mapping watersheds, coastal zones, flooding risks, forestry, and geological hazards (Mosaic Mapping Systems Inc., 2001; Haugerud et. al., 2003).

Development of soils is controlled by the way in which water moves through and over the landscape. This water movement is largely controlled by surface topography. DEMs provide valuable information about terrain attributes (i.e., elevation, slope, aspect, and surface curvature), which have a significant impact on soil-forming factors and processes (McKenzie et al., 2000). DEMs and terrain attributes derived from them represent the most common set of variables used in predictive soil mapping (Moore et al., 1993; Gessler et al., 1995; Skidmore et al., 1996; Scull et al., 2005).

Spatial statistical methods have significantly developed over the past few decades, moving from geostatistics (i.e., Kriging and CoKriging) to more sophisticated methods such as fuzzy logic and decision and regression tree analysis. Each of these advances in geostatistics has successfully been applied to PSM. Geostatistical analyses were initially used in soil science for the purpose of spatial interpolation of soil properties from intensive soil observations collected over small areas. A long applied technique, ordinary kriging, has been used in soil mapping (Odeh et al., 1992; Burrough et al., 1992) and salinity mapping (Bourgault et al., 1997). Major limitations to the application of kriging are: 1) the assumption that stationarity in data must be met by the field-sampled data sets, requiring great amounts of data to define the spatial autocorrelation and 2) the necessity for simple, flat terrain as only two dimensions [x, y] not [z], are statistically supported. Univariate geostatistics (e.g., ordinary and universal

Kriging) have been modified to accept secondary data (i.e., Cokriging). Cokriging is the multivariate extension of kriging, where available secondary data such as terrain attributes can be included in the prediction (Odeh et al., 1995). Still, the complexity of soil-forming and surficial processes and large survey areas obviate these statistical techniques.

Neural network is a recently developed technique that has been used in PSM. These techniques attempt to build a mathematical model that hypothetically works in an analogous way to the human brain (McBratney et al., 2000). Predicting continuous functions, such as soil hydraulic properties, represents the most common application of neural networks in DSM (e.g., Minasny and McBratney, 2002). Neural networks also were in predicting the probability of soil map classes from soil environmental variables (Zhu, 2000).

The concept of fuzzy logic was first introduced by Lotfi Zadeh in 1965. Fuzzy logic represents an alternative to Boolean logic. It provides an alternative conceptual paradigm within PSM research. The use of this theory has increased greatly in the last few years, making it an important component of PSM. It allows a partial class membership to a variable in contrast to traditional crisp or binary logic. According to crisp logic, a soil sample is either completely type A or it is not at all type A. On the other hand, fuzzy logic provides membership values ranging from zero-ital nonmembership to one-ital total membership within predictive soil models to express degrees of similarity. Fuzzy logic is useful in soil mapping because of the contiguous and complex nature of the soil across the landscape (Zhu et al., 2001; Sunila1, 2004). Many recent studies have been published on the application fuzzy logic to infer the

membership of a soil to particular classes on the basis of environmental variables, such as parent material, elevation, aspect, gradient, profile curvature and canopy coverage (Zhu et al. 1996; Burrough et al., 1997; De Gruijter et al., 1997; McBratney and Odeh, 1997; Zhu, 1997; Bui et al., 1999; Zhou et al., 2004; Scull et al., 2005).

Decision and regression trees have also been widely used in producing predictive soil maps. Prediction of soil classes (discrete or categorical values) is called a decision tree or a classification tree, whereas prediction of continuous soil attributes is called a regression tree. Classification and regression trees are machine or inductive learning methods where a set of automatically constructed rules are built up based on training dataset (i.e., data mining). A constructed decision tree consists of nodes (each representing an attribute), branches (each representing the attribute value), and leaves (each representing a class). A decision tree is built based on selecting the attribute that minimizes the amount of disorder in the sub-tree rooted at a given node. A training dataset is used to discover or exploit the unknown relationships between the predictor variables and the predicted variable. The theory behind this approach is based on the assumption that all the required information to establish soil predictions is contained in the data and can be extracted if a sufficient amount of training data can be collected (Dobos et al., 2006).

Decision and regression trees have many advantages over linear models and other PSM methods. Regression trees have the ability to address the nonlinear relationships between some soil properties, require no prior assumptions about the data, and they can use both categorical and continuous data for prediction of discrete soil classes or continuous soil attributes. One of the great advantages of tree models is that

they are easy to interpret when compared to methods like generalized linear models (GLMs), generalized additive models (GAMs), and neural networks (Clark and Pregibon, 1992). Prediction rules developed by decision and regression trees can be extrapolated to map soil properties or classes in similar landscapes (Venables and Ripley, 1994; Hansen et al., 1996; Huang et al., 2002). Because of these factors, decision and regression trees have been widely used in developing predictive soil maps over large areas (Hansen el al., 1996; Bui et al., 1999; Zhou el al., 2004; Scull et al., 2005). Also, they have been used to develop landcover and other natural resource maps (Friedl and Brodley, 1997; Friedl et al., 1999; Xian et al., 2002; Herold et al., 2003).

Geographic information systems (GIS) are computer-based systems for collecting, storing, analyzing and managing data and associated attributes which are spatially referenced to the Earth. Advances in GIS provide great assistance to the techniques used in producing maps. Data are represented in GIS using two data models: vector and raster. Vector data model is used to represent discrete features that can be identified using sharp boundaries such as state counties. Conversely, the raster data model is used to represent continuous numeric values (e.g., elevation, slope, and aspect) and continuous categories (e.g., vegetation types). Raster model also known as the field-view representation, is suitable for representing continuous spatial variations in soil properties across the landscape (Goodchild, 1992). It facilitates the integration of many environmental data and the sampling of data layers. Also, it is suitable for spatial analysis and modeling (Burrough and McDonnell, 1998).

Resolution of data stored in raster format depends on the grid cell size, where the smaller the grid size the higher the resolution and vice versa. Grid size or resolution

can be changed based on the details that need to be acquired. That is why the raster data model is most commonly used in PSM and other environmental resource mapping. Although vector data can be converted into a raster grid, they still inherit some of their original properties after conversion, such as the discrete boundaries. An example of this is the digitized version of a paper soil map. PSM techniques result in significant changes in soil data modeling from the dasymetric or area-class map, because the raster grid allows a better representation of the spatial variability in soil properties across the landscape.

Performance of the prediction methods is usually tested based on the root mean square error (RMSE) (McBratney et al., 2000). Descriptive and discrete-multivariate statistics described by Jensen (1996) are also used to assess the agreement between the existing and predicted soil classes or properties. Descriptive statistics include the calculations of the Overall, the Producer's and the User's accuracies in the error data matrix. Overall accuracy is calculated by dividing the total number of correctly predicted pixels by the total number of pixels in the error matrix. Producer accuracy, a measure of omission or exclusion errors, shows how successful the model is in prediction. It is calculated by dividing the total number of correctly predicted pixels of an individual category by the total number of pixels given to that category from the reference data. User accuracy, a measure of commission or inclusion errors, shows how well these map predictions are represented in reality. It is calculated by dividing the total number of correctly predicted pixels of a category by the total number of pixels that were actually classified in that category. Discrete-multivariate statistics, such as Kappa analysis (Cohen, 1960) are also very commonly used in evaluating predictive

and image classification maps (Jensen, 1996; Mather, 2004). Kappa coefficient is used to measure the accuracy or the agreement between the predicted and present categories in the reference data (Congalton, 1991).

Although several approaches have been used in PSM, continuous testing and developing of these methods over a wider variety of landscapes is very important, especially where the spatial distribution of soils is more complex (Scull et al., 2003). Some methods could be successful under certain environments, whereas others could fail (Scull et al., 2003). Also, the impact of the environmental variables that have been used as predictors of soil properties could change from one environment to another (e.g., rainy forests and dry deserts). Therefore, future studies should be continued to determine which methods and environmental attributes provide optimal results, and understand the appropriate environments for application.

The goal of this dissertation is to:

- retrieve soil-landscape model from old soil survey data using data mining techniques;
- evaluate environmental variables according to their significance in predicting soil map units in different landscapes;
- evaluate the consistency or uncertainty of conventional soil maps and facilitate their transition and update processes by use of predictive soil mapping;
- extrapolate models retrieved from published soil survey data to produce predictive soil maps for areas with similar landscapes;

- embed soil-surveyor knowledge in data-mining models to produce instant predictive soil maps and direct the soil survey plan based on mapping accuracy;
- compare the efficiency of conventional and predictive mapping techniques in producing soil salinity maps over large areas.

These objectives are studied in detail in the next three chapters of this dissertation. The three subject chapters are intended for publication in scientific journals. A summary chapter closes the dissertation.

References

Bourgault, G., Journel, A. G., Rhoades, J. D., Corwin, D. L., Lesch, S. M. 1997. Geostatistical analysis of a soil data set. Adv. Agron., 58:241-292.

Bui, E. N., Loughhead, A., Corner, R. 1999. Extracting soil-landscape rules from previous soil surveys. Aust. J. Soil Res., 37:495-508.

Burrough, P.A. 1986. Principles of geographical information systems for land resources assessment. New York: Oxford Univ. Press, p. 193.

Burrough, P. A. 1997. Environmental modeling with geographic information systems. *In:* Innovations in GIS 4, Kemp, Z. (ed) London: Taylor & Francis, pp. 143-153.

Burrough, P. A., Beckett, P. H. T., Jarvis, M. G. 1971. The relation between cost and utility in soil survey. J. Soil Sci., 22:368-381.

Burrough, P. A., MacMillian, R. A., van Deusen, W. 1992. Fuzzy classification methods for determining land suitability from soil profile observations and topography. J. Soil Sci., 43:193-210.

Burrough, P. A., McDonnell, R. A. 1998. Principles of geographical information systems. Oxford Univ. Press, Oxford.

Clark, L. A., Pregibon, D. 1992. Tree-based models. *In*: Chambers, J. M., Hastie, T. J. (Eds.), Statistical Models. S. Wadsworth and Brooks, California, USA, p. 377-420.

Cohen, J. 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20:37-46.

Congalton, R. G. 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of the Environment. 37:35-46.

Cook, S. E., Corner, R. J., Groves, P. R., Grealish, G. J. 1996. Use of gamma radiometric data for soil mapping. Aust. J. Soil Res., 34:183-194.

Csillag, F., Pasztor, L., Biehl, L. L. 1993. Spectral band selection for the characterization of salinity status of soils. Remote Sens. Environ., 43:231-242.

De Gruijter, J. J., Walvoort, D. J. J., van Gaans, P. F. M. 1997. Continuous soil maps - a fuzzy set approach to bridge the gap between aggregation levels of process and distribution models. Geoderma, 77:169-195.

Dobos, E., Carre, F., Hengl, T., Reuter, H. I., Toth, G. 2006. Digital soil mapping as a support to prediction of functional map. Digital Soil Mapping Working Group of the European Bureau Network. University of Miskolc, Miskolc-Egyetemváros, Hungary.

Franklin, J. 1995: Predictive vegetation mapping: geographic modeling of biospatial patterns in relation to environmental gradients. Progress in Physical Geography, 19:474–490.

Friedl, M. A., Brodley, C. E. 1997. Decision tree classification of land cover from remotely sensed data. Remote Sensing of the Environment. 61:399-409.

Friedl, M. A., Brodley, C. E., Strahler, A. H. 1999. Maximizing land cover classification accuracies produced by decision trees at continental to global scales. IEEE Transactions on Geoscience and Remote Sensing. 37(2):969-977.

Gessler, P. E., Moore, I. D., McKenzie, N. J., Ryan, P. J. 1995. Soil landscape modelling and spatial prediction of soil attributes. Int. J. Geogr. Info. Syst., 9:421–432.

Goodchild, M. F. 1992. Geographical data modeling. Computers and Geosciences, 18:401-408.

Hansen, M., Dubayah, R., DeFries, R. 1996. Decision trees: an alternative to traditional land cover classifiers. Int. J. Remote Sensing, 17(5):1075-1081.

Haugerud, R. A., Harding, D. J., Johnson, S. Y., Harless, J. L., Weaver, C. S., Sherrod, B. L. 2003. High - resolution LiDAR topography of the Puget Lowland, Washington-A bonanza for Earth Science. GSA Today, 13(6):4-10.

Herold, N. D., Koeln, G., Cunnigham, D. 2003. Mapping impervious surfaces and forest canopy using classification and regression tree (CART) analysis. ASPRS 2003 Annual Conference Proceedings. Anchorage, Alaska.

Huang, C., Davis, L. S., Townshend, J. R. G. 2002. An assessment of support vector machines for land cover classification. Int. J. Remote Sensing, 23(4): 725-749.

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

International Atomic Energy Agency (IAEA) 1991. Airborne gamma-ray spectrometer surveying. Technical Series 323, IAEA.

Jensen, J. R. 1996. Introductory digital image processing: A remote sensing perspective. 2nd Ed., Prentice Hall, Inc., Upper Saddle River, New Jersey, USA.

Joshi, M. D., Sahai, B. 1993. Mapping salt-affected land in Saurashtra coast using Landsat satellite data. Int. J. Remote Sensing, 14(10):1919-1029.

Khan, N. M., Rastoskuev, V. V., Shalina, E. V., Sato, Y. 2001. Mapping salt-affected soils using remote sensing indicators - a simple approach with the use of GIS-IDRSI. Paper presented at the 22nd Asian Conference of Remote Sensing. November 5-9, Singapore.

Manchanda, M. L. 1984. Use of remote sensing techniques in the study of distribution of salt-affected soils in north-west India. Indian Soc. Soil Sci., 32:701-706.

Mark, D. M., Csillag, F. 1989. The nature of boundaries on area-class maps. Cartographica, 21:65-78.

Mather, P. M. 2004. Computer processing of remotely-sensed images – an introduction. 3rd Ed., John Wiley and Sons Ltd., Chichester, England.

McBratney, A. B., Odeh, I. O. A. 1997. Applications of fuzzy sets in soil science: fuzzy logic, fuzzy measurements and fuzzy decisions. Geoderma, 77:85-113.

McBratney, A. B., Odeh, I. O .A., Bishop, T. F. A., Dunbar, M. S., Shatar, T. M. 2000. An overview of pedometric techniques for use in soil survey. Geoderma, 97:293-327.

McKenzie, N. J., Gessler, P. E., Ryan, P. J., O'Connel, D. 2000. The role of terrain analysis in soil mapping. *In:* Wilson, J.P., Gallant, J.C. (Eds.), Terrain Analysis-Principles and Applications. Wiley, New York, p. 245-265.

Metternicht, G. I. 1998. Fuzzy classification of JERS-1 SAR data: an evaluation of its performance for soil salinity mapping. Ecological Modeling, 111:61-74.

Minasny, B., McBratney, A. B. 2002. The neuro-m method for fitting neural network parametric pedotransfer functions. Soil Sci. Soc. Am. J., 66:352-361.

Moore, I. D., Gessler, P. E., Nielsen, G. A., Peterson, G. A. 1993. Soil attribute prediction using terrain analysis. Soil Sci. Soc. Am. J., 57:443-452.

Moreau, S.S. 1996. Application of remote sensing and GIS to the mapping of saline/sodic soils and evaluation of sodification risks in the province of Villarroel, Central Altiplano, Bolivia. Paper presented at the 4th International Symposium on High Mountain Remote Sensing Cartography, Karlstad - Kiruna - Troms, August 19-29.

Mosaic Mapping Systems Inc., 2001. A white paper on LiDAR mapping. Ottawa, ON, Canada. ftp://ftp-fc.sc.egov.usda.gov/NCGC/products/elevation/lidar-applications-whitepaper.pdf

Narayanan, R. M., Hirsave, P. P. 2001. Soil moisture estimation models using SIR-C SAR data: a case study in New Hampshire, USA. Remote Sens. Environ., 75:385–96.

Odeh, I. O. A., McBratney, A. B., Chittleborough, D. J., 1992. Fuzzy-c-means and kriging for mapping soil as a continuous system. Soil Sci. Soc. Am. J., 56:1848-1854.

Odeh, I. O. A., McBratney, A. B., Chittleborough, D. J. 1995. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. Geoderma, 67:215-225.

Roy, G., Vallee, G., Jean, M. 1993. Lidar-inversion technique based on total integrated backscatter calibrated curves. Applied Optics, 32(33):6754-6763.

Scull, P., Franklin, J., Chadwick, O. A. 2005. The application of decision tree analysis to soil type prediction in a desert landscape. Ecological Modeling, 181:1–15.

Scull, P., Franklin, J., Chadwick, O. A., McArthur, D. 2003. Predictive soil mapping: a review. Progress in Physical Geography, 27(2):171–197.

Sethi, M., Dasog, G. S., Lieshout, A. V., Salimath, S. B. 2006. Salinity appraisal using IRS images in Shorapur Taluka, Upper Krishna Irrigation Project, Phase I, Gulbarga District, Karnataka, India. Int. J. Remote Sensing, 27(14):2917–2926.

Sharma, R. C., Bhargawa, G. P. 1988. Landsat imagery for mapping saline soils and wetlands in north-west India. Int. J. Remote Sensing, 9:69-84.

Singh, A. N., Kristof, S. J., Baumgardner, M. F. 1977. Delineating salt-affected soils in the Ganges Plain, India by digital analysis of Landsat data. The Laboratory of Applications of Remote Sensing. Purdue University, West Lafayette, Indiana.

Skidmore, A. K., Watford, F., Luckananurug, P., Ryan, P. J. 1996. An operational GIS expert system for mapping forest soils. Photogrammetric Engineering and Remote Sensing, 62:501-511.

Spies, B., Woodgate, P. 2005. Salinity mapping methods in the Australian context. Published by the Department of the Environment and Heritage, and Agriculture, Fisheries and Forestry.

Sunila, R., Laine, E., Kremenova, O. 2004. Fuzzy model and kriging for imprecise soil polygon boundaries. Proc. 12th Int. Conf. on Geoinformatics − Geospatial Information Research: Bridging the Pacific and Atlantic, June 7-9, University of Gävle, Sweden.

Venables, W. N., Ripley, B. D. 1994. Modern applied statistics with S-PLUS. Springer-Verlag, New York.

Webster, R. 1994. The development of pedometrics. Geoderma, 62:1-15.

Wilding, L. P. 1985. Spatial variability: its documentation, accommodation and implication to soil surveys. *In:* Nielsen, D.R., Bouma, J. (Eds.), Soil Spatial Variability. Proceedings of a Workshop of ISSS and the SSSA, Les Vegas USA. Nov. 30-Dec. 1, 1984.

Xian, G., Zhu, Z., Hoppus, M., Fleming, M. 2002. Applications of decision-tree techniques to forest group and basal area mapping using satellite imagery and forest inventory data. Pecora 15/Land Satellite Information IV/ISPRS Commission I/FIEOS Conference Proceedings.

Zadeh, L. A. 1965. Fuzzy sets. Information and Control, 8:338-53.

Zhou, B., Zhang, X., Wang, R. 2004. Automated soil resources mapping based on decision tree and Bayesian predictive modeling. J. Zhejiang Univ. Sci., 5(7):782-795.

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

Zhu, A. X. 2000. Mapping soil landscape as spatial continua: the neural network approach. Water Resources Research, 36:663-677.

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

Zhu, A. X., Hudson, B., Burt, J., Lubich, K., Simonson, D. 2001. Soil mapping using GIS, expert knowledge, and fuzzy logic. Soil Sci. Soc. Am. J., 65:1463-1472.

CHAPTER 2

Assessing the Consistency of Conventional Soil Survey Data: Switching from Conventional to Digital Soil Mapping Techniques

Abdelhamid A. Elnaggar, Jay S. Noller

Prepared for submission to:

Soil Science Society of America Journal

Abstract

Conventional soil maps represent a valuable source of information about soil characteristics; however, some errors are associated with them. Common amongst errors are inaccurate boundaries, misidentified inclusions, and uncertainty with soil map units. This work aims to find a practical approach to handle these errors through using decision-tree analysis (DTA) as a predictive soil-mapping technique. In this technique the spatial relationships between known taxonomic classes and soil map units (SMUs) and their formative environmental variables are extracted and used in developing a predictive soil-landscape model. The predictive soil map differs from the original digitized area-class representation in that each soil map unit is predicted as a continuum or pixel by pixel. A comparison between the original and predicted soil taxonomic classes and soil map units was carried out to assess the original map consistency and the model efficiency. Study of a 278 km^2 region of Benton County, Oregon, shows a high correlation between predicted and original taxonomic classes and SMUs. In this study, six models were used to predict soils based on taxonomic class (soil orders, suborders, great groups, and subgroups) and soil components (major soil map units and all soil map units) in the study area. The developed digital soil maps provided valuable information about the prediction accuracy of each soil map unit and the areas where there is low confidence in predicting them. The predictive model of soil orders yields the highest prediction accuracy (90%), followed by soil suborders (85%), great groups (81%), subgroups (79%), major soil map units (78%) and all soil map units (74%), respectively. All of the predicted soil maps revealed similar characteristics between them, where the majority of taxonomic classes and/or SMUs that represent greater areal

extent have the highest prediction accuracy compared with those representing lesser extent. Results quantitatively reveal errors common to conventional soil maps. Confidence in the boundaries between taxonomic classes or SMUs had the greatest misclassification errors of all recognized elements in the mapping area. Also, the inclusion between soil map units was obvious especially among those representing smaller areas. The misclassification error was also higher among SMUs or classes at lower slope areas in the Willamette Valley. In these locations there are many delineated soil map units, whereas there is little or no significant change in terrain attributes and most of the environmental variables used in developing the prediction models.
Keywords: Accuracy assessment, predictive soil mapping, pedometrics, decision-tree analysis

2.1. Introduction

Digital soil maps are now the de facto reference with valuable information used by resource managers and decision makers. With the change to digital format and GIS from graphical techniques comes the opportunity to readily assess the reliability, accuracy, precision and meaning of a whole host of data, from inputs to processing to final products. Therefore, advancing new techniques to minimize uncertainty and accumulated errors in final map products is an important step. Conventional soil mapping techniques are very expensive, labor-intensive and time-consuming practice which are prove to a suite of errors, including incorrect labels, inclusions, and inaccurate class boundaries (Ehlschlaeger and Goodchild, 1994).

Traditionally and for the foreseeable future, spatial patterns have been captured and displayed as dasymetric maps with discrete boundary lines between soil map units, which implies homogeneity within such area-class map units. Two problems follow from this approach (Burrough, 1986): First, the lines drawn on the soil survey maps may not accurately depict the boundaries between map units. Boundaries between soil maps units are often diffused, not sharp (Mark and Csillag, 1989) leading to location errors. Second, the inferred homogeneities do not exist for many physical and chemical attributes that affect environmental modeling and soil management.

Conventional soil survey methods have been criticized for being too qualitative and subjective in character (McBratney et al., 2000; Qi and Zhu 2003; Scull et al., 2003; Prima et al., 2006), largely because soil boundaries are manually delineated on the basis of the mental soil-landscape model developed by a local soil expert. Complexity in spatial relationships between soils and their environmental variables may not be recognized because of scaling differences within the soil survey. Even in projects where there is scaling mismatch, soil surveyors may have a limited view over large areas of the spatial relationships that exist between all of the soils in their legend and the respective environmental factors.

Quantitative soil-landscape models are developed to describe, classify and analyze the spatial distribution patterns of soils using more objective, effective and less expensive means collectively called predictive or digital soil mapping techniques (McBratney et al., 2000; McBratney et al., 2003; Scull et al., 2003). Digital soil mapping (DSM) can be defined as the creation and population of spatial soil information by the use of field and laboratory observational methods coupled with

spatial and non-spatial soil inference systems (McBratney and Lagacherie, 2004). DSM, while de rigueur for soil survey in Australia and Canada (McBratney et al., 2003; MacMillan et al., 2005), is just now being institutionalized in the USA (USDA-NRCS, 2007).

Spatial variability represents a very important factor in mapping soils and other natural resources and has been subject of recent research. Study of soil variability has been approached by numerical classification, multivariate statistical methods, continuous (fuzzy) classification, geostatistics, fractal methods, mathematical morphology, and chaos theory (Burrough, 1993). Variability studies give consideration to the reliability or uncertainty associated with soil information; therefore they are an integral part of soil science. Soil information derived from soil maps should include an expression of uncertainty or variability, especially when used by soil managers and decision makers. Most soil maps developed using conventional methods have transect level and local assessments of reliability information (Soil Survey Division Staff, 1993), yet not equally over the entire study area. Soil maps developed using digital methods can have uncertainty assessment of one degree or another equally over the entire study area.

This paper reports on a method of assessing the uncertainty associated with soil maps which are produced by traditional soil survey techniques. To do so, we first used decision-tree analysis to retrieve the spatial relations between soil map units of a digitized conventional soil map and their environmental variables, and, second, we developed a predictive soil map of the study area using digital soil mapping techniques. We present our analysis of a reported technique – decision-tree (e.g. Hansen et al., 1996;

Scull et al., 2005) – and present potential uses thereof such as in updating soil surveys and in joining soil surveys for regional (e.g. MLRA) compilations and management.

2.2. Materials and methods

2.2.1. Description of the study area

Recent completion of the Benton County, Oregon, soil survey (USDA-NRCS, 2004a) affords a local, newly compiled updated soil survey map for this study. We chose a 278 km^2 area of maximum range in soil-forming factors across the county, an area that includes two 7.5 minute quadrangles called Flat Mountain and Greenberry. A soil map of the study area was clipped from the digitized (conventional) soil map of Benton County, Oregon (SSURGO database developed by USDA-NRCS, 2004). The study area encompasses 100 soil map units that represent Alfisols, Andisols, Inceptisols, Mollisols, Ultisols and Vertisols. Tables 2.1 and 2.2 show the map symbols and names, percentage of the area covered and taxonomic classification for some of the major-SMUs in the study area (refer to the Tables A.2.1 and A.2.2 for the complete list and descriptions).

The study area includes five different ecological regions: Prairie Terraces, Valley Foothills, Mid-Coastal Sedimentary, Willamette River and Tributaries Gallery Forest, and Volcanics (Clarke and Bryce, 1997) that compose this boundary area between the A2 and A4 regions of the Major Land Resource Areas (MLRA) (USDA, 2006). Dominant geology in the study area consists of alluvial deposits, lacustrine and fluvial sedimentary rocks, (tuffaceous) siltstone and sandstone, mafic intrusions, and

volcanic flows (Walker et al., 2003). Elevation varies from 63 to 831 m ($\bar{X}$=197 m) and slope ranges from 0 to 50^{o} ($\bar{X}$=7.29^{o}). Climate in the study area is humid Mediterranean, with mean annual precipitation (MAP) varying from 1092 mm on the valley floor to 2362 mm at peak elevations ($\bar{X}$=1503 mm), and Xeric to Udic (local aquic) soil moisture. Mean annual temperature (MAT) varies from 4^{o}C in the winter to 27 ^{o}C in the summer ($\bar{X}$= 17 ^{o}C), and Frigid to Mesic soil temperature.

Vegetation present in the area is: Oregon white oak *(Quercus garryana),* Douglas fir *(Pseodotsuga menziesii),* western hemlock *(Tsuga heterophylla)*, grand fir *(Abies grandis),* ponderosa pine *(Pinus ponderosa)*, Oregon ash *(Fraxinus oregona)*, and black cottonwood *(Populus trichocarpa)*. This is in addition to agricultural crops and pastures on the low-relief areas of the Willamette Valley (Kagan and Caicco, 1992).

2.2.2. Data sources and management

Environmental variables that were integrated in the soil prediction model (Table 2.3), include MAP from (1961 to 1990, 1:200,000, USDA-NRCS, 1999), geology (1: 500,000, USGS, 2003), and vegetation of Oregon (1:250,000; Kagan and Caicco, 1992) (Fig. 2.1). Terrain attributes (elevation, slope gradient, aspect, and plan and profile curvatures) were derived from the digital elevation model (DEM) of western Oregon (10 m cell size) using ArcGIS software (Fig. 2.2). Solar radiation (direct, diffuse, and globe) for the summer solstice, equinox and winter solstice was derived from the using the solar analyst ArcView extension developed by Fu and Rich (1999). Landscape position was classified by the DEM-derived Topographic Position Index (TPI) (Weiss, 2001; Jenness, 2005).

Processing of a Landsat ETM+ image (path 46 and row 29; acquired September 25, 2000) yielded the Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1973) and the Soil Adjusted Vegetation Index (SAVI) (Huete, 1988). These indices were used to identify vegetated areas and to distinguish between vegetation and soil background (Fig. 2.3).

Data layers are represented using the raster data model, which is more suitable for representing continuous spatial variations, data sampling, and data modeling. Accordingly, all data were converted to raster data of 30 m cell size.

2.2.3. Significance of the environmental variables

Three methods were used to select the most significant variable in predicting SMUs. First, Principal Component Analysis (PCA) of the Landsat ETM+ bands reduces the dimensionality of the dataset and identifies new typically meaningful underlying variables. Second, Tasseled Cap Transformation (TCT) analysis (Kauth and Thomas, 1976) derives from imagery an enhanced discrimination function between soil and vegetation. TCT provides three indices: soil brightness index (Brightness), green vegetation index (Greenness), and soil and vegetation moisture (Wetness). The first two indices (Brightness and Greenness) contain most of the scene information (95 to 98%) (Jensen, 1996). Third, winnowing attributes function in the See5 Program was used to sort environmental variables according to their significance in predicting SMUs, where decision trees and rulesets constructed by the See5 (Quinlan, 2001) do not generally use all of the input attributes. This function in the See5 program was used to pre-select a subset of the input attributes based on their significance in predicting SMUs

or groups in the study area. Only the environmental variables or attributes that provide predictive information were used in constructing the decision trees.

2.2.4. Sampling strategy

Soils are analyzed in this study as members of four groups based on taxonomic classes (soil order, suborder, great group, and subgroup) and two groups of soil map unit components, major-SMUs and all-SMUs to develop six prediction models. Soil orders group consists of 11 map classes; six soil orders, four complexes of soil orders, and one class for areas of open water. Soil suborders have 16 classes (10 individual suborders, five complexes of suborders, and one class for areas of open water). Soil great groups consist of 25 map classes (15 individual great groups, nine complexes of great groups, and one class for areas of open water). Soil subgroups have 41 map classes (25 individual subgroups, 15 complexes of subgroups, and one class for areas of open water). Major-SMUs refer to SMUs that have a combined polygon area of 0.45% or more of the study area. They include 48 SMUs and cover a total of about 91% of the study area. 52 SMUs represent a total of about 9% of the area and are included with major-SMUs for the all-SMUs (100 total SMUs) analysis. A full description of each studied group, its classes, and their given codes is provided in the Appendices.

Studied groups in the area of interest were randomly sampled based on their representative areas using the Classification and Regression Tree (CART) module designed by Earth Satellite Corporation (2003). The resulting output data matrix consists of about 90,000 random sample points as training data and about 30,000 random points as test data. Each random sample has information about the current SMU

or soil group and the different environmental variables at this point. An overview of the different sources of geodatabases and the analytical methods is illustrated in Figure 2.4.

2.2.5. Decision-tree analysis (DTA)

Decision-tree analysis is a predictive model that correlates several independent variables with direct or indirect relations to a certain subject or phenomenon, and then uses those in predicting that subject or dependent variable. Decision-tree analysis is the approach used in this study, applying the See5 program to develop prediction models (Breiman et al., 1984; Quinlan, 1993). Two advanced features in the See5 program called boosting and cross-validation were used to improve classification accuracy. With the boosting function, the program develops a sequence of decision trees; each subsequent tree attempts to fix the misclassification errors in the previous one. Each decision-tree makes a prediction and the final prediction is a weighted vote of the predictions of all trees (Freund and Schapire, 1996). This function often improves classification accuracy and reduces over-fitting of decision trees (Friedl et al., 1999; Moran and Bui, 2002). Cross-validation is designed to obtain a more reliable estimate of predictive accuracy assessment using a limited number of reference data samples for both training and accuracy assessment (Michie et al., 1994). For *f*-fold cross-validation, the training dataset is divided into *f* subsets of roughly the same size and class distribution. Predictive accuracy estimates are derived by using each subset to evaluate the classification developed by using the remaining training samples. Mean estimates represent the accuracy of the classification using all reference samples.

Prediction models for the taxonomic groups and SMUs were generated using about 75% of the random samples for training and about 25% of the samples for testing the model. The same number of the environmental variables (28 variables) was integrated in developing the six soil prediction models.

2.2.6. Models evaluation

Descriptive statistical methods (Jensen, 1996) were used to evaluate the agreement between the existing and predicted groups. In these methods, the overall accuracy is computed by dividing the total number of correctly predicted pixels in each SMU or taxonomic classes by the total number of pixels in the error matrix. Assessments for each SMU or group involved calculations of the producer's accuracy and the user's accuracy. Producer accuracy, which measures the exclusion errors, shows how successful the model is in prediction. User accuracy, a measure of inclusion errors, shows how well these map predictions are represented in reality.

In order to evaluate the accuracy of the six soil prediction models, both the actual and predicted soil maps were randomly sampled to collect about 30,000 points. As a starting assumption and condition for our analysis, the soil survey map of Benton County is deemed perfectly accurate. We recognize the improbability of this, as have other workers (Webster and Oliver, 1990; Brannon and Hajek, 2000; Rossiter, 2001) especially given the target accuracy for such SSURGO-certified maps are on the order of 75% (Soil Survey Division Staff, 1993). Henceforth, it is the *apparent* accuracy of the prediction models we report in this evaluation.

2.3. Results

2.3.1 Potential of the environmental variables

PCA on the subsetted Landsat ETM+ image revealed that the first four bands represent 99.86% of the variations within the images (Table 2.4). The first band, which covers the blue range of spectrum, has the highest eigenvalue and represents 88.66% of the image variations. On the other hand, bands 5 and 7 have the lowest eigenvalues and represent less than 0.5% of the variations within the images.

Preliminary results of the See5 winnowing option show that elevation, vegetation, geology, precipitation and slope gradient represent the most significant variables in predicting SMUs in the reference area. These variables are followed in significance by Landsat ETM+ bands 1, 3 and 4, slope aspect, landform, greenness, brightness, and NDVI. However, the remaining variables have very little contribution in predicting SMUs in the reference area. The surface curvature variables (pan and profile curvatures) did not show any significant effect on predicting SMUs or groups, which agrees with results obtained by Bui et al. (1999).

2.3.2. Model evaluation

Our model for soil orders successfully predicted all 11 classes (Fig. 2.5). It has the highest prediction accuracy compared with prediction models for other area-class groups of soils (Table 2.5). Prediction accuracy was enhanced by using a 10-fold boosting (14.2% and 11.6% of misclassification error without and with boosting, respectively) based on evaluating the test data. The prediction model of soil suborders (16 classes) came second in order of model accuracy (18.4% and 15.7% of

misclassification error without and with boosting, respectively). A comparison between the actual and predicted soil suborders is illustrated in Figure 2.6. Our model for soil great groups successfully predicts 24 groups out of 25. The unpredicted group (#27) is a complex of two soil great groups (Haploxerepts and Haploxerolls) and represents about 0.02% of the study area.

Our models for great groups, subgroups, and major-SMUs were efficiently predictive of all subject map units (Table 2.5). The predictive model of all-SMUs predicts 96 out of the 100 SMUs in the study area. The four unpredicted SMUs (51, 106, 114, and 119) represent only 0.04% of the studied area. Misclassification error of the all-SMUs model was the highest (32.8% and 28.1% without and with boosting, respectively) compared with the other models.

We noticed that the prediction accuracy was further enhanced by about 2% by using the majority filter in ArcGIS (3x3 kernel with eight neighboring cells) for all but predicted soil maps for soil suborders. In this case, the accuracy increased by less than 1% (0.62%). Filtering the generated soil maps removes scattered pixels in the soil prediction maps and improves the visualization of the output maps. The filter also clips out areas that are less than the minimum SMU size. With further field study, these outliers may be proved or disproved as inclusions or other SMU components.

Although the obtained results showed that soil orders have higher prediction accuracy than major or all-SMUs, predicting SMUs is more valuable. More information is associated with SMUs compared to soil taxonomic classes, although the latter are valuable for small-scale applications.

2.4. Discussion

2.4.1. Decision tree analysis and predicted soil map

Decision-tree analysis yields prediction models that exhibit consistent patterns regarding traditional soil survey maps and their taxonomic derivatives (e.g., distribution of soil suborders). As a general rule, based on our observations: the greater the map unit extent (area), the greater certainty or accuracy of the prediction. Ultisols and Mollisols, for example, are the dominant soils in the study area (37.7 and 26.7% of the area, respectively (Table A.2.3)) and show the highest user accuracy (98.6 and 92.6%, respectively) (Table 2.6). A map complex of Inceptisols and Mollisols has the smallest area (0.02%) and was not predicted in the filtered image. Vertisols represent 0.09% of the area and were correspondingly predicted with a low accuracy (28.6%).

Inclusion errors among suborders, which are represented by the calculated user accuracy, show the same trend as soil orders (Table 2.7). Humults is the most dominant soil suborder in the study area (37.7% of the area) and has the highest user accuracy (98.7%). Aquolls represent 0.09% of the area and have the lowest prediction accuracy (15.4%). Suborder 17 a complex of Xerepts with Xerolls, has the smallest area and was predicted with an accuracy of 40.0%. Moreover, user accuracies for soil great groups, subgroups, major-SMUs and all-SMUs showed the same trend, which led us to posit the general rule (Table 2.8).

This trend could be due to the fact that these SMUs or groups are represented in the training sample based on their area when the randomized sampling technique is used. Those units covering larger areas are represented by a larger number of sampling points than those representing smaller areas. Consequently, these larger units are well

characterized, whereas smaller units are poorly characterized in the output data matrix. Friedl and Brodley (1997) also reported that the decision-tree algorithm has a tendency to penalize classes with fewer observations in the training data. Another contributing factor could be the inconsistency in scale between the environmental data and the soil map. Geologic data, for example, have the coarsest scale (1:500,000); the soil map has a scale of (1:24,000), and the DEM and terrain attributes have a resolution of 10 m. Therefore, SMUs representing smaller areas are not well identified or discriminated from the input data. Also, the smaller units could be misclassified by the soil experts who developed the original soil map of the study area, where there are too many SMUs (e.g., 100 SMUs) in such a small area.

There are some exceptions for every rule; some SMUs or groups have higher prediction accuracies although they represent smaller areas and vice versa. This could be related to the strength of the relationships between the predicted unit and the environmental variables used in developing the prediction models. In other words, if the SMU or group has certain unique properties such as specific type of geology, vegetation or landform, it could be easily retrieved from one or more of the prediction variables. For example SMUs 109, 75, 148, 150, 174, and 119 represent small areas that comprise from 0.02 to 0.13% of the total area, they predicted with higher accuracies (100.0, 94.7, 90.5, 85.1, 62.5, and 60.0%, respectively).

There are some poorly predicted SMUs representing large areas. Soil map units 177, 8, 49, 157, 50, 154, and 29 cover relatively large areas and range between 0.4 and 3.5% but they were predicted with lower accuracy (39.4, 40.2, 44.3, 30.1, 36.2, 30.1, and 25.2%, respectively).

2.4.2. Uncertainty associated with conventional soil maps

Obtained results emphasize two of the most common issues associated with conventional soil mapping techniques: inaccurate boundaries and inclusions of soil map units. The confidence map derived from the prediction model of all-SMUs shows a relatively high misclassification error at the boundaries (Figures 2.7 and 2.8) and it also illustrates inclusions among SMUs. The error likely corresponds to imprecise traditional methods which may not accurately depict the boundaries between map units (Burrough, 1986). Zylman et al. (2005) also reported that erosional areas and transitional zones between soil map units display the greatest amount of variation in the SSURGO data. This work could help in reducing these location errors at the boundaries among SMUs by stratifying the sampling technique according to prediction confidence. We recommend that this work go parallel with the traditional field work for two reasons. First is to test the concurrent field or true accuracy. Second is to direct the sampling efforts to these locations where there are high misclassification errors to enhance their mapping accuracy. By doing this, the large number of samples required in the conventional soil survey can be reduced significantly. The same technique can be used to enhance and facilitate the update processes of soil survey data where most of the available data are outdated.

We found that the misclassification accuracies were high between the SMUs at low-slope to level areas in the Willamette Valley. This could be due to the relatively insignificant variations between most of the environmental variables used in developing the prediction models, especially terrain attributes on those areas, whereas there are many delineated SMUs. This agrees with the results obtained by Scull et al. (2005) who

found higher accuracies in predicting soil great groups in mountain areas compared to those in basin areas of the Mojave Desert, California. Misclassification error was high in areas that have a complex of more than one soil map unit or soil group in the study area. High misclassification errors also were found in deeply incised areas or drainage channels where surficial processes are active and soil development is correspondingly weak.

2.5. Conclusion

Soil-landscape models used in mapping soils can be successfully extracted using decision-tree analysis. The retrieved model can be used to assess the consistency of the original soil map. Also, it could be very helpful in transferring these digitized soil maps into more objective digital maps and facilitating the update process in soil survey. The decision-tree approach is flexible to train where environmental variables vary from one landscape to another. It can effectively minimize the large amount of field data required in conventional soil survey and consequently reduce the expenses and the time used in producing these maps.

Prediction accuracy of the developed models monotonically increases with increasing area of individual map units of area-class maps of soil taxa and soil map unit components. This trend depends on the number of classes (i.e., details) in each group, the proportional area of each class, and the accuracy and scale of contributing environmental variables. Soil orders had the highest prediction accuracy followed by soil suborders, great groups, subgroups, major and all soil map units, respectively. Map units that have a large areal extent and/or are well identified by more than one of the

environmental variables were predicted with higher accuracies. On the other hand, map units representing smaller areas or that poorly identified from the integrated variables were poorly predicted or entirely unpredicted. Selecting the appropriate model depends on the details that one wants to be retrieved from the generated soil map. Soil taxa maps would be useful in developing the preliminary soil maps at small scales (large areas) where detailed information is not necessarily required. Predicting soil subgroups and soil map units could be very valuable under larger scales (small areas) where more detailed knowledge about soil characteristics is required.

The predicted soil maps also revealed the common errors associated with the conventional soil maps: inaccurate boundaries and inclusions, especially between soil map units that cover small areas.

Acknowledgments

We thank Matthew Fillmore, NRCS scientist, who was the principal author of the map definitions we used from the Benton County soil survey. Also, we thank Joan Sandeno for editing an early version of the manuscript.

References

Brannon, G. R., Hajek, B. F. 2000. Update and recorrelation of soil surveys using GIS and statistical analysis. Soil Science Society of America Journal, 64(2):679–680.

Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J. 1984. Decision and regression trees. Wadsworth Belmont, CA.

Bui, E. N., Loughhead, A., Corner, R. 1999. Extracting soil-landscape rules from previous soil surveys. Aust. J. Soil Res., 37:495-508.

Burrough, P.A. 1986. Principles of geographical information systems for land resources assessment. New York: Oxford Univ. Press, p. 193.

Burrough, P.A. 1993. Soil variability revisited. Soils Fert., 56(5):529-562.

Clarke, S. E., Bryce, S. A. 1997. Hierarchical subdivisions of the Columbia Plateau and Blue Mountains Ecoregions, Oregon and Washington. Portland: U.S. Department of Agriculture-Forest Service General Technical Report PNW-GTR-395, p. 114.

Earth Satellite Corporation. 2003. CART Software User's Guide.

Ehlschlaeger, C. R., Goodchild, M. F. 1994. Dealing with uncertainty in categorical coverage maps: Defining, visualizing, and managing errors. In: Proceedings of the Workshop on Geographical Information Systems at the Conference on Information and Knowledge Management, Gaithersburg, Maryland: 86–91.

Freund, Y., Schapire, R. 1996. Experiments with a new boosting algorithm. *In* Machine Learning: Proceedings of the Thirteenth International Conference, July, 1996. San Mateo, California: Morgan Kaufmann.

Friedl, M. A., Brodley, C. E. 1997. Decision tree classification of land cover from remotely sensed data. Remote Sensing of the Environment. 61:399-409.

Friedl, M. A., Brodley, C. E., Strahler, A. H. 1999. Maximizing land cover classification accuracies produced by decision trees at continental to global scales. IEEE Transactions on Geoscience and Remote Sensing. 37(2):969-977.

Fu, P., Rich, P. M. 1999. The solar analyst 1.0 user manual. Helios Environmental Modeling Institute, LLC. http://www.hemisoft.com.

Hansen, M., Dubayah, R., DeFries, R. 1996. Decision trees: an alternative to traditional land cover classifiers. Int. J. Remote Sensing, 17(5):1075-1081.

Huete, A. 1988. A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment, 25:295-309.

Jenness, J. 2005. Topographic position index (tpi_jen.avx) extension for ArcView 3.x. Jenness Enterprises. http://www.jennessent.com/arcview/tpi.htm

Jensen, J. R. 1996. Introductory digital image processing: A remote sensing perspective. 2nd Ed., Prentice Hall, Inc., Upper Saddle River, New Jersey, USA.

Kagan, J., Caicco, S. 1992. Manual of Oregon actual vegetation. Idaho Cooperative Fish and Wildlife Research Unit, University of Idaho.

Kauth, R. J., Thomas, G. S. 1976. The tasseled cap: a graphic description of the spectral-temporal development of agricultural crops as seen by LANDSAT. Proceedings of the Symposium on Machine Processing of Remotely Sensed Data, Purdue University of West Lafayette, Indiana, p. 4B-41 to 4B-51.

MacMillan, R. A., Pettapiece, W. W., Brierley, J. A. 2005. An expert system for allocating soil to landforms through the application of soil survey tacit knowledge. Canadian Journal of Soil Science, 85(1):103-112.

Mark, D. M., Csillag, F. 1989. The nature of boundaries on area-class maps. Cartographica, 21:65-78.

McBratney, A. B., Lagacherie, P. 2004. Global workshop on digital soil mapping. Montpellier, France.

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

McBratney, A. B., Odeh, I. O. A., Bishop, T. F. A., Dunbar, M. S., Shatar, T. M. 2000. An overview of pedometric techniques for use in soil survey. Geoderma, 87:293-327.

Michie, D., Spiegelhalter, D. J., Taylor, C. C. 1994. Machine learning, neural and statistical classification. New York: Ellis Horwood, p. 289.

Moran, C. J., Bui, E. N. 2002. Spatial data mining for enhanced soil map modeling. Int. J. Geographical Information Science, 16(6):533-549.

Prima, O. D. A., Echigo, A., Yokoyama, R., Yoshida, T. 2006. Supervised landform classification of Northeast Honshu from DEM-derived thematic maps. Geomorphology, (78):373–386.

Qi, F., Zhu, A. X. 2003. Knowledge discovery from soil maps using inductive learning. Int. J. Geographical Information Science, 17(8):771–795.

Quinlan, J. R. 1993. C4.5: Programs for machine learning. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Quinlan, J. R. 2001. See5: An Informal Tutorial. http://www.rulequest.com.

Rossiter, D. G., 2001. Assessing thematic accuracy of area-class maps. Soil Science Division, ITC. Enschede, Netherlands.

Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W. 1973. Monitoring vegetation systems in the Great Plains with ERTS, Third ERTS Symposium, NASA SP-351, 1:309-317.

Scull, P., Franklin, J., Chadwick, O. A. 2005. The application of decision tree analysis to soil type prediction in a desert landscape. Ecological Modeling, 181:1–15.

Scull, P., Franklin, J., Chadwick, O. A., McArthur, D. 2003. Predictive soil mapping: a review. Progress in Physical Geography, 27(2):171–197.

Soil Survey Division Staff. 1993. Soil survey manual. Soil Conservation Service. U.S. Department of Agriculture Handbook 18.

USDA. 2006. Land Resource Regions and Major Land Resource Areas of the United States, the Caribbean and the Pacific Basin. Handbook 296.

USGS. 2003. Spatial digital database for geologic map of Oregon. U.S. Department of Interior, U.S. Geological Survey.

USGS-EROS Data Center. 1999. Oregon 10m DEM. U.S. Geological Survey, Sioux Falls, SD.

USGS-EROS Data center. 2000. Landsat ETM – Path: 46 Row: 29. U.S. Geological Survey and EROS Data Center, Sioux Falls, SD.

USDA-NRCS. 1999. Oregon annual precipitation. National Cartography and Geospatial Center. Fort Worth, TX.

USDA-NRCS. 2004a. Soil survey geographic (SSURGO) database for Benton County, Oregon. U.S. Department of Agriculture, Natural Resources Conservation Service, Fort Worth, Texas.

USDA-NRCS. 2007. Assessment and implementation of digital soil mapping in the national cooperative soil survey: a challenge dialogue with the soil

survey community. Workshop held January 9-11, USDA Lyng Service Center, Davis, California.

Walker, G.W., MacLeod, N.S., Miller, R.J., Raines, G.L., Connors, K.A., 2003, Spatial digital database for the geologic map of Oregon: U.S. Geological Survey, Open-File Report 03-67, ver. 2.0, 22 p.

Webster, R., Oliver, M. A., 1990. Statistical methods in soil and land resource survey. Oxford University Press, Oxford.

Weiss, A. 2001. Topographic position and landforms analysis. Poster Presentation, ESRI User Conference, San Diego, CA.

Zylman, J., Weindorf, D. C., Wittie, R., McFarland, A., Butler, T. 2005. Field-truthing of USDA-Natural Resources Conservation Service soil survey geographic data on Hunewell Ranch, Erath County, Texas. Soil Survey Horizons, 46(4):135-145.

Table 2.1. Map symbols and names of SMUs in the study area of Benton County, Oregon and their relative areas as percentage.

Map Symbol	Map Unit Name	%
1	Abiqua silty clay loam, 0 to 3 percent slopes	0.10
8	Amity silt loam, 0 to 3 percent slopes	4.24
9	Apt-McDuff complex, 5 to 30 percent slopes	1.92
10	Apt-McDuff complex, 30 to 50 percent slopes	0.29
12	Awbrig silty clay loam, 0 to 2 percent slopes	2.81
13	Bashaw clay, 3 to 12 percent slopes	0.05
17	Bellpine-Jory complex, 2 to 12 percent slopes	4.53
18	Bellpine-Jory complex, 12 to 20 percent slopes	6.83
19	Bellpine-Jory complex, 20 to 30 percent slopes	7.68
20	Bellpine-Jory complex, 30 to 60 percent slopes	10.79
21	Blachly-Kilowan complex, 5 to 30 percent slopes	0.42
22	Blachly-Kilowan complex, 30 to 60 percent slopes	0.24
23	Bohannon-Preacher complex, 30 to 60 percent slopes	3.96
24	Bohannon-Preacher complex, 60 to 90 percent slopes	1.56
27	Burntwoods-Oldblue complex, 30 to 60 percent slopes	0.81

(Refer to Table A.1.1 for a complete list of soil map units)

Table 2.2. Soil taxonomic classification of soil series in the study area.

Soil Name	Taxonomic Classification
Awbrig	Fine, smectitic, mesic Vertic Albaqualfs
Burntwoods	Medial-skeletal over loamy-skeletal, mixed over isotic, frigid Typic Fulvudands
Blachly	Fine, isotic, mesic Typic Dystrudepts
Bohannon	Fine-loamy, isotic, mesic Andic Dystrudepts
Oldblue	Fine-loamy, isotic, frigid Andic Dystrudepts
Preacher	Fine-loamy, isotic, mesic Andic Dystrudepts
Kilowan	Fine, isotic, mesic Typic Dystrudepts
Amity	Fine-silty, mixed, superactive, mesic Argiaquic Xeric Argialbolls
Apt	Fine, isotic, mesic Typic Haplohumults
McDuff	Fine, isotic, mesic Typic Haplohumults
Bellpine	Fine, mixed, active, mesic Xeric Haplohumults
Jory	Fine, mixed, active, mesic Xeric Palehumults
Bashaw	Very-fine, smectitic, mesic Xeric Endoaquerts

(Refer to Table A.1.2 for a complete taxonomic classification list of soil names)

Table 2.3. Environmental data used in developing the soil prediction model.

Variables	Data source	Resolution (Scale)	Type of Data
Landsat ETM+ (bands 1, 2, 3, 4, 5, and 7)	USGS-EROS Data Center (2000)	30 m	Continuous
NDVI	Derived from the Landsat image (Rouse et al., 1973)	30 m	Continuous
SAVI	Derived from the Landsat image (Huete, 1988)	30 m	Continuous
Tasseled Cap Transformation (Brightness, Greenness, and Wetness)	Derived from the Landsat image (Kauth and Thomas, 1976)	30 m	Continuous
Elevation (DEM) Slope gradient Slope aspect Surface curvature Profile curvature Plan curvature	USGS-EROS Data Center (1999)	30 m	Continuous
Landform classification	Derived from the DEM (Jenness, 2005)	30 m	10 classes
Solar radiation (in WH/m^2) (Diffuse, direct, and Globe radiation)	Drived from the DEM (Fu and Rich, 1999)	30 m	Continuous
Geology	USGS (2003)	1:500,000	7 classes
Vegetation	Kagan and Caicco, 1992	1:250,000	6 classes
Mean annual precipitation	USDA-NRCS (1999)	1:500,000	26 classes
Ecological regions	(Clarke and Bryce, 1997)	1:250,000	5 classes

Table 2.4. Results of principal component analysis of the Landsat ETM+ image (bands 1, 2, 3, 4, 5, and 7).

PC	Min	Max	Mean	Stdev	Eigenvalue	Percent (%)
1	47	195	63.27	11.75	3512.09	88.66
2	28	177	49.63	15.42	339.71	97.24
3	20	200	49.18	26.56	79.02	99.23
4	18	193	90.50	19.00	24.93	99.86
5	9	214	71.23	39.63	4.32	99.97
6	8	203	45.28	30.80	1.20	100.00

Table 2.5. Misclassification errors and prediction accuracy of training and test data without and with the use of boosting in See5 program for the six prediction models and calculated accuracies with and without using the majority filter.

Group Name	Without Boosting		With Boosting		Un-filtered map	Filtered map
	Training	Testing	Training	Testing		
Orders	11.6 (88.4)	14.2 (85.8)	7.0 (93.0)	11.6 (88.4)	89.00	90.02
Suborders	15.0 (85.0)	18.4 (81.6)	10.6 (89.4)	15.7 (84.3)	84.59	85.21
Great groups	19.7 (80.3)	25.3 (74.7)	13.9 (86.1)	21.2 (78.8)	79.05	81.22
Subgroups	20.9 (79.1)	27.1 (72.9)	15.2 (84.8)	22.7 (77.3)	77.41	79.43
Major-SMUs	23.2 (76.8)	29.4 (70.6)	17.1 (82.9)	24.6 (75.4)	75.61	77.71
All-SMUs	25.9 (74.1)	32.8 (67.2)	19.8 (80.2)	28.1 (71.9)	72.05	74.17

Table 2.6. Confusion matrix of soil Orders.

Orders	1	2	3	4	5	6	7	8	9	10	11	User Accuracy
1	3967	1		753	209						1	80.45
2		1637	79		183			2	19			85.26
3	1	38	1535		194			14	33			84.57
4	480			7006	71	2	3		2			92.62
5	60	14	34	32	10591			2	4		5	98.59
6				16	9	10						28.57
7				51			35					40.70
8		6	39	4	26			202				72.92
9		74	142		117			9	555			61.87
10					7					0		0.00
11	10			3	99						157	58.36
Producer Accuracy	87.80	92.49	83.93	89.08	92.05	83.33	92.11	88.21	90.54	0.00	96.32	90.02

Table 2.7. Confusion matrix of soil Suborders.

Sub-orders	1	3	4	6	7	8	9	10	11	12	13	14	15	16	17	18
1	3016	13				16	53	323								
3	5	1173				7	35	59	229							3
4			1585	115					204				21	3		
6			35	1423					134			1	38	21		
7					57				92							
8	202	7				247	123	172	1							
9	69	68				14	1758	432	31		1					
10	506	61				23	252	3528	73		5	1	1	1		
11	1	61	9	26	2		10	16	10656				5	1		11
12								14	8	4						
13							15	27			39					
14								4	5			30				
15			98	140					140				518	5		
16			11	35					26				1	168		
17					2				1						2	
18		16						7	86							176
Producer Accuracy	79.4	83.8	91.2	81.8	93.4	80.5	78.3	77.0	91.2	100	86.7	93.8	88.7	84.4	100	92.6
User Accuracy	88.16	77.6	82.2	86.1	38.3	32.9	74.1	79.3	98.7	15.4	48.2	76.9	57.5	69.7	40.0	61.6

Table 2.8. User accuracy of soil great groups, subgroups, major and all soil map units in the study area.

Great groups		Subgroups		Major soil map units		All soil map units	
GG	User accuracy	SG	User accuracy	SMU	User accuracy	SMU	User accuracy
1	88.37	1	92.40	8	42.30	1	45.45
4	85.81	5	85.87	9	75.67	8	40.24
6	84.36	6	71.60	12	44.25	9	67.87
8	84.30	8	87.99	17	84.33	10	31.25
9	74.07	13	63.89	18	76.10	12	54.97
10	56.34	19	48.89	19	73.61	13	0.00
11	37.47	20	42.90	20	83.43	17	81.48
12	50.16	21	31.75	23	71.01	18	68.03
13	75.37	22	49.37	24	73.95	19	71.68
14	75.31	23	53.71	40	91.15	20	83.48
15	68.38	24	60.87	48	54.23	21	36.54
16	91.88	25	84.22	49	50.35	22	61.76
17	62.12	26	28.11	50	40.74	23	66.42
18	83.61	27	69.01	52	60.23	24	67.91
19	25.00	28	39.41	53	95.09	27	79.66
20	63.54	29	46.42	56	82.04	28	55.81
21	67.50	31	71.62	57	77.33	29	25.20
22	81.28	32	73.05	61	73.43	30	69.09
23	72.31	33	80.75	68	77.08	32	68.89
24	64.64	34	43.11	86	94.32	33	53.66
25	50.00	36	69.57	87	51.72	36	58.82
26	19.05	38	67.83	90	93.06	37	81.25
27	0.00	40	92.69	91	83.55	38	30.00
28	92.86	41	81.18	94	77.86	40	85.20
29	75.87	42	58.62	95	74.68	46	58.33
Total	81.22		79.43		77.71		74.17

(Refer to Tables A.2.3 and A.2.4 for a complete accuracy assessment of the all groups)

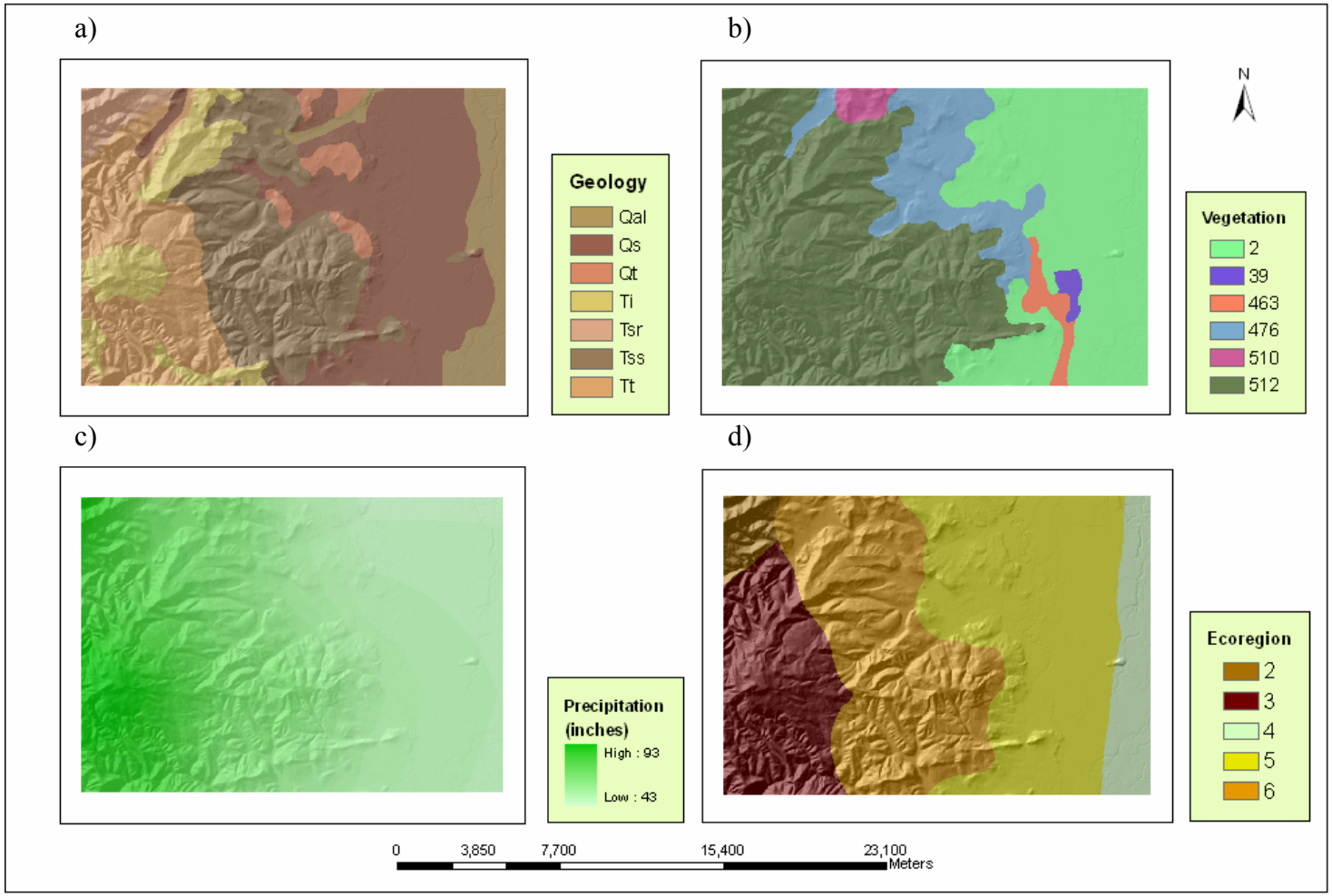

Figure 2.1. Environmental data used in developing soil prediction models: a) Geology, b) Vegetation, c) Precipitation, and d) Ecological regions.

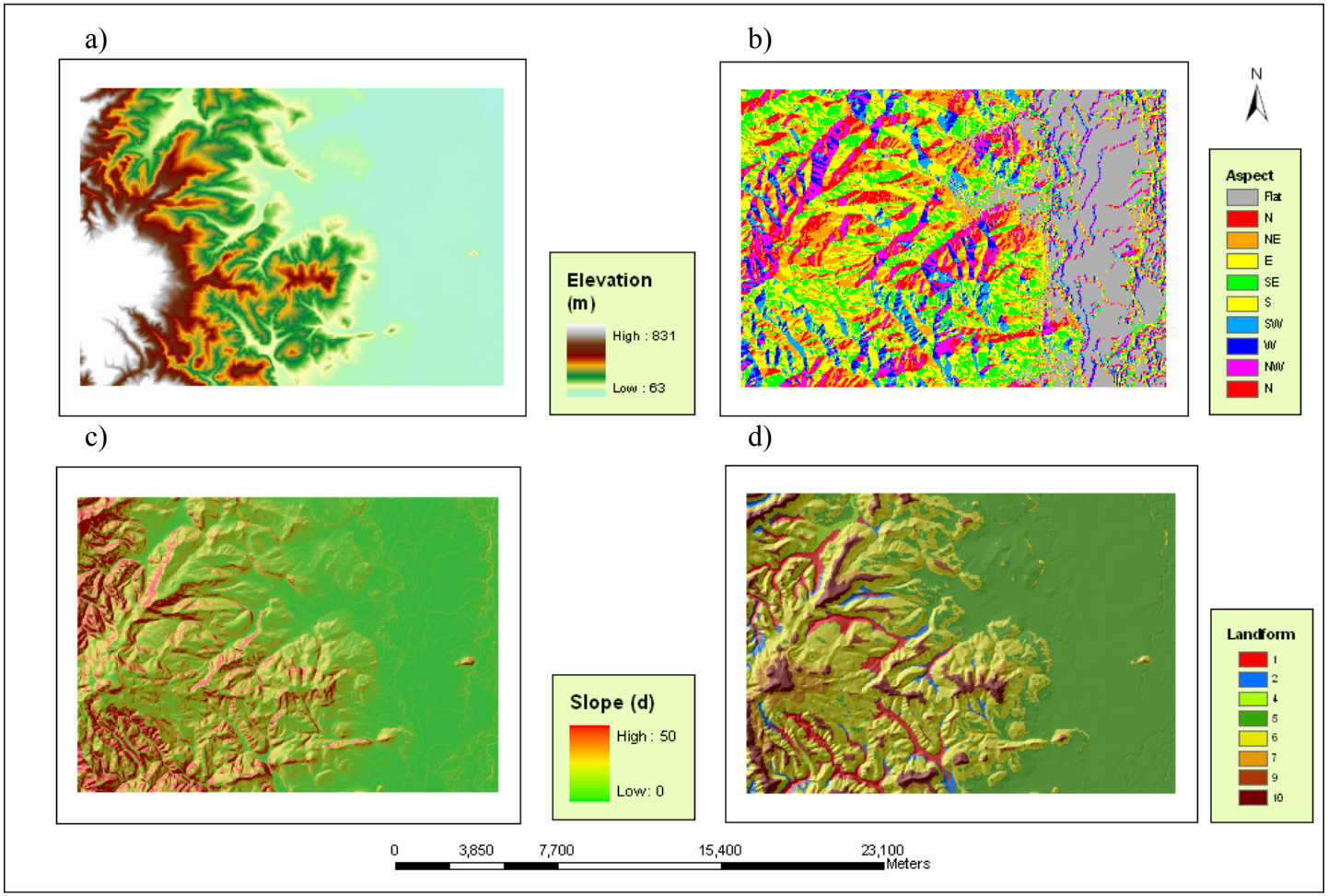

Figure 2.2. Terrain attributes developed from the digital elevation model (DEM): a) Elevation, b) Aspect, c) Slope, and d) Classified landforms.

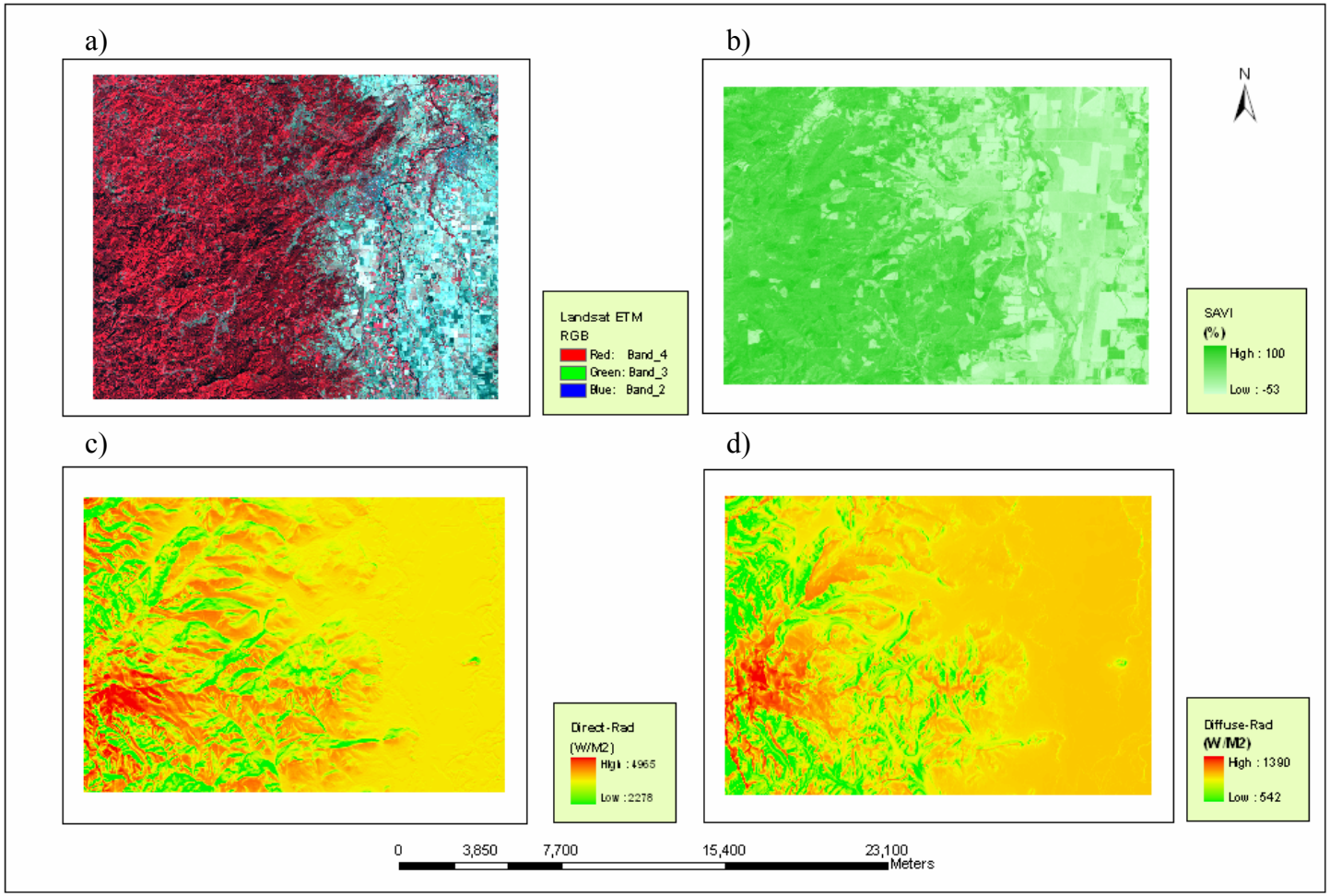

Figure 2.3. Landsat ETM+ and solar radiation data: a) False color composite of Landsat ETM+ image, b) SAVI index, c) Direct radiation, and d) Diffuse radiation.

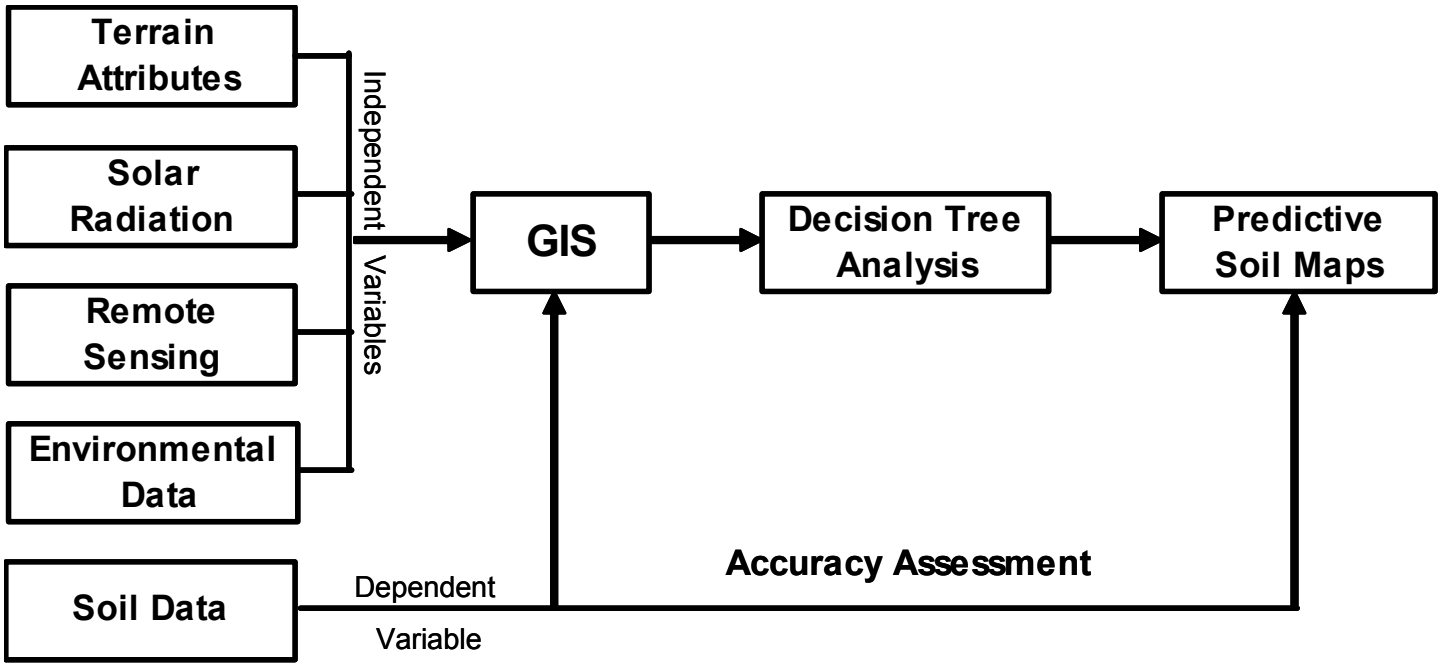

Figure 2.4. An overview of sources of geodatabases and the analytical methods

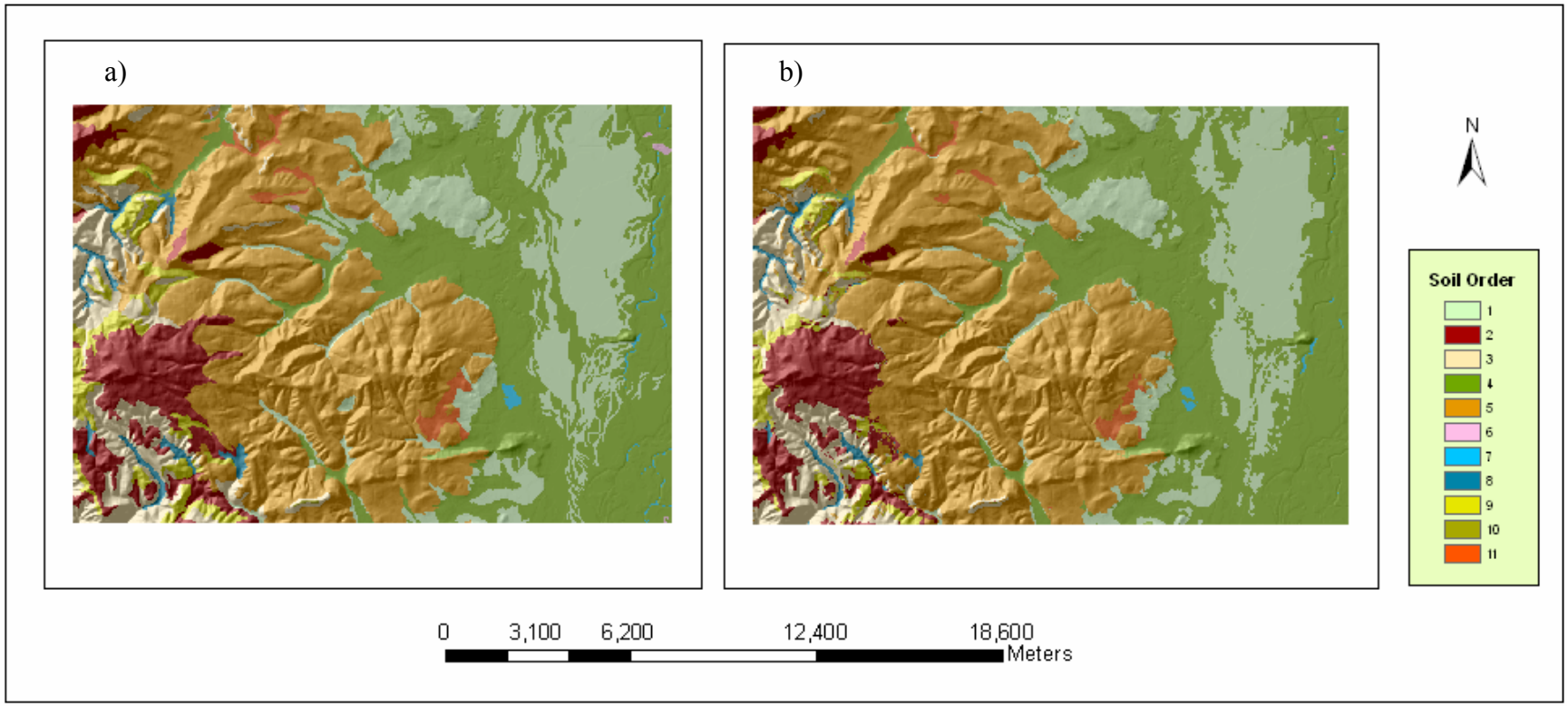

Figure 2.5. A comparison between the a) actual and b) predicted soil orders in the study area.
The codes refer to: 1. Alfisols; 2. Andisols; 3. Inceptisols; 4. Mollisols; 5. Ultisols; 6. Vertisols; 7. Water; 8. Complex of 1 and 3; 9. Complex of 2 and 3; 10. Complex of 3 and 4; and 11. Complex of 1 and 5.

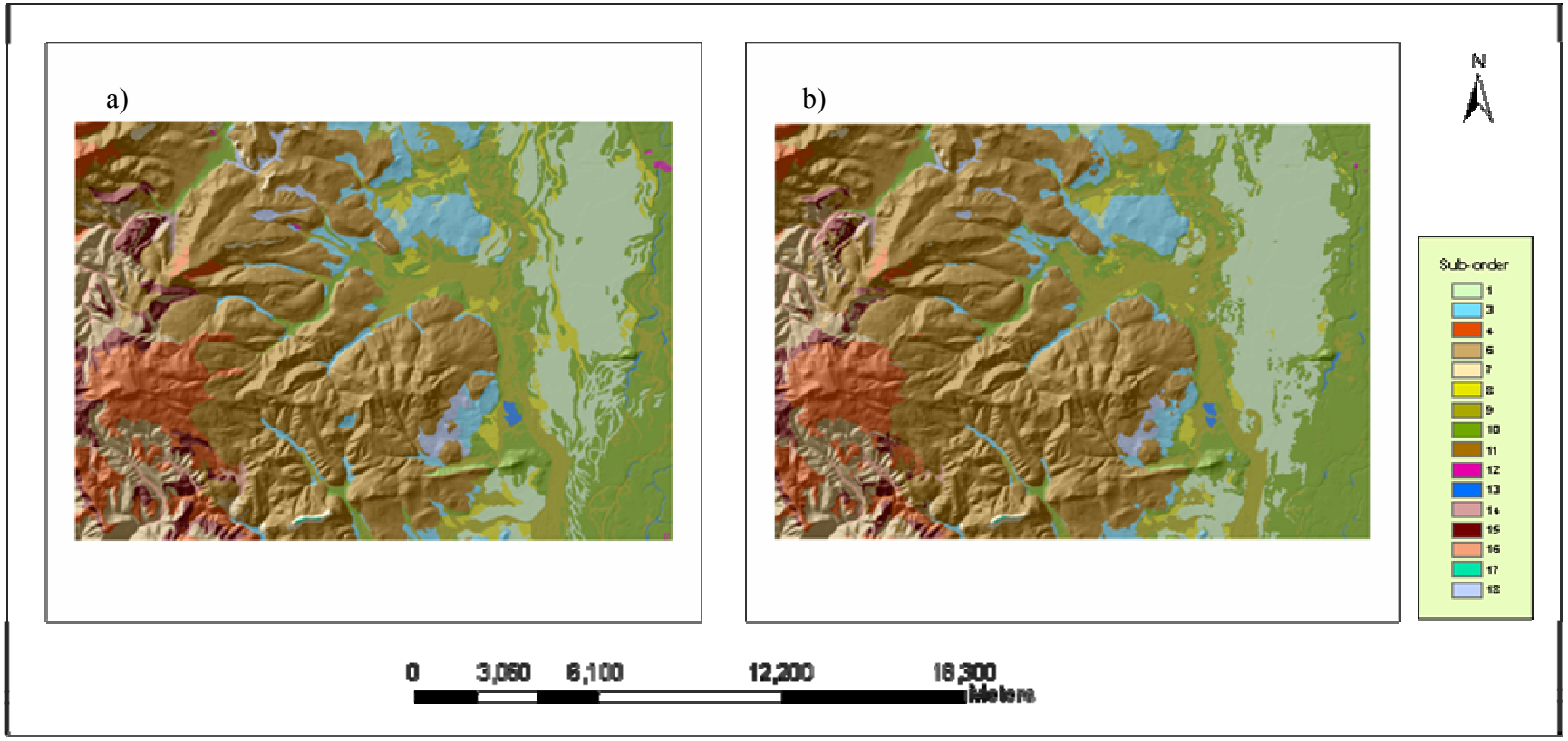

Figure 2.6. A comparison between the a) actual and b) predicted soil suborders in the study area.
The codes refer to: 1. Aqualfs; 2. Udalfs; 3. Xeralfs; 4. Udands; 5. Aquepts; 6. Udepts; 7. Xerepts; 8. Albolls; 9. Aquolls; 10. Xerolls; 11. Humults; 12. Aquerts; 13. Water; 14. Complex of 2 and 5; 15. Complex of 4 and 6; 16. Complex of 2, 5, and 6; and 18. Complex of 3 and 11.

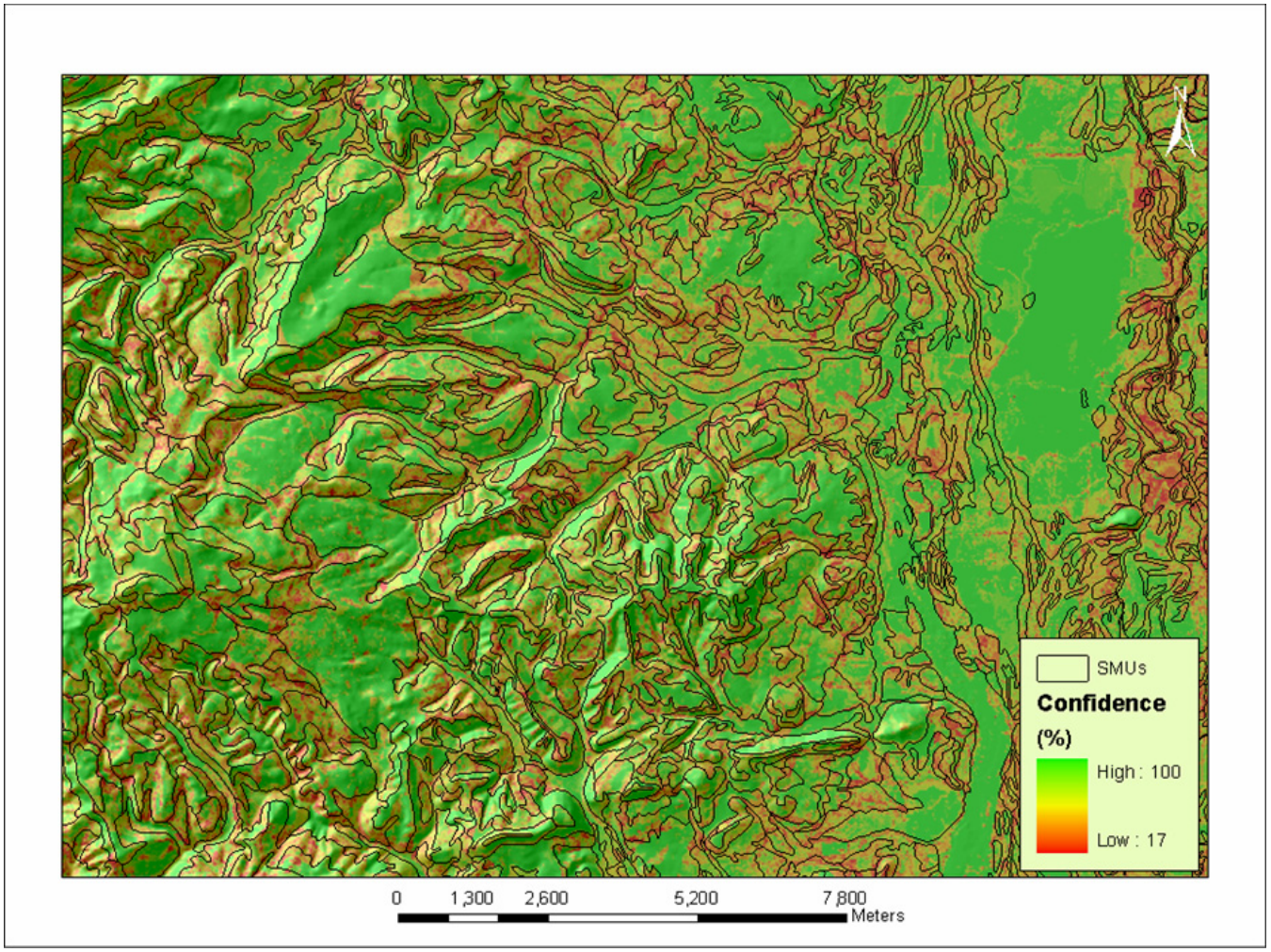

Figure 2.7. The boundaries between SMUs in the study area overlaid on the confidence data layer.

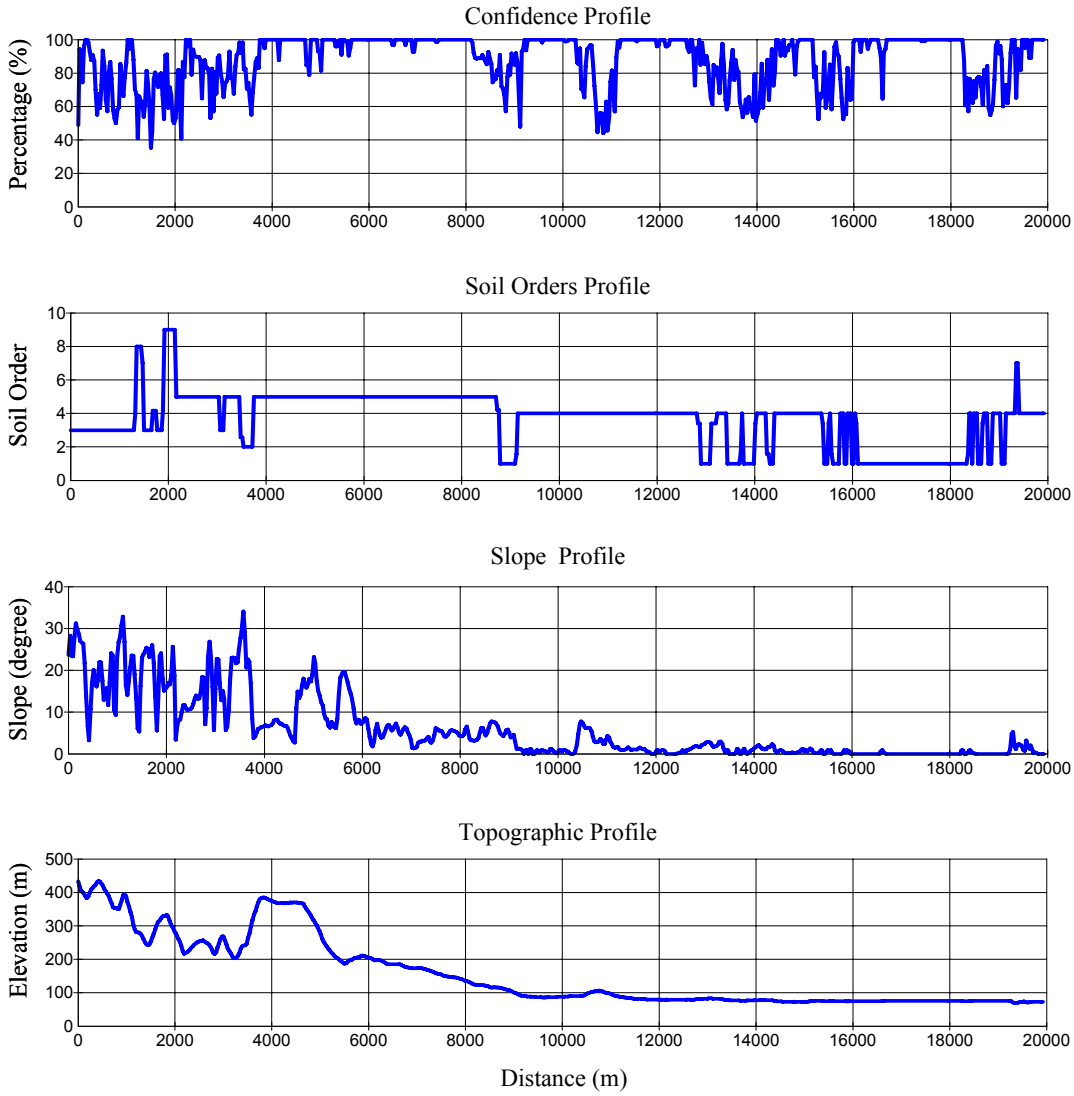

Figure 2.8. The influence of topographic and slope profiles on the prediction accuracy of soil orders.

CHAPTER 3

Spatial Data Mining and Soil-landscape Modeling Applied to Soil Survey

Abdelhamid A. Elnaggar and Jay S. Noller

Prepared for submission to:

Geoderma

Abstract

Data mining techniques are studied to recover knowledge from geodatabases in order to improve updates of existing soil maps and to help in developing a preliminary soil map for neighboring unmapped areas. Decision tree, one of the most widely used inductive learning methods, is used to retrieve the expert knowledge embedded in the soil-landscape model used by the Harney County, Oregon soil survey (ca. 1975-2003). The extracted model was extrapolated to develop a preliminary soil map for adjacent area in Malheur County, Oregon. Field data were used to test the prediction accuracy of the generated map. Also it was trained to develop a second model to produce another soil map for the study area in Malheur County. Spatial environmental data of geology, vegetation, precipitation, terrain attributes (elevation, slope, aspect and surface curvature), landforms, solar insolation and Landsat TM data at a resolution of 30 m were used to predict soil map units. Model efficiency was tested by making a comparison between the predicted and the present soil map for the reference area. Results show that 45 Soil Map Units (SMUs) out of 46 were successfully predicted with an over all accuracy of 92%. Prediction accuracy for the preliminary soil map of the unmapped area extrapolated based on the soil-landscape model of the reference area was very low. Few SMUs were predicted with significant accuracy, mostly those shallow SMUs that have either a lithic contact with the bedrock or developed on a duripan. On the other hand, the developed soil map based on field data was predicted with very high accuracy. The overall accuracy of that map was about 97%. Decision tree proved to be a powerful tool in retrieving the spatial relations between SMU and

the environmental variables. Also, it could be a very helpful approach in developing soil maps in more objective, effective, less-expensive and fast ways based on field data.

Keywords: Predictive soil mapping, Decision tree, Pedometrics, Soil-landscape models

3.1. Introduction

There is a growing demand for accurate and multi-resolution soil data in order to sustain agricultural and environmental development. Producing soil survey maps in the traditional way is an expensive, labor-intensive, and time-consuming process. Consequently, many areas worldwide still do not have soil maps. Also, updating of existing soil maps takes more than 15 years. This is in addition to errors associated with conventional soil survey maps such as inclusions and inaccurate boundaries (Burrough, 1986; Ehlschlaeger and Goodchild, 1994). However, locational errors are not only restricted to errors by experts but they also occur as a result of the nature of soils, where soil varies gradually at the boundaries. Accordingly, the boundaries between SMUs are often diffused not sharp (Mark and Csillag, 1989).

Traditional soil survey maps are developed based on an empirical model derived from the inductive reasoning of the field observations. This model is called the soil landscape model which is based on the interchangeable relationships between soils and their environmental variables. Unfortunately, the information about the soil landscape model used in developing soil survey maps is not well documented. Recently, different methods of inductive learning (i.e., decision tree, fuzzy logic, neural network) have been used to retrieve most of the information about the soil-landscape model used to

create the soil maps (Zhu et al., 2001; Moran and Bui, 2002; Scull et al., 2003). These methods incorporate different environmental variables including geology, vegetation, and terrain attributes to predict the soil mapping units. Most of these data are available nowadays in digital form through geodatabase clearinghouses.

Decision-tree analysis (DTA) of spatial data is an approach widely used in natural resource mapping and it has potential for land cover mapping problems using remote sensing data (Friedl and Brodley, 1997; Friedl et al., 1999; Xian et al., 2002; Herold et al., 2003). It also has been successfully used in developing predictive soil maps for large areas (Hansen el al., 1996; Bui et al., 1999; Zhou el al., 2004; Scull et al., 2005). Decision tree analysis is used in this study for these reasons: 1) it is a non-parametric method for analyzing hierarchical relationships between variables; 2) it deals with the nonlinear relationships between some soil properties; 3) it handles both continuous and categorical variables; and 4) it develops interpretable prediction rules that can be extrapolated to similar landscapes (Venables and Ripley, 1994; Hansen et al., 1996; Huang et al., 2002). It is economically wise to benefit from the existing soil maps 20 or more years old and/or less-intensive field data in updating or generating preliminary soil maps for unmapped areas that developed under the same soil forming conditions.

The objectives of this study are to retrieve the spatial relations between soil map units (SMUs) and their environmental variables in a reference area (Harney County, OR); develop a soil prediction model; predict SMUs for a neighboring unmapped area (Malheur County, OR) and, finally, test the model efficiency and evaluate the prediction accuracy of the predicted soil map.

3.2. Materials and methods

3.2.1. Description of the study area

Two areas were used in this study (Fig. 3.1). One is the reference area (about 977 km^2), located in the south-eastern part of Harney County. Second is an adjacent unmapped area (about 1160 km^2), located in Malheur County, Oregon, USA. Soil map of the reference area was clipped from the digitized soil map of Harney County (SSURGO database developed by USDA-NRCS, 2004b)**.** There are 46 SMUs in the reference area; their map symbols; names; representing areas and taxonomic classification are present in Tables 3.1 and 3.2.

The common landforms in the area consist of rock pediments, alluvial fans, fan remnants, and playas. Most of the soils in the study area are developed on volcanic parent materials (basalt and andesite, tuffaceous sedimentary rocks, welded tuffs), lacustrine deposits and fluvial sedimentary rocks (Clarke and Bryce, 1997). Elevation varies from 1175 m to 2079 m ($\bar{X}$=1307 m) and slope ranges from 0 to 59^o ($\bar{X}$= 2.83^o). Mean annual precipitation (MAP) varies from 178 mm on low-leveled areas to 838 mm on high elevations ($\bar{X}$= 238.5 mm), soil moisture regime is mainly aridic. Minimum annual temperature is about 0^{o}C and the maximum annual temperature is about 17^{o}C. Aridisols is the most dominant soil order in the studied area, however there are some areas with weakly to slightly developed soils (Entisols and Inceptisols) and moderately developed soils (Mollisols).

Wyoming big sagebrush (*Artemisia tridentata wyomingensis*) represents the most prevalent vegetation in the studied area. It occurs on flat areas as well as gentle and steep slope areas. Shadscale saltbush (*Atriplex confertifolia*), greasewood

(*Sarcobatus vermiculatus*), black sagebrush (*Artemisia nova*), basin big sagebrush (*Artemisia tridentata tridentata*), and low sagebrush (*Artemisia arbuscula arbuscula*) also are dominant in some areas.

3.2.2. Data collection and preparation

Soil-forming factor model developed by Jenny (1941) represents the theoretical basis for this study, just as it has for many digital soil mapping research (Bui et al., 1999; Mckenzie and Ryan, 1999; McBratney et al., 2000; McBratney et al., 2003; Henderson et al., 2005; Zhou et al., 2004; Scull et al., 2005). Accordingly, environmental variables or attributes that have significant influence on soil development were integrated in developing the soil prediction models. Environmental variables used in this work are represented in Table 3.3. Terrain attributes (elevation, slope gradient, aspect, and plan and profile curvatures) were derived from the digital elevation model DEM of south and south east Oregon (10 m cell size) using the ArcGIS software Package (Fig. 3.2). Classified landforms were derived from the DEM using Topographic Position Index (TPI) (Weiss, 2001; Jenness, 2005). Solar insolation data were calculated during the summer solstice, equinox and winter solstice from slope gradient and aspect using the ArcView solar analyst extension (Fu and Rich, 1999). Study area covers parts of two Landsat TM images (P42R30 and P42R31) acquired on August 17, 2005 (Fig. 3.3). Images were processed to calculate the Normalized Vegetation Index (NDVI) (Rouse et al., 1973), Soil Adjusted Vegetation Index (SAVI) (Huete, 1988), Brightness, Greenness, and Wetness indices (Kauth and Thomas, 1976). Vegetation indices were used to distinguish between the densely and sparsely vegetated areas. Landsat Bands (1,

2, 3, 4, 5 and 7) and band ratios (b3/b1, b3/b4, b5/b3, and b5/b7) were also calculated and combined in the prediction model. Band ratios were used to interpret soil properties. For example, band ratio 3 to 1 is known to reflect iron content (Rowan et al., 1977). Band ratio 5 to 7 is found to have a strong correlation with clay minerals content in areas where vegetation is absent (Riaza et al., 2000). Band ratio 5 and 4 gave differences between iron oxide dominance and hydroxyl with areas of high oxides giving brighter pixels due to stronger absorption of the band 4. Other data include MAP from 1961 to 1990 (1:200,000, USDA-NRCS, 1999), geology (1: 500,000, USGS, 2003), historic vegetation (1:100,000, Tobalske, 2002) and Landfire vegetation (USFS, 2006) maps of Oregon.

Data sources were projected to UTM Zone 11, Datum NAD 83 and clipped to cover both the reference and the unmapped areas. Data layers were represented using the raster data model with 30m cell size. Vector data were recoded and converted to raster data. All raster data layers were created or converted to Imagine file format, so as to be sampled using the CART model under ERDAS imagine (Earth Satellite Corporation, 2003). Soil map units in the reference area were randomly sampled for all environmental variables to obtain the output data matrix for decision tree analysis. The output data matrix consists of 72,545 random sample points as training data and 24,718 random points as test data.

3.2.3. Decision tree analysis

Decision tree approach (Breiman et al., 1984; Quinlan, 1993) was used to retrieve the soil-landscape model used in producing the soil map of the reference area. Retrieved model was then used to develop a prediction of the reference soil map. After analysis and adjustments, prediction model was extrapolated to the neighboring unmapped area in Malheur County to generate the preliminary soil map.

Decision tree analysis was carried out using the See5 program (Quinlan, 2001). Soil-landscape models were retrieved using 10 of boosting[1] to enhance the prediction accuracy (Friedl et al., 1999; Moran and Bui, 2002). Classifying observations in the training dataset may not be expressed as a function of the attribute values. These observations could occur as a result of either an error in the attribute values or the attributes do not provide sufficient information to classify the object. As a result, continuous division of the training dataset until all subsets contain members of a single class may be impossible. Although this division could be possible, it may be inadvisable. Dealing with such a problem is to allow the tree to grow up and then remove unimportant portions by pruning it (Quinlan, 1990; Eklund et al., 1998). In pruning function a subtree is replaced by one of its branches or by a leaf, which is very common, resulting in a smaller tree but with greater accuracy. Developed tree was pruned by 35% to reduce over-fitting of decision trees. Decision tree produced by See5 program was applied using the CART model under ERDAS image (Earth Satellite Corporation, 2003) to develop the predicted soil maps.

[1] In boosting, a sequence of decision trees is developed; each subsequent tree attempts to fix the misclassification errors in the previous one. Each decision tree makes a prediction and the final prediction is a weighted vote of the predictions of all trees.

3.2.4. Field Work

Field data were used in this work to evaluate the prediction accuracy of the predicted soil map extrapolated from the reference area for the unmapped area in Malheur County. A random sampling technique was used in the beginning to select sampling locations, although it was difficult or undesirable to reach all these locations (De Gruitjer, 2000). However, we tried to get as close as possible to the sampling location using the dirt road map that we developed for the area from the digital ortho quads (DOQ) (USGS, 2000-2001). About 210 soil profiles (Fig. 3.4) were described and classified using established methods (USDA-NRCS, 2002).

Field data also were used in developing a predictive soil map for the unmapped area in Malheur County. To accomplish this, each sampling location in the field data was assumed representative and buffered using a stratified buffer based on the topographic and geologic properties at that location. Buffered locations were randomly sampled and used to train the predictive model.

3.2.5. Model Evaluation

Both simple-descriptive and discrete-multivariate statistics (Jensen, 1996) were used to evaluate the predictive maps of the reference and the unmapped areas. Descriptive statistics include producer's accuracy, user's accuracy and overall accuracy. Overall accuracy is computed by dividing the total of correctly predicted SMUs by the total number of pixels in the error matrix. Producer's accuracy measures the exclusion errors and user's accuracy measures the inclusion errors were calculated for each SMU.

Kappa analysis (Cohen, 1960) a discrete multivariate technique, was used to measure the agreement between the model predicted SMUs and the real SMUs. Kappa is computed from (Mather, 2004):

$$K = \frac{N\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r}\left(x_{i+} * x_{+i}\right)}{N^2 - \sum_{i=1}^{r}\left(x_{i+} * x_{+i}\right)} \quad (1)$$

where N is the total number of sites in the error matrix, r is the number of rows in the matrix, x_{ii} are the diagonal entries of the error matrix, x_{i+} and x_{+i} indicate the sum of row i and the sum of column i of the error matrix, respectively. The K value is a measure of agreement or accuracy. It varies from zero to one, where zero indicates no agreement and one indicates total agreement (Congalton, 1991). Kappa equals zero when the estimates and field data are statistically independent.

Model efficiency was tested first by making a comparison between predicted and present SMUs in the Harney County reference map. The efficiency test was performed with about 50,000 random sample points of the soil map of the reference area, and it was assumed that this map is accurate. Accuracy of the predictive soil map produced for the unmapped area, in all iterations, was tested using field observations.

3.3. Results

3.3.1 Predictor variables and their significance

Although there were many input variables used as predictors (Table 3.3), not all of them have a significant influence in predicting SMUs in the reference area. Accordingly, input variables were arranged by their predictive significance using the

"Winnow Attributes" function in the See5 program. Preliminary results showed that only 21 variables have significant influence in predicting SMUs in the reference area. Historic vegetation, geology, elevation, ecological habitat, precipitation, slope gradient, distance from rivers, slope aspect, greenness index, and NDVI were the most significant variables, which agrees with Jenny's model. Those variables were followed by Landsat band 1, land fire vegetation, classified landforms, wetness index, diffuse radiation in equinox, Landsat bands 7, 5 and 4, diffuse radiation in winter solstice, and Landsat bands 3 and 2, respectively. Only variables with significant influence in predicting SMUs were integrated in this study to the retrieve soil-landscape model; all others were omitted.

3.3.2. Predictive soil map of the reference area in Harney County

Soil map of the reference area was successfully produced using decision tree analysis (Fig. 3.5), as demonstrated by a low (0.5%) misclassification error within the training data of about 150,000 cases. Further, the misclassification error within the validation data of about 50,000 cases was 8.4%, again a value below our initial error threshold of 10%.

Agreement between the predicted and present SMUs in the reference area was tested using the producer's accuracy, user's accuracy, overall accuracy and Kappa statistics. About 50,000 random sampling points were used to create the confusion matrix and calculate the prediction accuracy. Forty-five SMUs were predicted out of 46 SMUs in the reference area. Overall accuracy was about 92% and the Kappa coefficient was 0.91 (Table 3.4). Producer's accuracy varied from 54.97 to 100%, with an average

of 89.85%. User's accuracy ranged between 54.61 and 99.42%, with an average of 89.04%.

3.3.3. Predictive soil map of the of the unmapped area in Malheur County

3.3.3.1. First map derived from the reference map

Soil-landscape model retrieved from the reference area was extrapolated to develop a preliminary soil map for the adjacent unmapped area in Malheur County (Fig. 3.6). Confidence in predicting SMUs is inversely related with distance from the reference area (Fig. 3.7).

Comparing field data with the predicted soil map for the unmapped area in Malheur County shows that most SMUs were not predicted correctly. However, few of these SMUs were predicted with significant confidence. For example, SMUs 21, 23, 24, 73, and 259 were predicted with accuracy of 67, 67, 42, and 89%, respectively. SMUs 21, 23, and 24 (Atlow, Atlow- rock outcrop complex, and Atlow and Skedaddle complex) are shallow soils to bedrock formed in residuum from chert, argillite, shale, altered rhyolitic tuff and andesite. Mostly, they are located on mountain and hill summits, crests, shoulders and sideslopes. Soil map unit 73 which is a "complex" of Deppy and Tumtum soil series is also a shallow soil to a duripan. It is developed in alluvium derived from volcanic rocks. Soil map unit 259 or playa (alkali flat or sabkha) is bare soil with white evaporates.

3.3.3.2. Second map derived from the field data

Because the results obtained from the first map were not satisfactory, we produced a second soil map derived from the collected field data (Fig. 3.8). In this version, all SMUs in the developed map were predicted with accuracies higher than 81% except for SMUs 152 and 288 (accuracies of 46 and 33%, respectively). Overall accuracy of SMUs in the unmapped area was 97.38% and the Kappa coefficient was 0.97. Producer's and user's accuracies are represented in Table 3.5. Producer's accuracy varied from 87.10 to 100%, with an average of 97.38%. User's accuracy ranged between 33.33 and 99.92%, with an average of 88.70%.

3.4. Discussion

Observations from the predicted soil map of the reference area show a consistent pattern in prediction accuracy among SMUs based on the representative areas. Soil map units that represent larger areas were predicted with higher accuracy compared to those representing smaller areas (Table 3.4). On the other hand, SMUs (e.g., 36, 249 and 283) represent proportionally smaller areas and were predicted with lower accuracies. Smaller SMUs are represented by fewer numbers of sampling points when the randomized sampling technique is used due to their small areas. As a result, they are poorly characterized in the output data matrix and they also penalize by the decision tree algorithm (Friedl and Brodley 1997; Elnaggar and Noller, 2007). However, prediction accuracy of all SMUs in the reference map was greater than 50% except for SMU 137. Soil map unit 137 (Hackwood) represents the smallest SMU in the reference

map (covers only 7 pixels (6300m^2)), which could be the main reason for it not being well predicted.

Soil map units predicted with higher accuracy could be a result of their association with distinct features in the input data such as certain types of geology or vegetation which make it from easy them to be distinguished from the other SMUs. Playas, for example, have unique properties that could be well identified by more than one variable such as geology, vegetation, terrain attributes (elevation, slope and aspect), Landsat images and indices (bands (1, 2 and 3), brightness, greenness, and wetness). In contrast, SMUs are predicted with lower confidence because: 1) it was hard to discriminate them from the input data; 2) they share some properties with other SMUs; or 3) they are integrated in soil complexes. Soil map unit 21 is Atlow, where SMUs 22, 23, 24, 245, 300, 301, and 302 are complexes of Atlow with other SMUs and those share some, but not all, of its properties.

Scale inconsistency among the input data could be another factor causing significant reduction in prediction accuracy (Bui et al., 1999; Zhou et al., 2004; Scull et al., 2005). Geologic data have a scale of 1:500,000, whereas soil map has a scale of 1:24,000. When we started our field work, we found this coarse-scale geologic data to be one of the main reasons for having an overlap among SMUs and misdelineation of their boundaries. Coarse-scale data have great influence on the misclassification of smaller SMUs.

Prediction accuracy of the preliminary soil map derived from the reference area was very low when evaluated using the field data. This could be due to one or more of these factors: First, the reference map and the other environmental data, especially those

area-class maps such as geology and vegetation maps, were treated as accurate maps in this work. This could be a wrong assumption, where the soil map of the reference area goes back to 1975 to 1977 and it has not been updated until now. Also, it does not have information about their accuracy or the empirical model used in producing it. Second, predictor variables with coarser scales could result in misallocation, inclusion and overlapping errors as mentioned before. Third, environmental variables that cover the study area in both counties may not be developed using a seamless approach. Seamless approach means data are continuous across the political or artificial boundaries and data are mapped using the same criteria and the same mapping scale. Fourth, the study area may be less represented by the sampling points (either the number of the sampling points or their distribution) because of the accessibility problem.

It was noticed that soil map units predicted with higher accuracy are mostly shallow soils such as Altow. This could be because they are in a direct contact with their bedrock or parent material. Once these parent materials are located it is easy to predict SMUs developed on them. Playa was the other SMU predicted with high confidence, which also could easily be distinguished from the input data.

One of the most important pieces of information revealed from the confidence map of the unmapped area was the relationship between the prediction confidence of SMUs and their distances from the reference area. This indicates that the extrapolated soil-landscape model works well in neighboring areas which may share similar environmental conditions.

Results of using field data in developing soil map for the unmapped area in Malheur County are significant. Here, field data more accurately reflect relevant

information about soil properties and their formative environment. However, fewer soil map units (21 SMUs) were predicted in the unmapped area. This could be due to the limited number of sampling points, and/or the uneven distribution of these points.

Confidence mapping shows the model to be poor in predicting SMUs located at high elevation areas (i.e., mountain ridges and mesas) where there are no sampling points. Low confidence was found in predicting SMUs in areas that have significant changes in slope (narrow valleys). This could be due to the higher activity of surficial processes (erosion and deposition) at these locations which result in weakly developed soils (Elnaggar and Noller, 2007). Also, low confidence was found at low-elevation areas and along the boundaries between SMUs. This could be related to the coarser scale of some environmental data, where small details in soil topography and geology could not be supported by these data. Also, it could be inherited from the locational errors associated with area-class maps (i.e., geology) used in the prediction model.

3.5. Conclusion

Decision tree can be a helpful approach in retrieving soil-landscape models embedded in old soil survey maps. Under certain circumstances, retrieved models can be extrapolated to develop preliminary maps over areas having similar landscapes. By using decision tree analysis, a wide variety of environmental variables can be integrated and mined to predict the spatial distribution of soils and their properties.

Results show that using decision tree analysis in developing predictive soil maps from field data provides better results compared to extrapolating retrieved soil-landscapes models from reference areas. These results could be very helpful in reducing

the great amount of field data required in the conventional soil mapping techniques. It could significantly help in reducing the great amount of time consumed in developing soil maps in the traditional ways and facilitate their updates. The most important goals are producing soil maps in more objective, quantitative and less-expensive ways, and providing information about accuracy of the developed maps which are not available in current soil maps.

Results also revealed that field conditions could restrain collecting more representative data. Therefore, new sampling techniques should be developed to facilitate the process of collecting and testing soil properties under field conditions. These techniques should take in consideration field accessibility problems and also they should not significantly impair the accuracy assessment of the results.

Certain points should be considered to enhance the results of using the decision tree approach in predictive soil mapping. First, assuming that available soil survey maps and environmental data reference areas are accurate could be untrue for some areas. Accordingly, we recommend that soil maps and all other data integrated in predictive soil mapping techniques should be evaluated in advance for accuracy or their accuracy should be certified. Second, extrapolating retrieved soil-landscape models has spatial limits. They should be applied to neighboring areas that experience similar environmental conditions and topography.

Acknowledgments

We gratefully acknowledge Mark Keller, soil survey project leader, Alina Rice, soil scientist, and Charlie Tackman, BLM scientist, in the BLM Office in Vale, Oregon for

their great support and help with the field work and accommodation. Also, we thank

Joan Sandeno for editing the manuscript.

References

Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J. 1984. Decision and regression trees. Wadsworth Belmont, CA.

Bui, E. N., Loughhead, A., Corner, R. 1999. Extracting soil-landscape rules from previous soil surveys. Aust. J. Soil Res., 37:495-508.

Burrough, P.A. 1986. Principles of geographical information systems for land resources assessment. New York: Oxford Univ. Press, p. 193.

Clarke, S. E., Bryce, S. A. 1997. Hierarchical subdivisions of the Columbia Plateau and Blue Mountains Ecoregions, Oregon and Washington. Portland: U.S. Department of Agriculture-Forest Service General Technical Report PNW-GTR-395, 114 p.

Cohen, J. 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20:37-46.

Congalton, R. G. 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of the Environment. 37:35-46.

De Gruitjer, J.J., 2000. Sampling for spatial inventory and monitoring of natural resources. Technical report, Alterrrarapport 070, Wageningen, Alterra, Green World Research.

Earth Satellite Corporation 2003. CART software user's guide. U.S. Geological Survey – National Land Cover Database (NLCD).

Ehlschlaeger, C. R., Goodchild, M. F. 1994. Dealing with uncertainty in categorical coverage maps: Defining, visualizing, and managing errors. In Proceedings of the Workshop on Geographical Information Systems at the Conference on Information and Knowledge Management, Gaithersburg, Maryland: 86–91.

Eklund, D P. W., Kirkby, S. D., Salim, A. 1998. Data mining and soil salinity analysis. Int. j. geographical information science, 12(3):247-268.

Elnaggar, A. A., Noller, J. S. 2007. Assessing the consistency of conventional soil survey data: Switching from Conventional to Digital Soil Mapping Techniques. In press.

Friedl, M. A., Brodley, C. E. 1997. Decision tree classification of land cover from remotely sensed data. Remote Sensing of the Environment. 61:399-409.

Friedl, M. A., Brodley, C. E., Strahler, A. H. 1999. Maximizing land cover classification accuracies produced by decision trees at continental to global scales. IEEE Transactions on Geoscience and Remote Sensing. 37(2):969-977.

Fu, P., Rich, P. M. 1999. The solar analyst 1.0 User's manual. Helios Environmental Modeling Institute, LLC. http://www.hemisoft.com.

Hansen, M., Dubayah, R., DeFries, R. 1996. Decision trees: an alternative to traditional land cover classifiers. Int. J. Remote Sensing, 17(5):1075-1081.

Henderson, B. L., Bui, E. N., Moran, C. J., Simon, D. A. P. 2005. Australia-wide predictions of soil properties using decision trees. Geoderma, 124: 383-398.

Herold, N. D., Koeln, G., Cunnigham, D. 2003. Mapping impervious surfaces and forest canopy using classification and regression tree (CART) analysis. ASPRS 2003 Annual Conference Proceedings. Anchorage, Alaska.

Huang, C., Davis, L. S., Townshend, J. R. G. 2002. An assessment of support vector machines for land cover classification. Int. J. Remote Sensing, 23(4): 725-749.

Huete, A. 1988. A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment, 25:295-309.

Jenness, J. 2005. Topographic position index (tpi_jen.avx) extension for ArcView 3.x. Jenness Enterprises. http://www.jennessent.com/arcview/tpi.htm

Jenny, H., 1941, Factors in soil formation: New York, McGraw-Hill, 281 p.

Jensen, J. R. 1996. Introductory digital image processing: A remote sensing perspective. 2nd Ed., Prentice Hall, Inc., Upper Saddle River, New Jersey, USA.

Kauth, R. J., Thomas, G. S. 1976. The tasseled cap: a graphic description of the spectral-temporal development of agricultural crops as seen by LANDSAT. Proceedings of the Symposium on Machine Processing of Remotely Sensed Data, Purdue University of West Lafayette, Indiana, p. 4B-41 to 4B-51.

Mark, D. M., Csillag, F. 1989. The nature of boundaries on area-class maps. Cartographica, 21:65-78.

Mather, P. M. 2004. Computer processing of remotely-sensed images – an introduction. 3rd Ed., John Wiley and Sons Ltd., Chichester, England.

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

McBratney, A. B., Odeh, I. O. A., Bishop, T. F. A., Dunbar, M. S., Shatar, T. M. 2000. An overview of pedometric techniques for use in soil survey. Geoderma, 87:293-327.

Mckenzie, N. J., Ryan, P. J. 1999. Spatial prediction of soil properties using environmental correlation. Geoderma, 89:67-94.

Moran, C. J., Bui, E. N. 2002. Spatial data mining for enhanced soil map modeling. Int. J. Geographical Information Science, 16(6):533-549.

Quinlan, J. R. 1990. Learning Logical Definitions from Relations. Machine Learning, 5(3): 239-266.

Quinlan, J. R. 1993. C4.5: Programs for machine learning. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Quinlan, J. R. 2001. See5: An Informal Tutorial. http://www.rulequest.com.

Riaza, A., Mediavilla, R., Santistieban, J. I. 2000. Mapping geological stages of climate-dependent iron and clay weathering alteration on lithologically uniform sedimentary units using Thematic Mapper imagery (Tertiary Duero Basin, Spain). Int. J. Remote Sensing, 21(5):937-950.

Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W. 1973. Monitoring vegetation systems in the Great Plains with ERTS, Third ERTS Symposium, NASA SP-351, 1:309-317.

Rowan, L. C., Goetz, A. F. H., Ashley, R. P. 1977. Discrimination of hydrothermally altered and unaltered rocks in the visible and the near infrared multispectral images. Geophysics, 42:522-535.

Scull, P., Franklin, J., Chadwick, O. A. 2005. The application of decision tree analysis to soil type prediction in a desert landscape. Ecological Modelling. 181:1–15.

Scull, P., Franklin, J., Chadwick, O. A., McArthur, D. 2003. Predictive soil mapping: a review. Progress in Physical Geography, 27(2):171–197.

Tobalske, C. 2002. Map of historic vegetation for the State of Oregon. Oregon Natural Heritage Program. Portland, Oregon.

USDA-NRCS. 1999. Oregon annual precipitation. National Cartography and Geospatial Center. Fort Worth, TX.

USDA-NRCS. 2002. Field book for describing and sampling soils. Version 2. National Soil Survey Center - Natural Resources Conservation Service - U.S. Department of Agriculture.

USDA-NRCS. 2004b. Soil survey geographic (SSURGO) database for Harney County area, Oregon. U.S. Department of Agriculture, Natural Resources Conservation Service, Fort Worth, Texas.

USFS. 2006. Landfire existing vegetation cover. USDA forest Service. Missoula, Montana.

USGS. 2000-2001. Digital Orthophoto Quadrangles. U.S. Geological Survey, Reston, VA.

USGS. 2003. Spatial digital database for geologic map of Oregon. U.S. Department of Interior, U.S. Geological Survey.

USGS. 2005. LANDSAT TM – Path: 42 Row: 30 for Scene: 5042030000519810 and Path: 42 Row: 31 for Scene: 5042031000519810. U.S. Geological Survey Center for Earth Resources Observation and Science (EROS). Sioux Falls, SD.

USGS-EROS Data Center. 1999. Oregon 10m DEM. U.S. Geological Survey, Sioux Falls, SD.

Venables, W. N., Ripley, B. D. 1994. Modern Applied Statistics with S-PLUS. Springer-Verlag: New York.

Weiss, A. 2001. Topographic position and landforms analysis. Poster Presentation, ESRI User Conference, San Diego, CA.

Xian, G., Zhu, Z., Hoppus, M., Fleming, M. 2002. Applications of decision-tree techniques to forest group and basal area mapping using satellite imagery and forest inventory data. Pecora 15/Land Satellite Information IV/ISPRS Commission I/FIEOS Conference Proceedings.

Zhou, B., Zhang, X., Wang, R. 2004. Automated soil resources mapping based on decision tree and Bayesian predictive modeling. J. Zhejiang Univ. Sci., 5(7):782-795.

Zhu, A. X., Hudson, B., Burt, J., Lubich, K., Simonson, D. 2001. Soil mapping using GIS, expert knowledge, and fuzzy logic. Soil Sci. Soc. Am. J., 65:1463-1472.

Table 3.1. Map symbols, names and percentages of SMUs in the reference area.

Symbol	Map Unit Name	Area (%)
4	Alvodest silty clay loam, 0 to 3 percent slopes	7.16
5	Alvodest-Playas complex, 0 to 2 percent slopes	0.35
21	Atlow very stony loam, 5 to 30 percent slopes	0.88
22	Atlow-Rock outcrop complex, 5 to 30 percent slopes	1.49
23	Atlow-Rock outcrop complex, 30 to 50 percent slopes	1.15
24	Atlow-Skedaddle complex, 5 to 30 percent slopes	14.17
36	Berdugo silt loam, 0 to 3 percent slopes	0.23
40	Boravall-Playas complex, 0 to 3 percent slopes	0.94
45	Brabble-Calderwood complex, 5 to 25 percent slopes	1.26
69	Davey sandy loam, 0 to 8 percent slopes	3.72
70	Davey-Oreanna complex, 0 to 8 percent slopes	1.17
72	Deppy very gravelly loam, 5 to 15 percent slopes	1.46
73	Deppy-Tumtum complex, 5 to 15 percent slopes	11.46
76	Dixon gravelly fine sandy loam, alkali, 0 to 2 percent slopes	1.19
77	Dixon gravelly sandy clay loam, 3 to 15 percent slopes	1.69
78	Dixon-Droval complex, 0 to 2 percent slopes	2.33
86	Droval loam, 0 to 3 percent slopes	4.29
93	Enko loamy sand, 2 to 8 percent slopes	2.93
96	Enko-Catlow association, 2 to 20 percent slopes	0.69
103	Felcher-Rock outcrop complex, 40 to 70 percent south slopes	1.53
131	Goldrun-Alvodest complex, 0 to 12 percent slopes	7.13
137	Hackwood gravelly loam, 20 to 35 percent slopes	0.00
152	Kerrfield loam, 3 to 20 percent slopes	1.04

Table 3.1. Map symbols, names and percentages of SMUs in the reference area (Continued).

Symbol	Map Unit Name	Area (%)
165	Langslet silty clay, 0 to 2 percent slopes	0.09
178	Lonely-Robson association, 5 to 25 percent slopes	1.01
182	Madeline very stony loam, 15 to 40 percent south slopes	0.47
192	McConnel cobbly sandy loam, 3 to 8 percent slopes	4.58
216	Nevador very gravelly sandy loam, 3 to 12 percent slopes	2.53
235	Norad silt loam, 0 to 1 percent slopes	0.05
245	Olac-Atlow complex, 2 to 10 percent slopes	0.20
248	Outerkirk sandy loam, 1 to 4 percent slopes	0.40
249	Outerkirk sandy loam, silty substratum, 2 to 6 percent slopes	0.32
250	Outerkirk-Defenbaugh association, 1 to 4 percent slopes	0.04
251	Ozamis silt loam, 0 to 1 percent slopes	1.58
256	Pernty-Rock outcrop complex, 30 to 70 percent south slopes	0.23
259	Playas	9.92
272	Raz-Brace complex, 2 to 20 percent slopes	0.42
282	Rio King loam, 1 to 6 percent slopes	0.67
283	Rio King-Droval complex, 0 to 2 percent slopes	0.08
288	Robson-Fourwheel complex, 3 to 30 percent slopes	0.98
291	Rock outcrop and Rubble land, 20 to 60 percent slopes	0.53
300	Skedaddle-Atlow-Rock outcrop complex, 5 to 30 percent slopes	1.90
301	Skedaddle-Atlow-Rock outcrop complex, 30 to 50 percent slopes	3.00
302	Skedaddle-Rock outcrop complex, 40 to 70 percent slopes	0.71
312	Spangenburg silty clay loam, thick surface, 0 to 2 percent slopes	0.66
334	Tumtum cobbly loam, 4 to 15 percent slopes	1.39

Table 3.2. Taxonomic classification of soils series in the reference area of Harney County.

Soil Name	Taxonomic Classification
Alvodest	Fine, montmorillonitic, mesic Sodic Aquicambids
Atlow	Loamy-skeletal, mixed, mesic Lithic Xeric Haplargids
Berdugo	Fine, montmorillonitic, mesic Xeric Paleargids
Boravall	Fine, montmorillonitic (calcareous), mesic Aeric Halaquepts
Brabble	Fine-loamy, mixed, mesic Xeric Haplodurids
Brace	Fine-loamy, mixed, frigid Xeric Argidurids
Calderwood	Loamy-skeletal, mixed, mesic Lithic Xeric Haplocambids
Catlow	Loamy-skeletal, mixed, mesic Durinodic Xeric Haplocambids
Davey	Sandy, mixed, mesic Xeric Haplocambids
Defenbaugh	Fine-loamy, mixed, mesic Typic Haplocambids
Deppy	Loamy, mixed, mesic, shallow Argidic Argidurids
Dixon	Fine-loamy over sandy or sandy-skeletal, mixed, mesic Xeric Haplocambids
Droval	Fine, montmorillonitic, mesic Sodic Aquicambids
Enko	Coarse-loamy, mixed, mesic Durinodic Xeric Haplocambids
Felcher	Loamy-skeletal, mixed, mesic Xeric Haplocambids
Fourwheel	Fine, montmorillonitic, frigid Vertic Paleargids
Goldrun	Mixed, mesic Xeric Torripsamments
Hackwood	Fine-loamy, mixed Pachic Cryoborolls
Kerrfield	Coarse-loamy, mixed, mesic Durinodic Xeric Haplocambids
Langslet	Fine, montmorillonitic, frigid Xeric Aquicambids
Lonely	Fine-loamy, mixed, frigid Xeric Haplocambids
Madeline	Clayey, montmorillonitic, frigid Lithic Argixerolls
McConnel	Sandy-skeletal, mixed, mesic Xeric Haplocambids
Nevador	Fine-loamy, mixed, mesic Durinodic Xeric Haplargids
Norad	Fine-silty, mixed, mesic Xeric Haplargids
Olac	Loamy-skeletal, mixed, mesic Lithic Xeric Haplargids
Oreanna	Fine-loamy over sandy or sandy-skeletal, mixed, mesic Typic Haplocambids
Outerkirk	Coarse-loamy, mixed, mesic Durinodic Haplocalcids
Ozamis	Fine-loamy, mixed, mesic Fluvaquentic Endoaquolls
Pernty	Loamy-skeletal, mixed, frigid Lithic Argixerolls
Raz	Loamy, mixed, frigid, shallow Xeric Haplodurids
Rio King	Coarse-loamy, mixed, mesic Aridic Haploxerolls
Robson	Clayey-skeletal, montmorillonitic, frigid Lithic Xeric Haplargids
Skedaddle	Loamy-skeletal, mixed, nonacid, mesic Lithic Xeric Torriorthents
Spangenburg	Fine, montmorillonitic, mesic Xeric Paleargids
Tumtum	Loamy, mixed, mesic, shallow Typic Argidurids

Table 3.3. Input data integrated in developing soil prediction models and their properties.

Variables	Data source	Resolution (Scale)	Type of Data
Landsat TM (bands 1, 2, 3, 4, 5, and 7)	USGS (2005)	30 m	Continuous
NDVI	Derived from the Landsat image (Rouse et al., 1973)	30 m	Continuous
SAVI	Derived from the Landsat image (Huete, 1988)	30 m	Continuous
Tasseled Cap Transformation (Brightness, Greenness, and Wetness)	Derived from the Landsat image (Kauth and Thomas, 1976)	30 m	Continuous
Terrain attributes (elevation, slope, aspect, and surface, profile and plan curvatures)	Derived from the DEM (USGS-EROS Data Center, 1999) using AcrGIS	10 m	Continuous
Landform classification	Derived from the DEM (Jenness, 2005)	30 m	10 classes
Solar radiation (in WH/m^2) (Diffuse, direct, and Globe radiation)	Drived from the DEM (Fu and Rich, 1999)	30 m	Continuous
Geology	USGS (2003)	1:500,000	15 classes
Historic vegetation	Tobalske (2002)	1:100,000	18 classes
Landfire vegetation	USFS (2006)	30 m	34 classes
Mean annual precipitation	USDA-NRCS (1999)	1: 200,000	14 classes
Ecological habitat	Clarke and Bryce (1997)	1:250,000	9 classes
*Distance from streams	Created using multi-ring buffer in ArcGIS	1:24,000	7 classes

*Distance from streams = (<300, 300-900, 900-1800, 1800-2700, 2700-3600, 3600-4500, and 4500-6000m)

Table 3.4. Producer's, user's, overall accuracy and Kappa coefficient for predicted soil map units in the reference area.

SMU	Producer's Accuracy			User's Accuracy		
	Predicted	Total	%	Predicted	Total	%
4	3208	3482	92.13	3208	3705	86.59
5	94	171	54.97	94	116	81.03
21	414	437	94.74	414	425	97.41
22	699	720	97.08	699	723	96.68
23	532	573	92.84	532	556	95.68
24	6414	6868	93.39	6414	6697	95.77
36	83	106	78.30	83	152	54.61
40	439	466	94.21	439	527	83.30
45	610	614	99.35	610	630	96.83
69	1668	1830	91.15	1668	1783	93.55
70	564	575	98.09	564	574	98.26
72	707	731	96.72	707	970	72.89
73	5363	5583	96.06	5363	5556	96.53
76	468	592	79.05	468	578	80.97
77	723	814	88.82	723	868	83.29
78	1030	1166	88.34	1030	1272	80.97
86	1792	2100	85.33	1792	2103	85.21
93	1238	1374	90.10	1238	1309	94.58
96	296	349	84.81	296	346	85.55
103	684	754	90.72	684	732	93.44
131	3094	3395	91.13	3094	3322	93.14
152	478	526	90.87	478	508	94.09

Table 3.4. Producer's, user's, overall accuracy and Kappa coefficient for predicted soil map units in the reference area (Continued).

SMU	Producer's Accuracy			User's Accuracy		
	Predicted	Total	%	Predicted	Total	%
165	45	45	100.00	45	49	91.84
178	464	493	94.12	464	493	94.12
182	235	258	91.09	235	239	98.33
192	2115	2211	95.66	2115	2220	95.27
216	1037	1203	86.20	1037	1207	85.92
235	20	20	100.00	20	21	95.24
245	91	99	91.92	91	92	98.91
248	154	185	83.24	154	190	81.05
249	151	168	89.88	151	230	65.65
250	14	18	77.78	14	16	87.50
251	698	785	88.92	698	740	94.32
256	113	118	95.76	113	115	98.26
259	4608	4762	96.77	4608	4635	99.42
272	185	207	89.37	185	200	92.50
282	277	323	85.76	277	318	87.11
283	26	31	83.87	26	44	59.09
288	423	478	88.49	423	444	95.27
291	247	259	95.37	247	263	93.92
300	881	944	93.33	881	1037	84.96
301	1341	1434	93.51	1341	1374	97.60
302	357	370	96.49	357	366	97.54
312	265	335	79.10	265	329	80.55
334	536	682	78.59	536	582	92.10
Total	44881	48656	92.24	44881	48656	92.24
Average			89.85			89.04
Kappa						0.9172

Table 3.5. Producer's, user's, overall accuracy and Kappa coefficient for predicted soil map units in the unmapped area using field data.

SMU	Producer's accuracy			User's accuracy		
	Predicted	Total	%	Predicted	Total	%
21	577	550	95.32	619	550	88.85
23	5635	5531	98.15	5537	5531	99.89
24	12058	11815	97.98	12007	11815	98.40
45	830	808	97.35	880	808	91.82
69	1078	1061	98.42	1144	1061	92.74
70	48	48	100.00	53	48	90.57
72	1447	1313	90.74	1352	1313	97.12
73	3103	2956	95.26	3037	2956	97.33
77	661	628	95.01	689	628	91.15
93	1086	1052	96.87	1129	1052	93.18
131	439	428	97.49	459	428	93.25
152	12	12	100.00	26	12	46.15
192	62	54	87.10	58	54	93.10
216	1273	1260	98.98	1291	1260	97.60
235	351	347	98.86	349	347	99.43
245	450	445	98.89	453	445	98.23
251	30	30	100.00	37	30	81.08
259	1299	1299	100.00	1300	1299	99.92
272	827	810	97.94	842	810	96.20
283	10	10	100.00	12	10	83.33
288	1	1	100.00	3	1	33.33
Total	31277	30458	97.38	31277	30458	97.38
Average			97.35			88.70
Kappa						0.9673

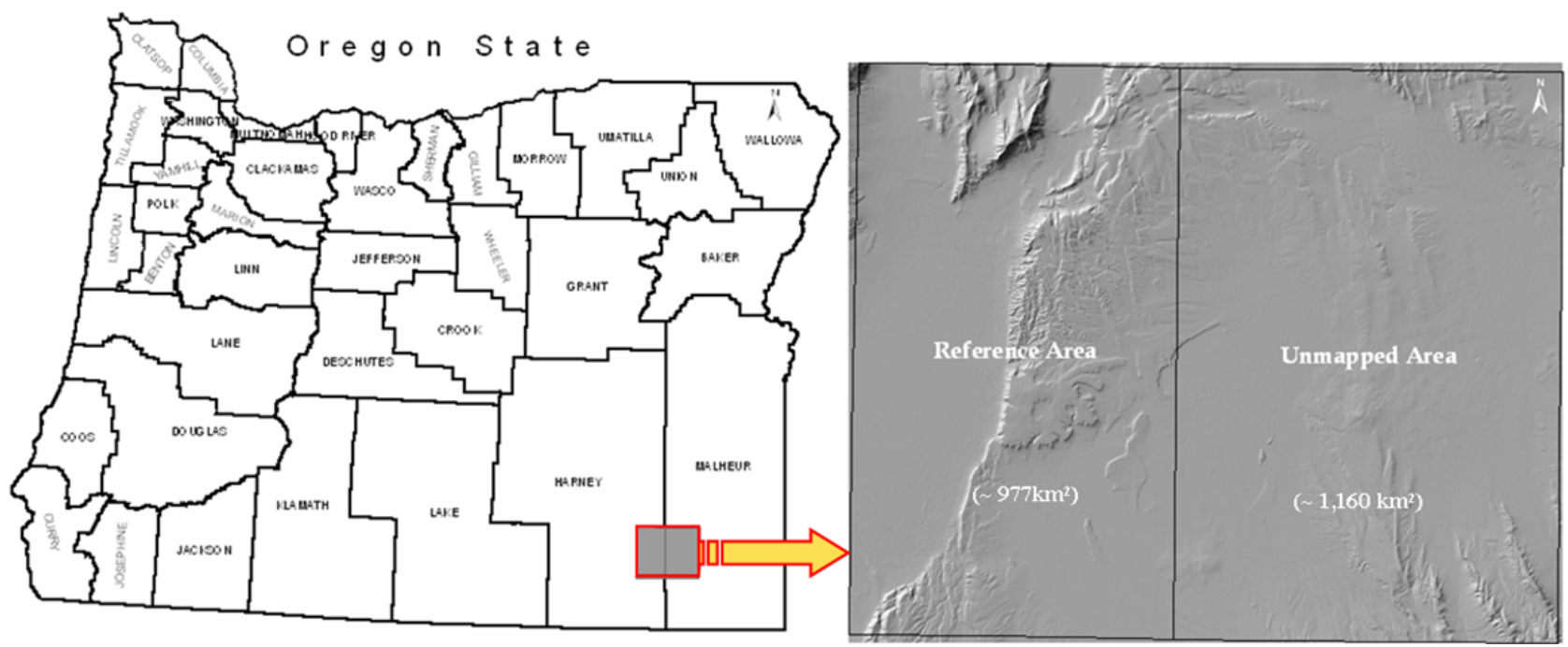

Figure 3.1. Study areas in Harney and Malheur Counties, Oregon.

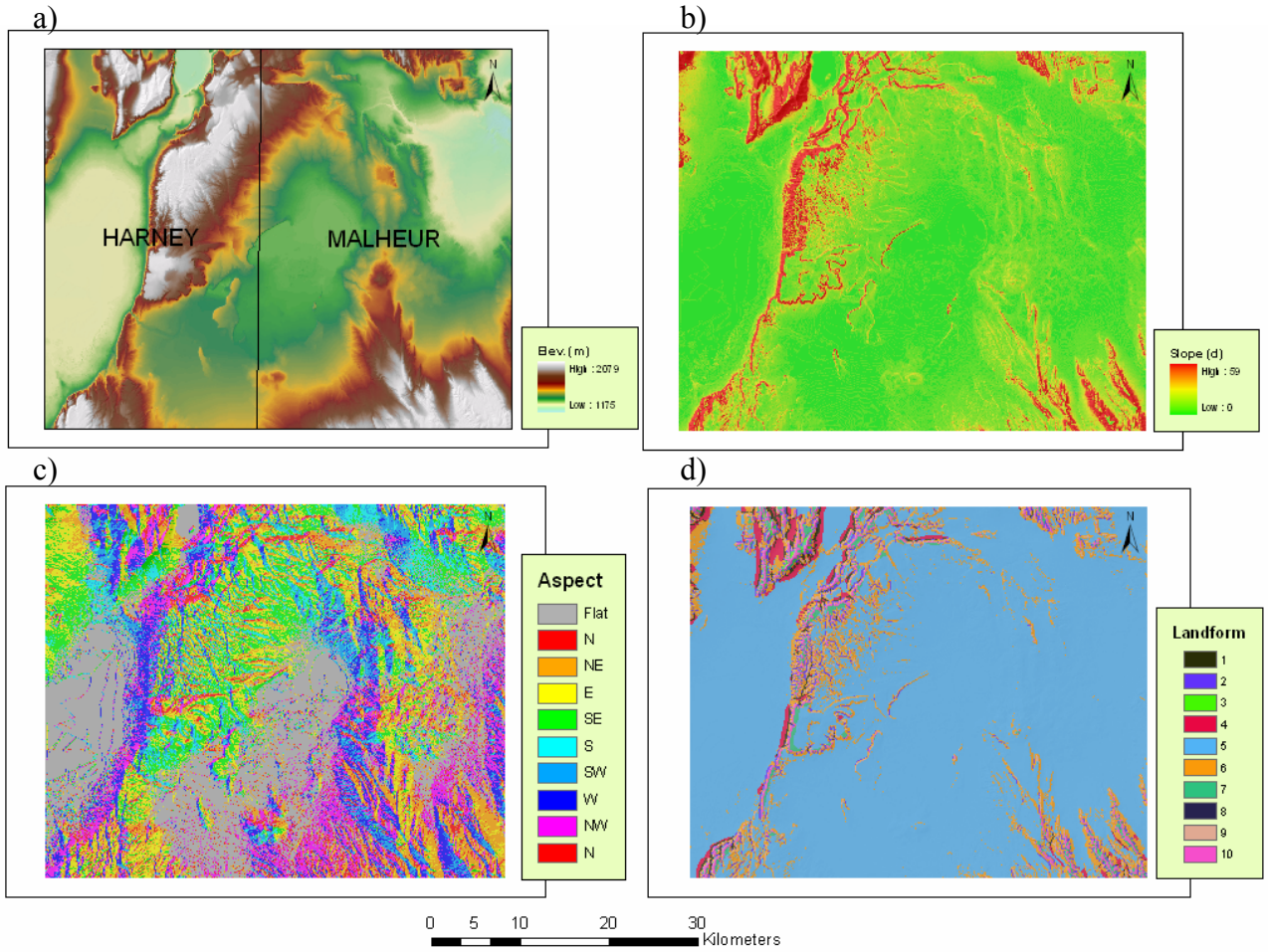

Figure 3.2. Terrain attributes developed from the digital elevation model (DEM): a) Elevation, b) Slope, c) Aspect, and d) Classified landforms.

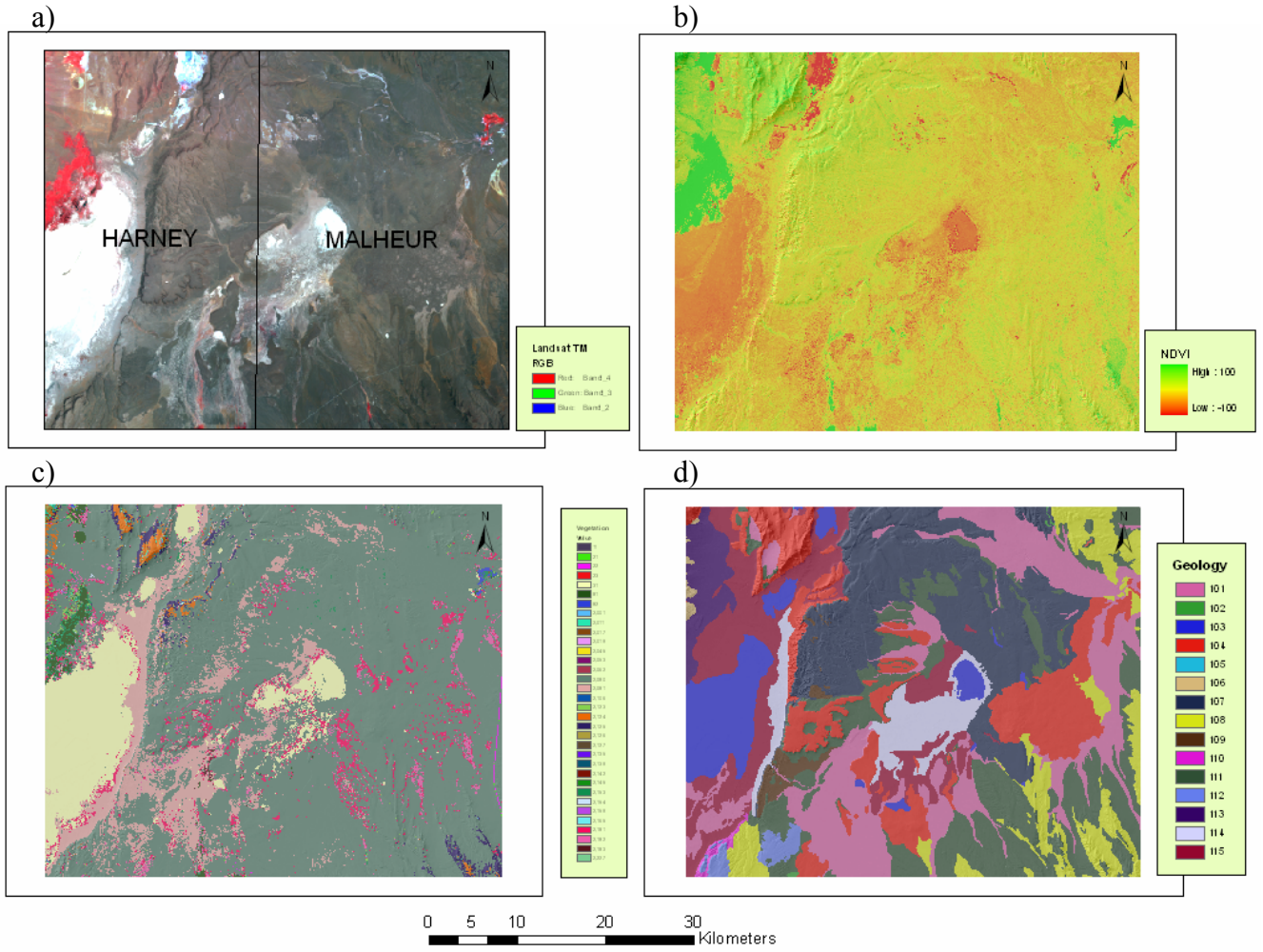

Figure 3.3. Other environmental: a) False color composite of the Landsat TM images, b) NDVI, c) Vegetation, and d) Geology.

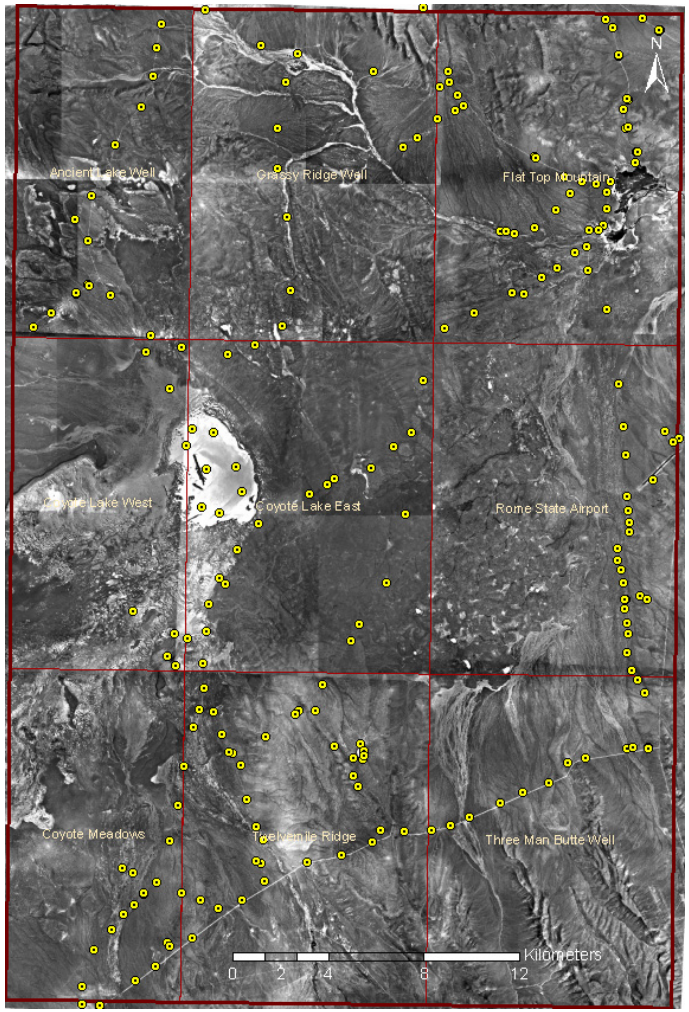

Figure 3.4. Quadrangles and sampling point distribution over the DOQ of the unmapped area in Malheur County.

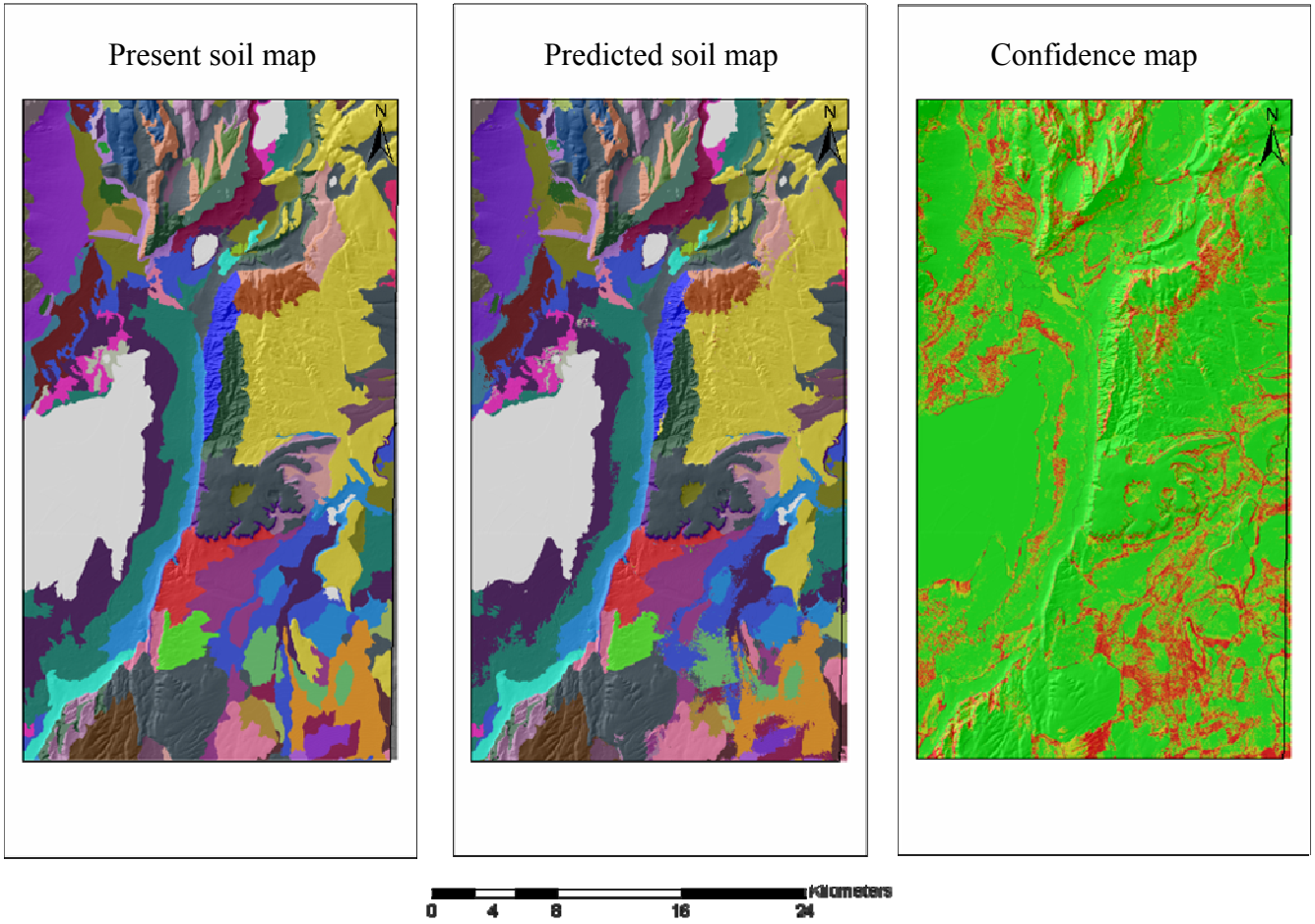

Figure 3.5. A comparison between present and predicted soil maps of the reference area in Harney County and their prediction confidence.

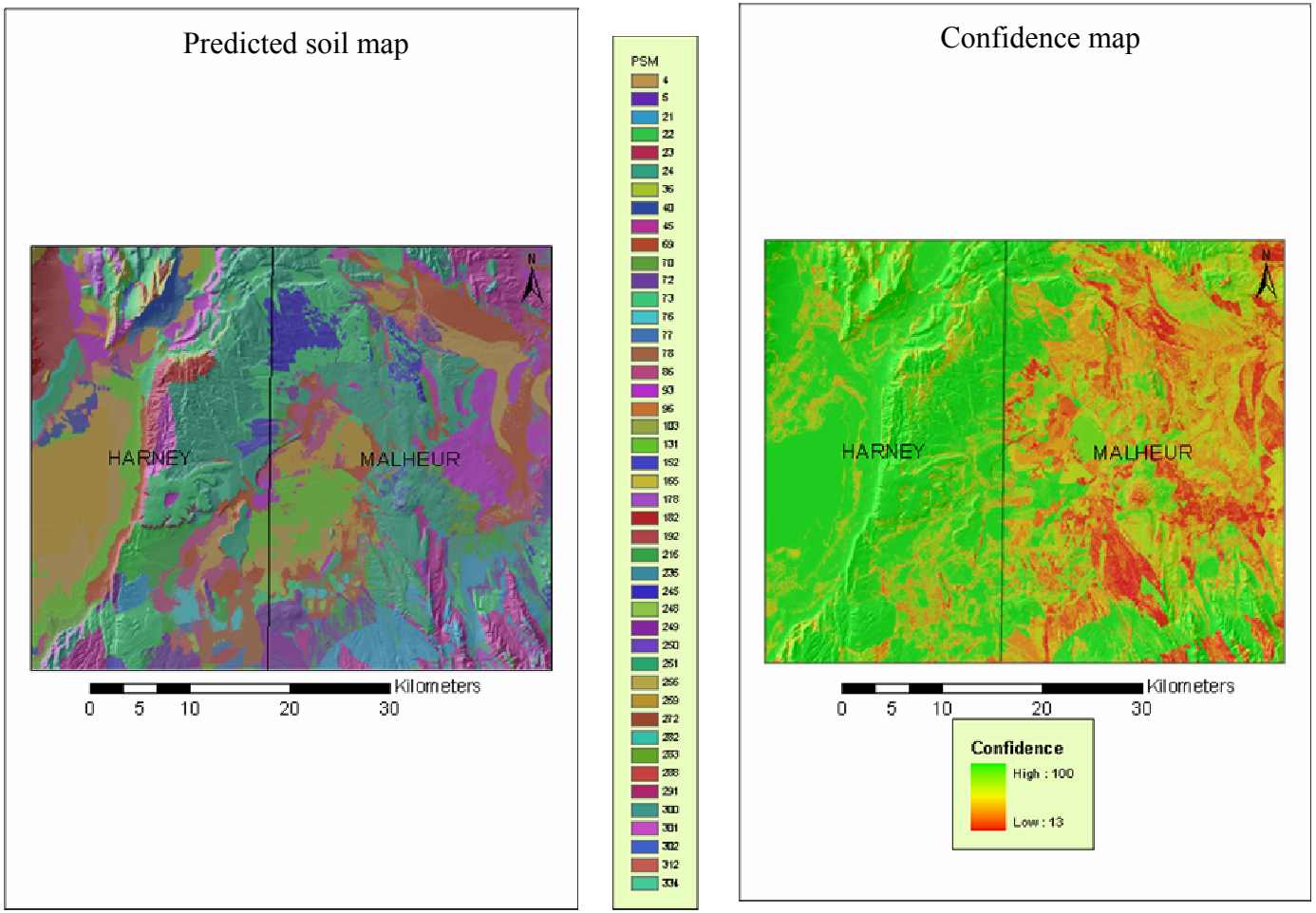

Figure 3.6. Predicted soil map for both the reference and the unmapped areas generated from extrapolating soil-landscape model derived from the reference map and its prediction confidence.

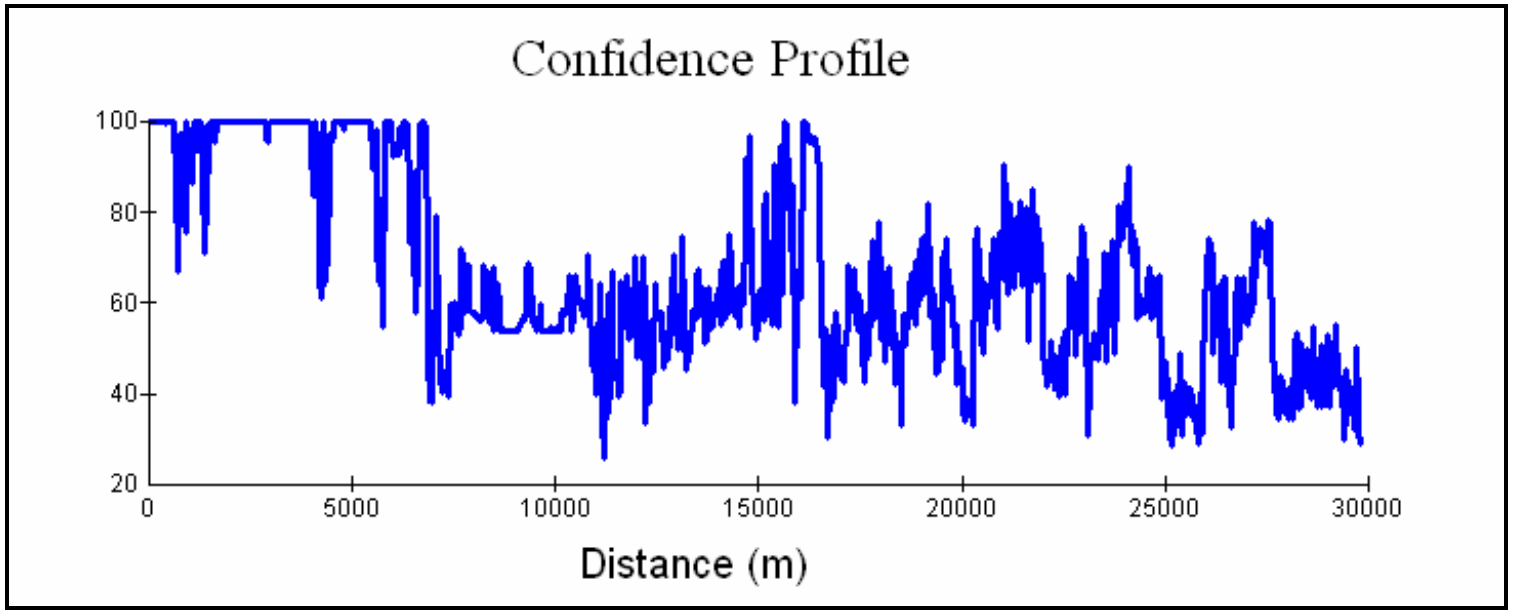

Figure 3.7. Relationship between prediction confidence of SMUs and distance from reference area.

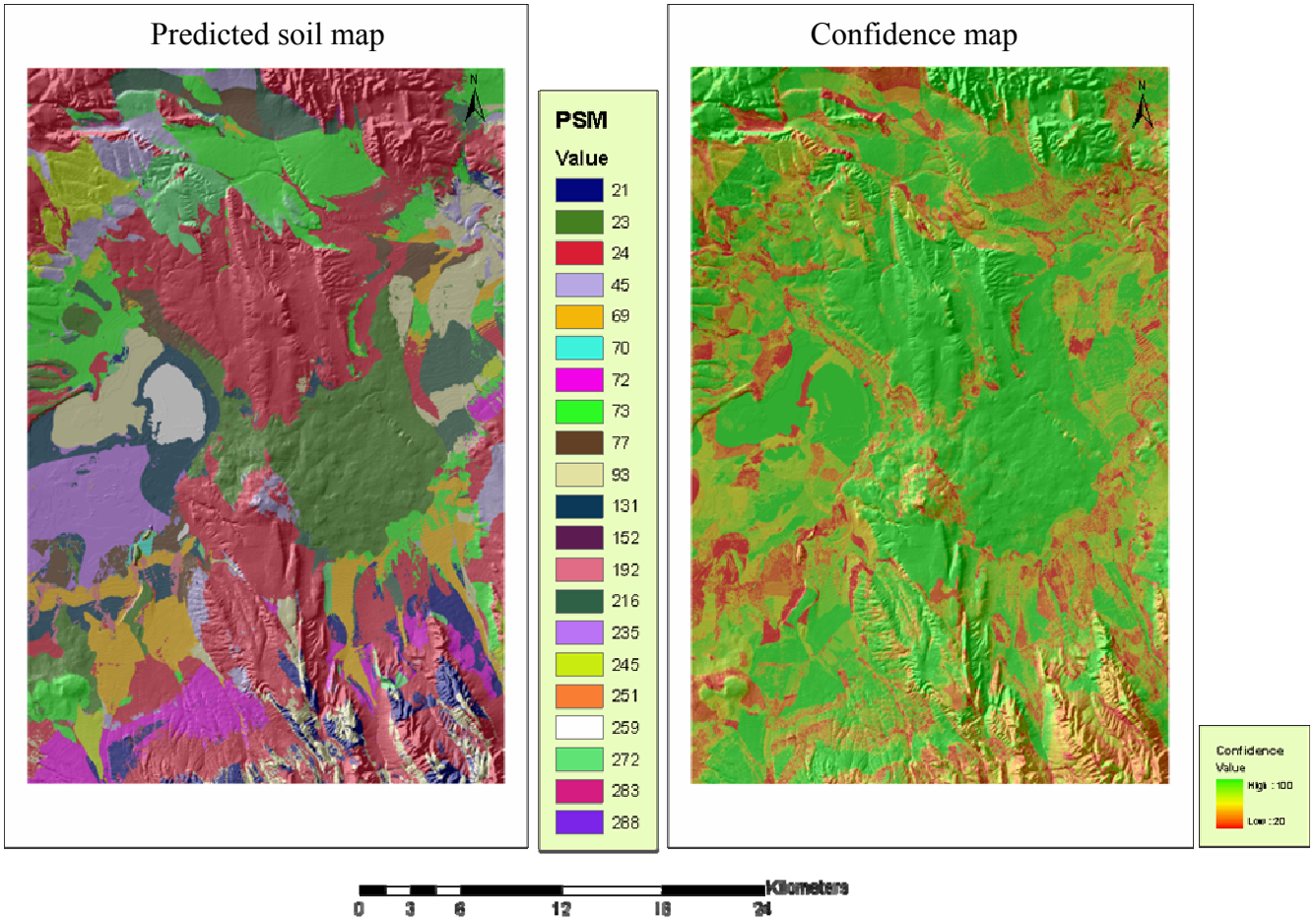

Figure 3.8. Predicted soil map for the unmapped area in Malheur County developed from field data and its prediction confidence.

CHAPTER 4

Application of Remote Sensing Data and Decision Tree Analysis to Mapping Salt-affected Soils over Large Areas

Abdelhamid A. Elnaggar and Jay S. Noller

Prepared for submission to:

International Journal of Remote Sensing

Abstract

This study deals with the problem of mapping soil salinity over large areas in arid and semi-arid environments. Remote sensing data and decision tree analysis (DTA) in conjunction with field data were integrated in this work to generate soil salinity maps of the study area in Malheur County, Oregon. Salinity developed in the study area is mainly either due to water logging or to the introduction of irrigation to some areas. A significant correlation was found between electrical conductivity (EC) values and surface elevation, bands 1, 2, 3 and 4 of the Landsat TM image, and brightness and wetness indices. Salt-affected areas were indicated by their high spectral reflectance and they were easily discriminated from the remote sensing data. However, remote sensing data failed to distinguish between the different classes of soil salinity. The prediction accuracy of non-saline soils (EC<4 dSm^{-1}) mapped by classifying the Landsat images was 97%; but it was 60% for saline soils (EC>4 dSm^{-1}), with an overall accuracy of about 95%. On the other hand, the five classes of soil salinity were successfully predicted using DTA with and overall accuracy of about 99%. Moreover, the calculated area of salt-affected soil was overestimated when mapped using remote sensing data compared to that predicted by using DTA. DTA proved to be a promising approach for mapping soil salinity in the study area in more productive and accurate ways compared to only using remote sensing data.

Keywords: soil salinity, salinity mapping, digital soil mapping, remote sensing, geostatistics, classification tree

4.1. Introduction

Soil salinization is a major land-degradation problem in arid and semi-arid environments and, wherever irrigation systems are introduced (Moreau, 1996; Dwivedi and Sreenivas, 1998; Khan et al., 2001). Mapping techniques that can be used to inventory and monitor soil salinization over large areas in more efficient, time-effective and less expensive ways are required for precision agriculture and sustaining soil productivity in many parts of the world. Predictive mapping techniques, such as linear and multiple regression, geostatistics (i.e., Kriging and CoKriging), fuzzy logic, neural network, and classification and regression trees (Burrough, 1986; Hansen et al., 1996; McBratney et al., 2003; Qi and Zhu, 2003; Scull et al., 2005) have been used to develop soil and natural resource maps. Each of these techniques provides optimal results under certain circumstances. Geostatistics, for example, yield significant results when data are normally distributed and stationary (mean and variance do not vary significantly in space); where significant deviations from normality and stationarity arise, the analysis becomes problematic (Olea, 1999; Pozdnyakova and Zhang, 1999). This normality issue is difficult to constrain at smaller scales, especially when values of environmental parameters and soil properties dramatically change from one location to another across the soilscape. It is just such a case in mapping soil salinity, where EC values change significantly between salt-affected and normal soils over relatively short distances within large areas. Using geostatistics in this case will result in significant errors and the predicted values will depart significantly from the original. Mapping salt-affected soils in the field is difficult as they are interspersed with normal soils and form no contiguous pattern (Sethi et al., 2006).

Geostatistical techniques such as cokriging could benefit from the availability of secondary data in developing prediction maps, but all of the data have to be numerical and normally distributed, not nominal or categorical data. Thus, valuable data such as geology and vegetation which either have a significant influence on salt accumulation or influenced by soil salinity, could not be used in producing predictive soil maps. This issue is in addition to the requirement of geostatistical techniques to large amounts of field data in order to obtain optimal results.

Decision-tree analysis (DTA), on the other hand, is a predictive mapping technique that can be used in developing soil salinity maps over large areas. It is a non-parametric or distribution-free statistical method, which means data are not required to fit a normal distribution curve. It can be used with ordinal as well as categorical attributes. Moreover, it requires no assumptions about the data and provides interpretable prediction rules that can be extrapolated to similar areas (Venables and Ripley, 1994; Hansen et al., 1996; Huang et al., 2002).

Remote-sensing data have been used successfully in mapping soil salinity for decades (Singh et al., 1977; Manchanda, 1984; Sharma and Bhargawa, 1988; Csillag et al., 1993; Joshi and Sahai, 1993; Moreau, 1996; Khan et al., 2001; Spies and Woodgate, 2005; Sethi et al., 2006). The principle behind this success is based on the dramatic effects that soil salinity has on soil physical, chemical and biological properties. The quantities and changes in soil properties can be monitored using remote sensing.

The objectives of this study are to demonstrate a combined method involving remote sensing data and DTA in developing soil salinity maps. A comparative study is

performed to measure efficiency gains that are expected to support land managers with the required information for future land management practices in Southern Oregon.

4.2. Materials and Methods

4.2.1. Site description

Saline soils and intergrades are abundant in the study area, located in Malheur County, Oregon (Fig. 4.1). For this experiment, 1160 km^2 area was chosen, with a surface elevation of 1175 to 1771m ($\bar{X}$=1297m) above sea level and slopes that range between 0 to 57^o ($\bar{X}$= 2^o). Aridisol is the dominant soil order in the area as the soil moisture regime is aridic and the mean annual precipitation (MAP) varies from 178mm in low areas to 330mm on high elevations ($\bar{X}$= 254mm). Minimum annual temperature is about 0^oC and the maximum annual temperature is about 17^oC.

Soils in the area are developed on Tertiary basalt and andesite, tuffaceous sedimentary rocks, lacustrine deposits and fluvial sedimentary rocks (Clarke and Bryce, 1997). Prevalent landforms in the area are alluvial fans, fan remnants, basins, flood plains, pediments, and playas. Vegetation varies from native vegetation to agricultural pasture land (Kagan and Caicco, 1992). Dominant native types of vegetation in the area are Wyoming big sagebrush (*Artemisia tridentata wyomingensis*), Shadscale saltbush (*Atriplex confertifolia*), greasewood (*Sarcobatus vermiculatus*), black sagebrush (*Artemisia nova*), basin big sagebrush (*Artemisia tridentata tridentata*), and low sagebrush (*Artemisia arbuscula arbuscula*).

Most soils in the area are well drained and depth to the water table is far away from the soil surface except for areas close to agriculture pasture land and the internal

playas. Agricultural lands are pump-irrigated and irrigation water flows to nearby low-lying areas, resulting in higher water tables. Coyote Lake, in addition to many smaller playas, is an internal-draining playa (alkali flats or sabkha) that is periodically water-logged during the winter and spring seasons, and dries out during the summer. Playas are bare and shallow depressions with a high content of soluble salts and high alkalinity.

4.2.2. Data sources and description

A wide variety of data were used in this work including: IFSAR data (Fig. 4.2), remote-sensing data, geology, vegetation, and precipitation. Description of the data layers and their sources is presented in Table 4.1 and a more detailed description of the attribute-values for each layer is given in Table 4.2. Spatial data were represented using the raster data model in ArcGIS which is helpful in data modeling and data manipulation. Data layers were resized to have a spatial resolution of 30m, which represent most of the data used in this study.

4.2.3. Soil samples and analysis

About 210 surface soil samples (nominally15cm depth) were collected from the study area during the months of July and August (dry season) 2006, where salt efflorescence reaches its maximum (M. Keller, 2006, personal comm.). Samples were collected using dirt road transects and stratified random sampling methods, depending on landscape complexity and representative areas in the Landsat images (Moreau, 1996; De Gruitjer, 2000). Samples were air dried, crushed and sieved to pass through a 2mm

sieve. Electrical Conductivity (EC) was measured in the soil-saturation extract in deciSiemans per meter (dSm^{-1}) according to Richards (1954) (Table 4.3). Soil reaction (pH) was also measured in the soil paste. No data were available about water table depth and water salinity content for this remote area.

4.2.4. Mapping methods

Two methods were used in this paper to develop soil salinity maps: remote sensing data (Landsat images) and DTA. Results were obtained by each method, combined, and then compared.

4.2.4.1. Remote sensing

The study area covers parts of two Landsat TM images acquired on August 17, 2005 (Fig. 4.3). The images were mosaiked and subsetted to cover the area of interest. Salt-affected soils are usually poorly vegetated areas and stressed vegetation could be used as indirect sign for the presence of salinity. Two vegetation indices were therefore integrated in the analysis: Normalized Vegetation Index (NDVI) (Rouse et al., 1973) and Soil Adjusted Vegetation Index (SAVI) (Huete, 1988). Tassel Cap Transformation (TCT) indices (brightness, greenness, and wetness (Kauth and Thomas, 1976)) were used to distinguish areas with high spectral reflectance, green vegetated areas and soil and vegetation moisture. Band ratios such as b3/b1 and b5/b7 also were calculated and used to interpret some soil properties. Band ratio 3 to 1 is found to reflect iron content as reported by Rowan et al. (1977), whereas band ratio 5 to 7 is found to have a strong correlation with clay mineral content in poorly vegetated areas (Riaza et al., 2000).

4.2.4.2 Decision tree analysis

Several environmental variables were incorporated in developing soil-salinity prediction maps for the study area using DTA. DTA was carried out using the See5/C4.5 algorithm (Quinlan, 1993). See5/C4.5 is a system for automated knowledge acquisition for knowledge base and other artificial intelligence (AI) applications (Eklund et al., 1998). A constructed decision tree consists of nodes representing variables or attributes, branches representing attribute values, and leaves representing classes. Decision tree is built based on selecting the attribute that minimizes the amount of disorder in the sub-tree rooted at a given node.

Soil samples were classified according their EC values into five classes: (1) EC < 2 dsm^{-1} very low; (2) EC from 2 to 4 dsm^{-1} low; (3) EC from 4 to 8 dsm^{-1} moderate; (4) EC from 8 to 16 dsm^{-1} high; and (5) EC > 16 dsm^{-1} very high. Sampling points were buffered using a 300m (10 pixels) buffering distance on the GIS map to collect other local environmental data used in training the model. These locations were randomly sampled using the Classification and Regression Tree (CART) model (Earth Satellite Corporation, 2003). About 21,412 sampling points were used to train the model, whereas about 7,641 sampling points were used to validate the model. Training and validation data were boosted using 10 trials to enhance the prediction accuracy. Using this function results in creating a sequence of decision trees, where each subsequent tree attempts to fix the misclassification errors in the previous one. Each decision tree makes a prediction and the final prediction is a weighted vote of the predictions of all trees (Freund and Schapire, 1996). Also, the growing tree was pruned by 30% to reduce the over-fitting problem and increase the model efficiency (Eklund et al., 1998).

4.3. Results

4.3.1. Field observations

Field observations in the study area indicate the presence of salt accumulation in certain landforms across the landscape. Salts effloresced on the soil surface in the floodplain of Crooked Creek (close to the agriculture land). EC values were very high at that location and varied from 12.46 to 82.80dSm^{-1}. This high salt content is associated salt-tolerant vegetation (halophytes) such as salt grass (*Distichlis spicata var. stricta.*) and greasewood. Also, this location represents a low-leveled area and the ground water table was encountered at about 60cm. This result is supported by the significant correlation between EC values and surface elevation (Table 4.4).

Higher salt content was also observed in the playas; however, there was no salt accumulation on the playa surface compared to Crooked Creek floodplain. No vegetation is growing in these areas, soils are strongly compacted, and pH values are greater than 8.5.

4.3.2. Image analysis and visual interpretation

Salt-affected soils could easily be visually identified from the Landsat images using the false color composite (RGB 432) and brightness and wetness indices. The spectral reflectance curve (Fig. 4.4) shows that severely salt-affected soils have a high reflectance in the visual (bands 1, 2 and 3) and near infrared (band 4) parts of spectrum and relatively low reflectance in the mid-infrared parts of spectrum (bands 5 and 7). A significant correlation was found between the EC values and bands 1, 2, 3, and 4 of the

TM images (Table 4.4). Also, a significant correlation was found between the EC values and the brightness and wetness indices.

Landsat images were classified using the maximum likelihood supervised classifier in the ENVI program into 5 classes (1. Saline soil; 2. Agriculture land; 3. Inter-mountain basins big sage steppe; 4. Low sage brush steppe; and 5. Inter-mountain big sage brush shrubland) (Fig. 4.5). The output map was reclassified into two classes: saline (class 1) EC>4 dSm^{-1} and non-saline (classes 2, 3, 4, and 5) EC<4 dSm^{-1}. Prediction accuracy of salt-affected soils was about 60% and non-saline soils was about 97%, with an overall accuracy was about 95%. Classified saline area represents 6.67% of the total area, whereas the non-saline area represents 93.33%.

4.3.3. Decision tree and predicted soil salinity map

Decision tree created by the See5 program shows a high degree of confidence in the classification accuracy (Fig. 4.6). The overall accuracy of the decision tree produced without boosting was 98.40% and Kappa coefficient was 0.90 (Table 4.5). Producer's accuracy varied from 77.68 to 99.17% with a mean value of 87.61%, whereas the user's accuracy varied from 78.95 to 99.22% with a mean value of 85.82%. The prediction accuracy was enhanced by using 10 trials of boosting. The overall accuracy was 98.81% and Kappa coefficient was 0.92 (Table 4.6). Producer's accuracy varied from 78.75 to 99.2% with a mean value of 91.39%, whereas the user's accuracy varied from 71.05 to 99.59% with a mean value of 85.96%. The calculated area of saline soils represents 1.86% of the total area, whereas the non-saline area represents 97.40%.

4.4. Discussion

This study reveals that salt accumulates in the study area as a result of either water logging or the introduction of irrigation to neighboring areas. Salt-affected soils close to the agriculture area mostly originates from evaporation of the shallow water table (upward movement) in these low-lying locations. They could also be due to the weathering of salt-bearing minerals such as feldspars which are abundant in volcanic rocks such as basalt that dominates the area. Although the artificially irrigated area represents a very small percent of the total area, it should raise the concern about the development of salinity when soils in the area are introduced to irrigation in the future. Much research has been performed on the relationship between the development of secondary salinization and the introduction of irrigation to increase crop yield especially under arid and semi-arid environments. It has been found in most irrigated areas that the introduction of irrigation has resulted in a build up of secondary salinization and water logging (Dwivedi and Sreenivas, 1998; Furby et al., 1998; Khan et al., 2001; Sukchan and Yamamoto, 2002).

Results indicate that Landsat images of the study area well identify bare areas that have a high reflectance due to their high salinity content and/or salt-efflorescence on the soil surface. This result agrees with that obtained by Everitt et al. (1988) who reported that salt-affected soils with salt encrustation at the surface are, generally, smoother than non-saline surfaces and cause high reflectance in the visible and near-infrared bands. It was also noticed that some spotted areas observe high spectral reflectance due to the yellowish dust blown from the playa and deposited there,

resulting in misclassifying these locations as highly saline soils. Accordingly, classifying salt-affected soils based on spectral signatures could overestimate their areas.

Vegetation indices (NDVI, SAVI, and Greenness index) did not have a significant correlation with the EC values, which indicates that halophytes couldn't be used in identifying salt-affected areas under vegetation cover. This could also be due to the coarse resolution of the Landsat image (30m pixel size) and the smaller size of the salt-affected area. Landsat data could only distinguish between highly saline soils and normal or non-saline soils but salinity classes or degrees in between could not be discriminated. Similar results were obtained by Moreau (1996) and Sethi et al. (2006). It also found that the wetness index has a significant correlation with the EC values which could be due to the tendency of salt-affected soils to retain high moisture content.

The soil salinity map developed by DTA successfully predicts five classes of salinity levels, a significant increase of the standard remote-sensing methods. This could be due to ability of DTA to integrate other environmental variables that have significant influence on the development of secondary salinization. Elevation and slope, for example, are very important variables in predicting soil salinity. Secondary salinization mostly occurs in low-land areas, where groundwater frequently rises up through the soil profile in these locations (Eklund, 1998; Sethi et al., 2006). Therefore, it is important to identify these recharge locations in the study area using the DEM. Also, the accumulation of soil salinity is not only influenced by the morphology of the soil profile but also by the soil physical, chemical and biological properties (Szabolcs, 1987; Sethi et al., 2006). Bedrock geology and its chemical composition is another valuable variable that was integrated in the analysis and could result in enhancing the

prediction accuracy. Soils in the study area are developed on volcanic rocks, mostly basalt and andesite, which are rich in feldspars and salt-bearing minerals. Dominant vegetation is another variable that is directly influenced by higher soil salinity levels. This was also integrated in the analysis; however the vegetation map has a coarser scale that does not represent the vegetation types associated with soil salinity (halophytes), especially over smaller areas.

4.5. Conclusion

Current remote-sensing methodology used in mapping soil salinity could be significantly improved if Decision-tree analysis (DTA) is incorporated in such efforts. Water accumulated in low-lying areas and salt effloresces on the soil surface due to the upward movement of water and its evaporation, presenting a problem to most land uses. Remote-sensing data alone have been a nominally successful tool in mapping soil salinity over large areas, as it can distinguish between only highly salt-affected soils indicated by poor and sparse vegetation and high spectral reflectance and non-saline soils indicated by healthy vegetation. This is insufficient for modern soil salinity management with its finer classes of salinity. Moreover, salt-affected areas may be overestimated when mapped using only spectral signatures. Further study is recommended to determine if, as suggested by this study, introducing irrigation to the study area in Malheur County without an appropriate drainage system could result in severe salinity problems.

As demonstrated here, the development of a soil salinity map using DTA successfully distinguishes between the five broadly used classes of salinity in the study

area. DTA proved to be an efficient, useful approach for mapping soil salinity over large areas compared to traditional remote-sensing data. DTA incorporates several environmental variables that significantly influence the development of soil salinity and not only the spectral properties of the soil surface. Using this technique could significantly enhance the productivity and the accuracy of soil salinity mapping compared to conventional mapping methods especially in such remote inhospitable areas. Global predictive maps of soil salinity should now be closer to obtain.

Acknowledgments

We gratefully acknowledge Mark Keller, the soil survey project leader; Alina Rice the scientist; and Charlie Tackman the BLM scientist in the BLM Office in Vale, Oregon for their great support and help with the field work and accommodation. We thank Anne Nolin, associate professor in the department of Geosciences, for her great help with the digital image analysis. We would like to thank Will Austin the Central Laboratory Director for letting us use the laboratory for salinity analysis. We greatly thank Joan Sandeno for editing this manuscript. Also, we thank Elizabeth Cervelli and Nathan Goodson, students at Oregon State University for their great help with the field work and the Laboratory analysis.

References

Burrough, P.A. 1986. Principles of geographical information systems for land resources assessment. New York: Oxford Univ. Press, p. 193.

Clarke, S. E., Bryce, S. A. 1997. Hierarchical subdivisions of the Columbia Plateau and Blue Mountains Ecoregions, Oregon and Washington. Portland: U.S. Department of Agriculture-Forest Service General Technical Report PNW-GTR-395, 114 p.

Csillag, F., Pasztor, L., Biehl, L. L. 1993. Spectral band selection for the characterization of salinity status of soils. Remote Sens. Environ., 43:231-242.

De Gruitjer, J.J., 2000. Sampling for spatial inventory and monitoring of natural resources. Technical report, Alterrrarapport 070, Wageningen, Alterra, Green World Research.

Dwivedi, R. S., Sreenivas, K. 1998. Image transforms as a tool for the study of soil salinity and alkalinity dynamics. Int. J. Remote Sensing, 19(4):605-619.

Earth Satellite Corporation 2003. CART software user's guide. U.S. Geological Survey – National Land Cover Database (NLCD).

Eklund, D P. W., Kirkby, S. D., Salim, A. 1998. Data mining and soil salinity analysis. Int. j. geographical information science, 12(3):247-268.

Everitt, J., Escobar, D., Gerbermann, A. L., Alaniz, M. 1988. Detecting saline soils with video imagery. Photogrammetric Engineering and Remote Sensing, 54:1283–1287.

Freund, Y., Schapire, R. 1996. Experiments with a new boosting algorithm. *In* Machine Learning: Proceedings of the Thirteenth International Conference, July, 1996. San Mateo, California: Morgan Kaufmann.

Furby, S., Flavel, R., Sherrah, M., McFarlane, J. 1998. Mapping salinity in the upper south east catchment in South Australia. A report from the LWWRDC project, Mapping dryland salinity (CDM2). CSIRO Mathematical and Information Sciences Primary Industries South Australia. CMIS 98/104.

Hansen, M., Dubayah, R., DeFries, R. 1996. Decision trees: an alternative to traditional land cover classifiers. Int. J. Remote Sensing, 17(5):1075-1081.

Huang, C., Davis, L. S., Townshend, J. R. G. 2002. An assessment of support vector machines for land cover classification. Int. J. Remote Sensing, 23(4): 725-749.

Huete, A. 1988. A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment, 25:295-309.

Joshi, M. D., Sahai, B. 1993. Mapping salt-affected land in Saurashtra coast using Landsat satellite data. Int. J. Remote Sensing, 14(10):1919-1029.

Kagan, J., Caicco, S. 1992. Manual of Oregon actual vegetation. Idaho Cooperative Fish and Wildlife Research Unit, University of Idaho.

Kauth, R. J., Thomas, G. S. 1976. The tasseled cap: a graphic description of the spectral-temporal development of agricultural crops as seen by LANDSAT. Proceedings of the Symposium on Machine Processing of Remotely Sensed Data, Purdue University of West Lafayette, Indiana, p. 4B-41 to 4B-51.

Khan, N. M., Rastoskuev, V. V., Shalina, E. V., Sato, Y. 2001. Mapping salt-affected soils using remote sensing indicators - a simple approach with the use of GIS-IDRSI. Paper presented at the 22nd Asian Conference of Remote Sensing. November 5-9, Singapore.

Manchanda, M. L. 1984. Use of remote sensing techniques in the study of distribution of salt-affected soils in north-west India. Indian Soc. Soil Sci., 32:701-706.

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

Moreau, S.S. 1996. Application of remote sensing and GIS to the mapping of saline/sodic soils and evaluation of sodification risks in the province of Villarroel, Central Altiplano, Bolivia. Paper presented at the 4th International Symposium on High Mountain Remote Sensing Cartography, Karlstad - Kiruna - Troms, August 19-29.

NRCS 2007. InterMap IFSAR elevation data, Malheur County, Oregon. National Cartography Geospatial Center, Fort Worth, Texas.

Olea, R.A. 1999. Geostatistics for engineers and Earth scientists. Kluwer Academic Publishers, Boston.

Pozdnyakova, L., Zhang, R. 1999. Geostatistical analyses of soil salinity in a large field. Precision Agriculture, 1:153-65.

Qi, F., Zhu, A. X. 2003. Knowledge discovery from soil maps using inductive learning. Int. J. Geographical Information Science, 17(8):771–795.

Quinlan, J. R. 1993. C4.5: Programs for machine learning. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Riaza, A., Mediavilla, R., Santistieban, J. I. 2000. Mapping geological stages of climate-dependent iron and clay weathering alteration on lithologically uniform sedimentary units using Thematic Mapper imagery (Tertiary Duero Basin, Spain). Int. J. Remote Sensing, 21(5):937-950.

Richards, L. A. 1954. Diagnosis and improvement of saline and alkali soils. U.S. Dep. Agric. Handb. No. 60.

Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W. 1973. Monitoring vegetation systems in the Great Plains with ERTS, Third ERTS Symposium, NASA SP-351, 1:309-317.

Rowan, L. C., Goetz, A. F. H., Ashley, R. P. 1977. Discrimination of hydrothermally altered and unaltered rocks in the visible and the near infrared multispectral images. Geophysics, 42:522-535.

Scull, P., Franklin, J., Chadwick, O. A. 2005. The application of decision tree analysis to soil type prediction in a desert landscape. Ecological Modeling, 181:1–15.

Sethi, M., Dasog, G. S., Lieshout, A. V., Salimath, S. B. 2006. Salinity appraisal using IRS images in Shorapur Taluka, Upper Krishna Irrigation Project, Phase I, Gulbarga District, Karnataka, India. Int. J. Remote Sensing, 27(14):2917–2926.

Sharma, R. C., Bhargawa, G. P. 1988. Landsat imagery for mapping saline soils and wetlands in north-west India. Int. J. Remote Sensing, 9:69-84.

Singh, A. N., Kristof, S. J., Baumgardner, M. F. 1977. Delineating salt-affected soils in the Ganges Plain, India by digital analysis of Landsat data. The Laboratory of Applications of Remote Sensing. Purdue University, West Lafayette, Indiana.

Spies, B., Woodgate, P. 2005. Salinity mapping methods in the Australian context. Published by the Department of the Environment and Heritage, and Agriculture, Fisheries and Forestry.

Sukchan, S., Yamamoto, Y. 2002. Classification of salt affected areas using remote sensing and GIS. JIRCAS Working Report No. 30. http://www.jircas.affrc.go.jp/english/publication/working/30/30-01-02.pdf

Szabolcs, I. 1987. The global problems of salt-affected soils. Acta Agronomica Hungarica, 36(1-2):159-172.

USDA-NRCS 1999. Oregon annual precipitation. National Cartography and Geospatial Center. Fort Worth, TX.

USFS 2006. Landfire existing vegetation cover. USDA forest Service. Missoula, Montana.

USGS 2003. Spatial digital database for geologic map of Oregon. U.S. Department of Interior, U.S. Geological Survey.

USGS 2005. LANDSAT TM – Path: 42 Row: 30 for Scene: 5042030000519810 and Path: 42 Row: 31 for Scene: 5042031000519810. U.S. Geological Survey Center for Earth Resources Observation and Science (EROS). Sioux Falls, SD.

Venables, W. N., Ripley, B. D. 1994. Modern applied statistics with S-PLUS. Springer-Verlag, New York.

Table 4.1. Databases and their sources.

Variables	Data source	Resolution (Scale)	Data type
Landsat TM (7 bands)	USGS (2005)	30 m	Raster
Normalized Difference Vegetation Index (NDVI)	Derived from the Landsat image (Rouse et al., 1973)	30 m	Raster
Soil Adjusted Vegetation Index (SAVI)	Derived from the Landsat image (Huete, 1988)	30 m	Raster
Normalized Difference Salinity Index (NDSI)	Derived from the Landsat image (Khan et al., 2001)	30 m	Raster
Brightness, Greenness, and Wetness	Derived from the Landsat image (Kauth and Thomas, 1976)	30 m	Raster
Band ratios B3/b1 and b5/b7	Derived from the Landsat image	30 m	Raster
Terrain attributes (elevation, slope and aspect)	Derived from the IFSAR data (NRCS, 2007)	5 m	Raster
Geology	USGS (2003)	1:500,000	Vector
Landfire vegetation	USFS (2006)	30 m	Raster
Mean annual precipitation	USDA-NRCS (1999)	1:200,000	Vector
Distance from streams	Created using multi-ring buffer in ArcGIS	1:24,000	Vector
Ecological habitat	Clarke and Bryce (1997)	1:250,000	Vector

Table 4.2. Environmental variables and their properties.

Variable	Type	Number of classes	Value range
Landsat TM bands	Continuous	Continuous	0-255
NDVI	Continuous	Continuous	0-100
SAVI	Continuous	Continuous	0-100
Brightness	Continuous	Continuous	42-462
Greenness	Continuous	Continuous	-60-112
Wetness	Continuous	Continuous	-96-53
Elevation	Continuous	Continuous	1175-1771
Slope gradient	Continuous	Continuous	0-57
Slope aspect	Continuous	Continuous	-1-360
Landform	Discrete	10 classes	1-10
Geology	Discrete	9 classes	101-111
Historic vegetation	Discrete	10 classes	Wide range (1-45)
Landfire vegetation	Discrete	26 classes	Wide range (1-2227)
Precipitation	Discrete	4 classes	7, 9, 11 and 13
*Distance from streams	Discrete	1 classes	1-7
Habitat	Discrete	6 classes	1-6

*Distance from streams = (<300, 300-900, 900-1800, 1800-2700, 2700-3600, 3600-4500, and 4500-6000m)

Table 4.3. Soil samples and their XY coordinates, saturation percentage (SP), field capacity (FC), pH and EC values.

Sample No	Latitude (X)	Longitude (Y)	SP (%)	FC (%)	pH	EC (uS/m)	EC (dS/m)
1	427377.94	4732683.5	23	11.5	8.61	520	0.52
2	427408.01	4732738.4	22	11	7.95	440	0.44
3	425123.02	4733171.7	23	11.5	7.7	420	0.42
4	426066.03	4729825.9	22	11	6.5	430	0.43
5	425888.5	4729396.5	20	10	7.28	490	0.49
6	426422.18	4727613.1	17	8.5	7.38	360	0.36
8	423415.2	4726609.6	20	10	7.3	580	0.58
9	426927.68	4702650.1	14.6	7.3	7.29	630	0.63
10	426253.98	4702676.5	21.78	10.89	7.04	700	0.7
11	426017.04	4702615.3	20.76	10.38	6.68	290	0.29
12	424288.7	4702239.8	30.08	15.04	8.24	670	0.67
13	423567	4702049.6	27.44	13.72	8.22	750	0.75
14	421671.78	4700790.2	19.34	9.67	7.24	370	0.37
15	419418.24	4699755.5	21.51	10.75	6.72	260	0.26
16	418614.62	4699396.5	25	12.5	6.5	590	0.59
17	416709.02	4699157.9	24.8	12.4	6.65	180	0.18
18	415714.33	4699196.4	31.13	15.56	7.64	590	0.59
19	414747.16	4701045	24	12	7.32	460	0.46
20	414546.51	4701464.7	24	12	7.32	490	0.49
21	415354.86	4698688.2	23	11.5	7.5	450	0.45
22	414068.77	4698174.9	23	11.5	7.37	530	0.53
23	412616.4	4697856.8	24	12	7.47	290	0.29
24	410844.16	4697052.9	23	11.5	7.61	1170	1.17
25	410501.52	4697923.7	21	10.5	7.44	480	0.48

SP= Saturation Percentage
FC= Field Capacity= ½ SP
(A complete list of soil samples and their properties is presented in Table A.3.1)

Table 4.4. Correlation between EC values and numerical environmental variables.

Variable	EC
Elevation	-0.1805 *
Slope	-0.1378 ns
Band1	0.2498 **
Band2	0.2902 **
Band3	0.2562 **
Band4	0.3837 **
Band5	0.1283 ns
Band7	0.0581 ns
NDVI	-0.1445 ns
NDSI	0.1445 ns
SAVI	-0.1328 ns
Brightness	0.2570 **
Greenness	-0.0590 ns
Wetness	0.4537 **

Number of observations is 210
* Significant at confidence level of 95%
** Very significant at confidence level of 95%
ns Non significant

Table 4.5. Producer's, user's, overall accuracy and Kappa coefficient for predicted soil salinity classes developed be the See5 without boosting.

Salinity class	Class1	Class2	Class3	Class4	Class5	Total of rows	User's accuracy
Class1	6961	15	5	12	23	7016	99.22
Class2	27	61				88	69.32
Class3	8		30			38	78.95
Class4	16			380	2	398	95.48
Class5	7			7	87	101	86.14
Total of columns	7019	76	35	399	112	7641	
Producer's accuracy	99.17	80.26	85.71	95.24	77.68		98.40
Kappa							0.90

Table 4.6. Producer's, user's, overall accuracy and Kappa coefficient for predicted soil salinity classes developed be the See5 with 10 trails of boosting.

Salinity class	Class1	Class2	Class3	Class4	Class5	Total of rows	User's accuracy
Class1	6987	17	3	4	5	7016	99.59
Class2	25	63				88	71.59
Class3	11		27			38	71.05
Class4	13			380	5	398	95.48
Class5	7			1	93	101	92.08
Total of columns	7043	80	30	385	103	7641	
Producer's accuracy	99.20	78.75	90.00	98.70	90.29		98.81
Kappa							0.92

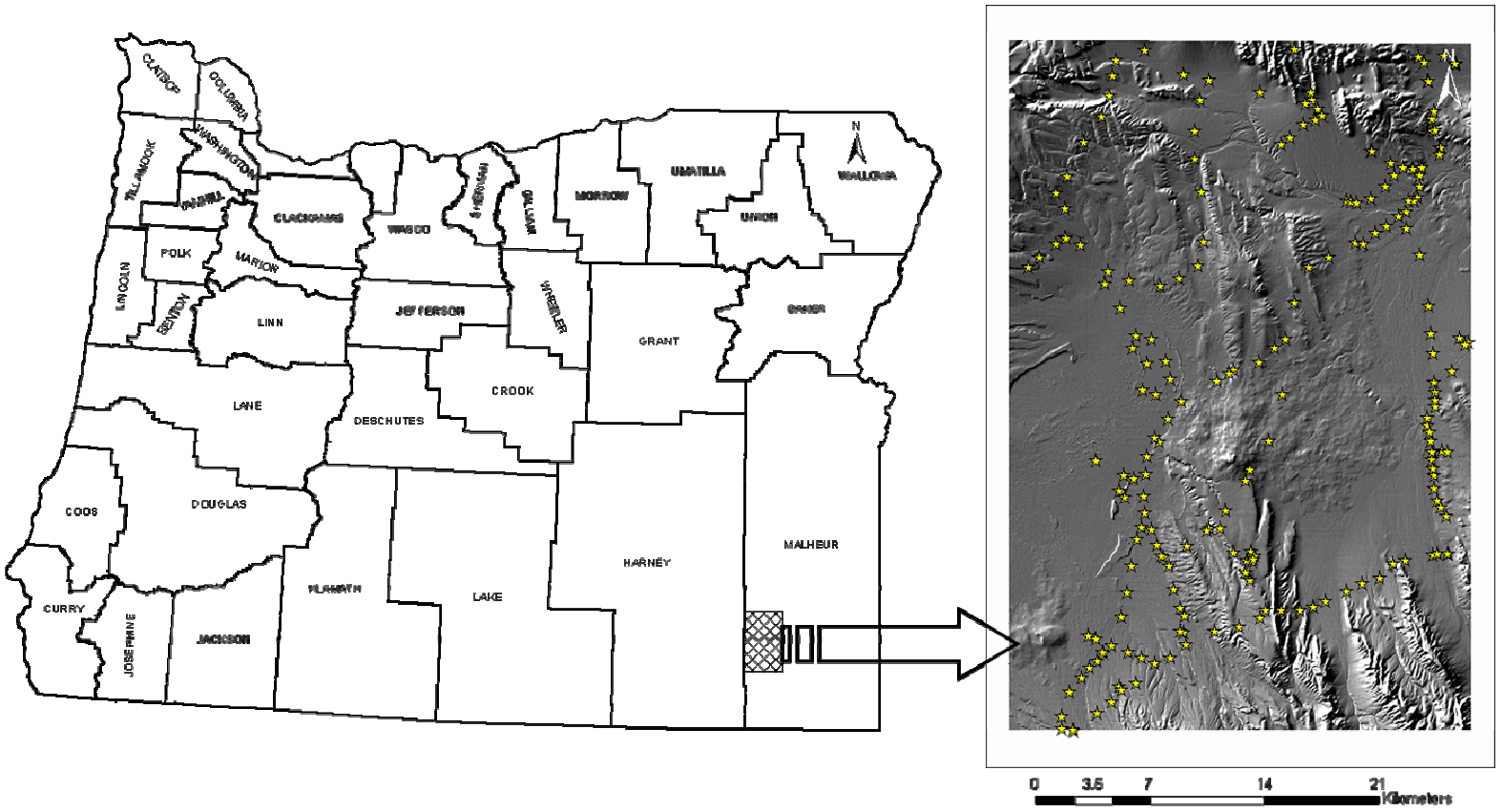

Figure 4.1. Study area and sampling points in Malheur County, Oregon.

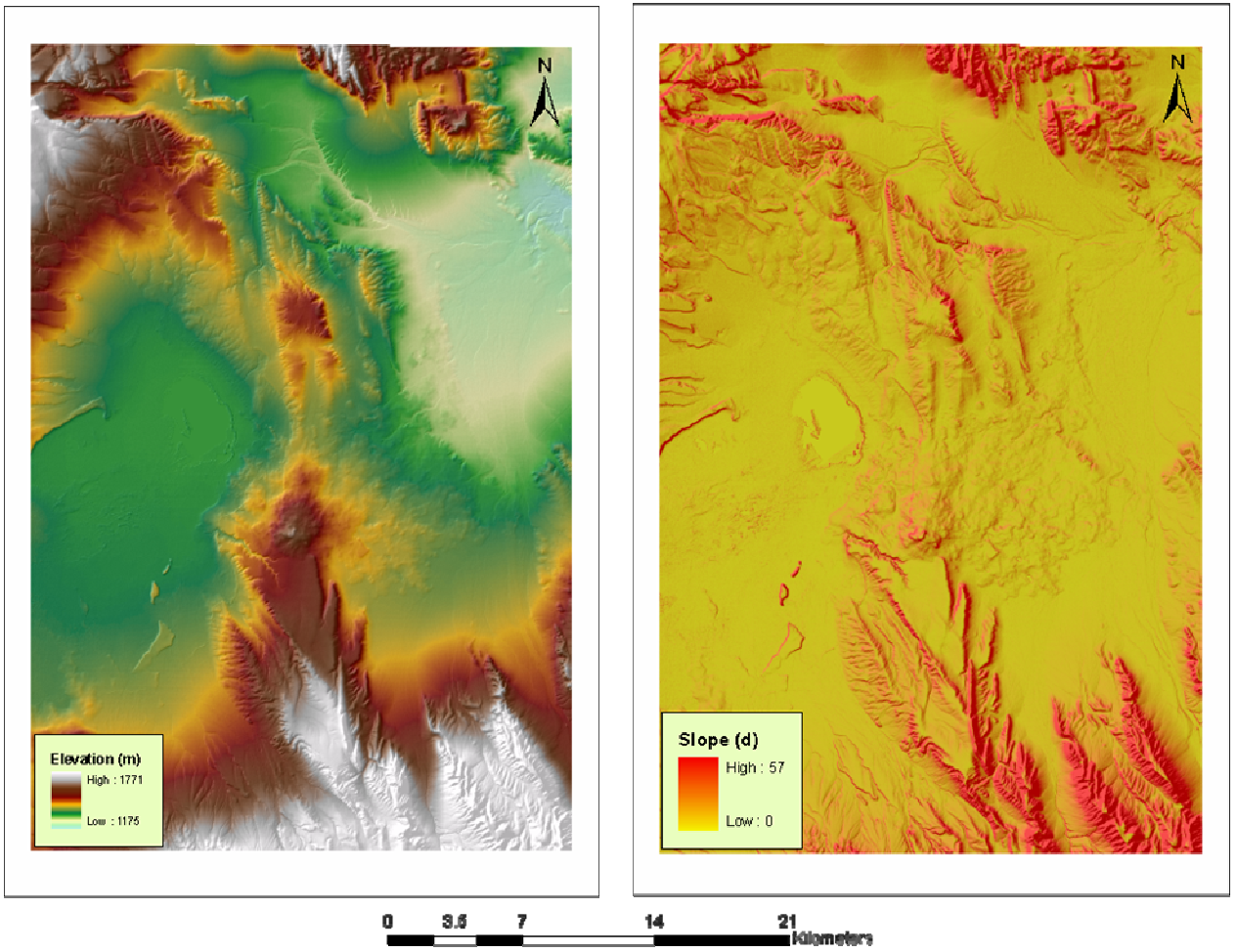

Figure 4.2. Digital terrain model (DTM) and slope gradient in the study area.

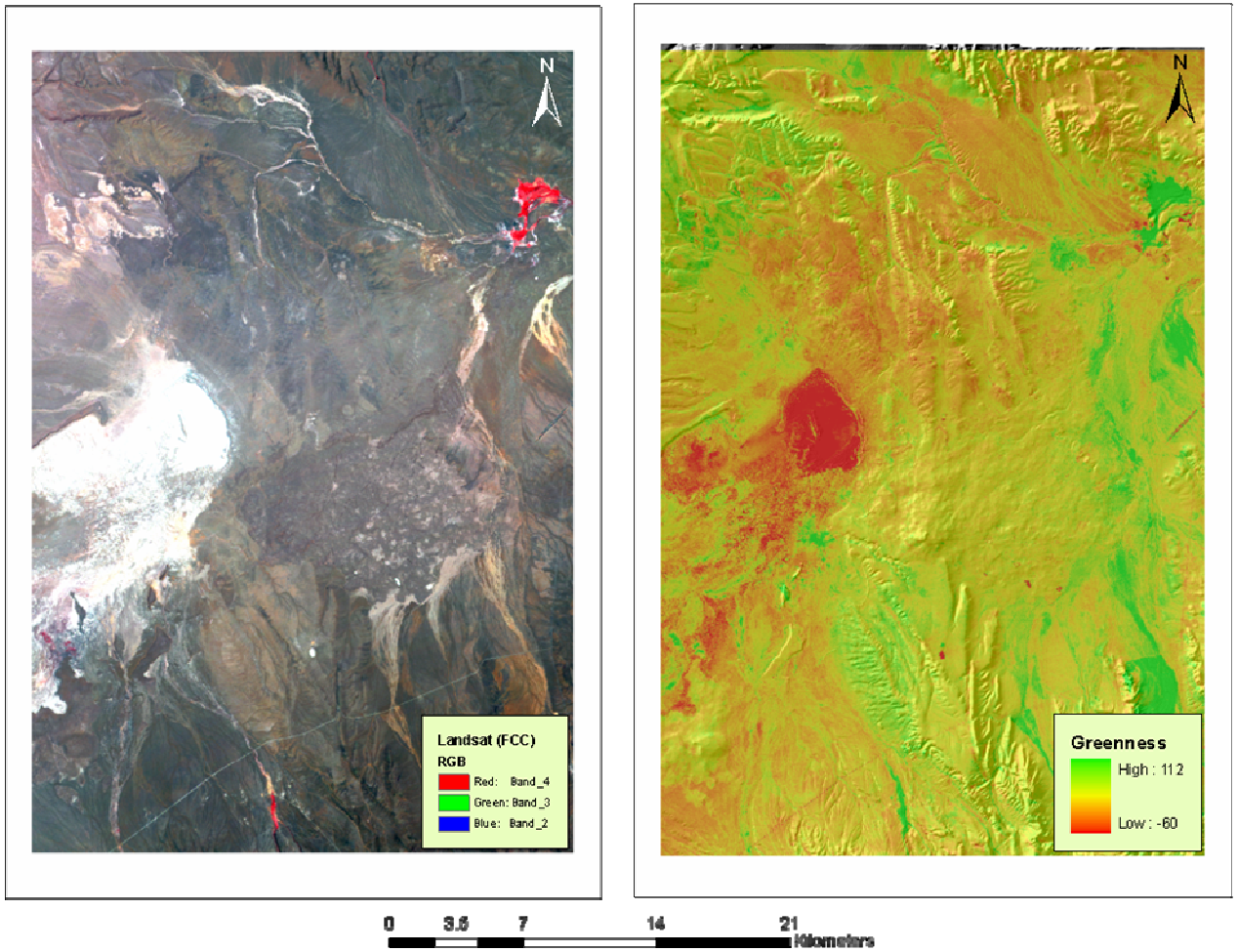

Figure 4.3. False color composite of the Landsat TM (RGB 432) and greenness index.

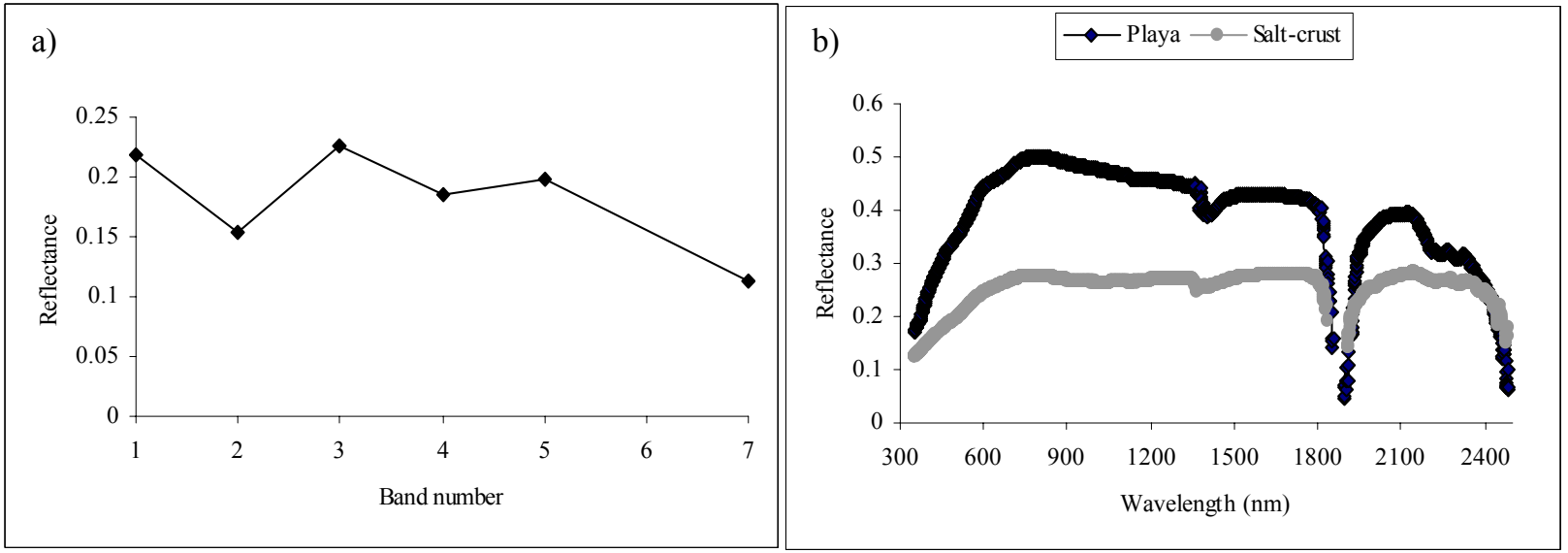

Figure 4.4. Spectral reflectance of salt-affected soils collected by using a) Landsat TM image and b) Spectroradiometer[2].

[2] Spectral properties of salt-affected soils in the study area were measured in the field almost at the same acquisition time of Landsat TM images (August 17, 2005). The spectral reflectance was measured using the FieldspecfiPro, manufactured by Analytical Spectral Devices of Boulder Colorado. The instrument has a field of view of 25 mm. It covers the spectral range from 350 to 2500nm with an average bandwidth of 1nm.

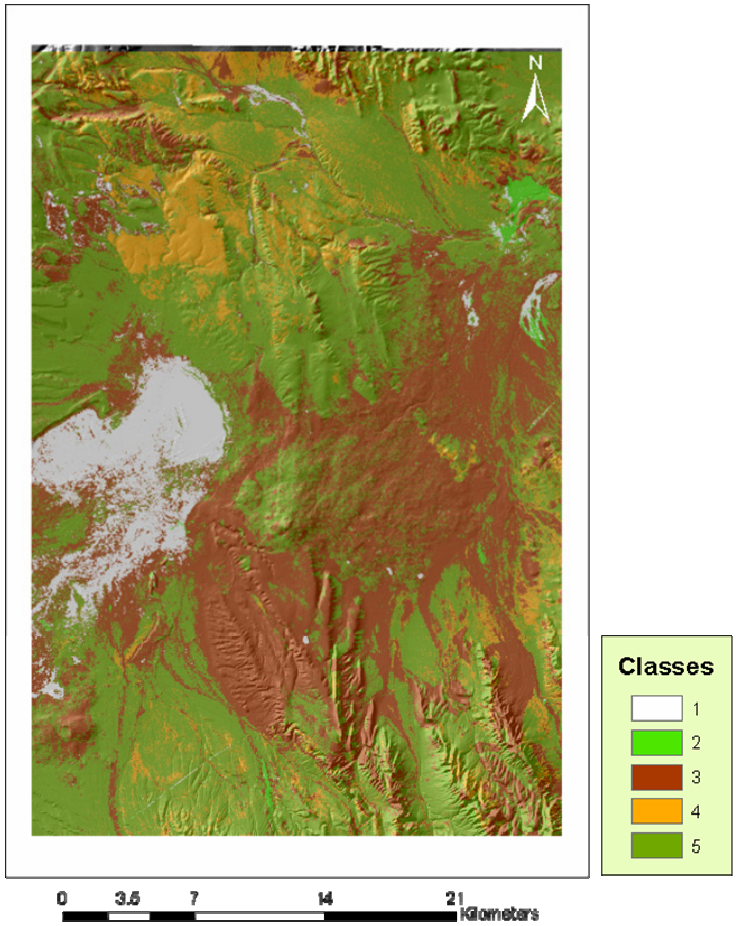

Figure 4.5. Supervised classification map of the Landsat TM image (1. Saline soil, 2. Agriculture land, 3. Inter-mountain basins big sage steppe, 4. Low sage brush steppe, and 5. Inter-mountain big sage brush shrubland).

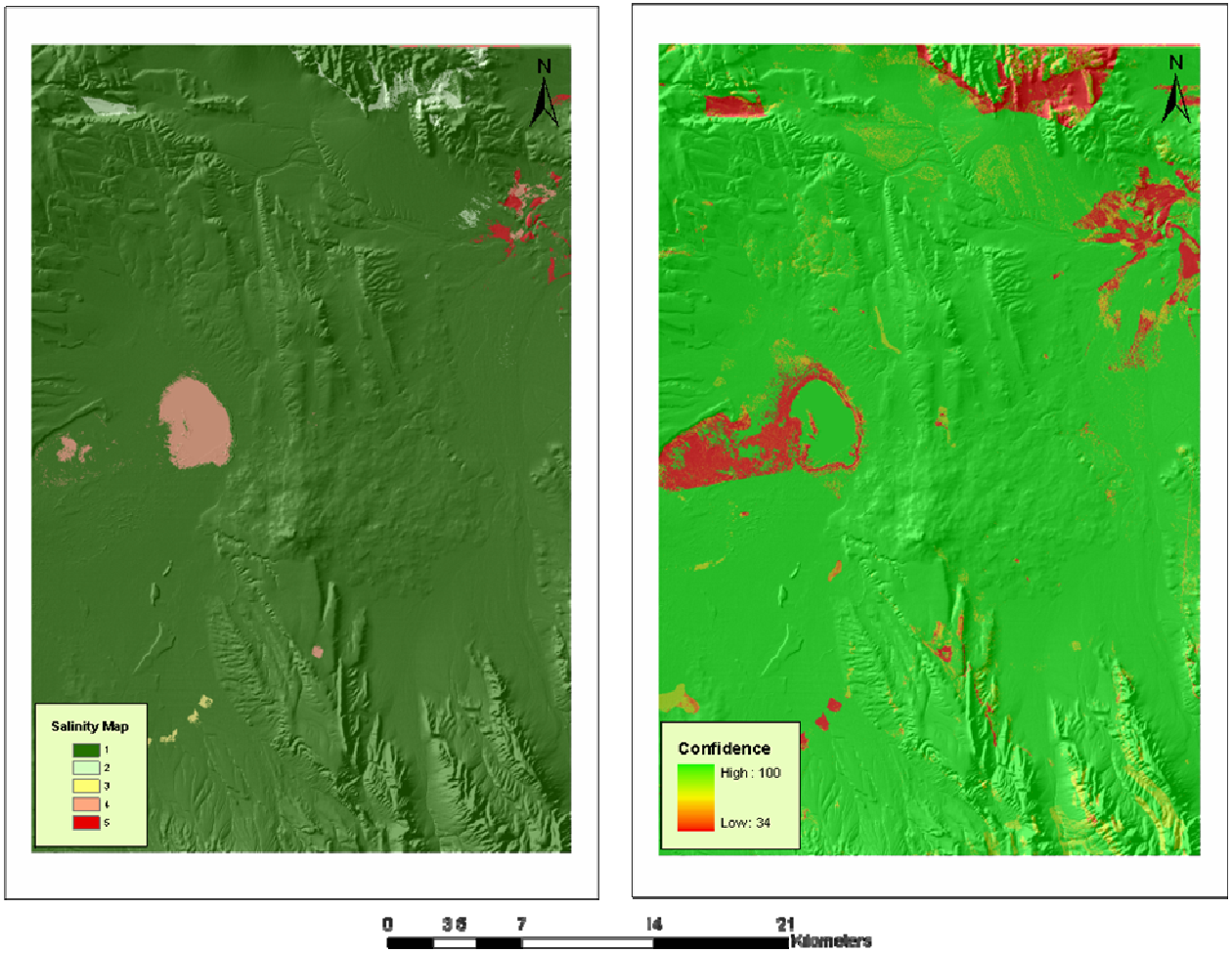

Figure 4.6. Predicted soil salinity map using decision tree analysis and its prediction confidence.

CHAPTER 5

General Conclusion

Soil survey maps represent one of the most significant sources of information used by land managers. Nowadays, demand is growing for accurate and multi-resolution soil data to sustain agriculture and the environment. Producing soil survey maps by traditional methods is an expensive, labor-intensive, and time-consuming process which is no longer supportable. Therefore, developing new techniques that can utilize the wealth of available data and technology in producing soil maps in more objective, effective, less-expensive, and faster ways should meet the necessity for enhancing soil sustainability and productivity in the future. This dissertation studies predictive soil mapping (PSM) techniques as a means to this end.

A soil-landscape model can successfully be extracted from old soil survey data using decision-tree analysis and environmental data. Such models were efficiently used in transferring the digitized soil maps of the study areas in Benton and Harney Counties into more objective digital maps. Assessing the consistency of the present soil map of the study area in Benton County resulted in obtaining valuable information. Common errors in conventional soil maps (inaccurate boundaries and inclusions) were revealed from the predicted soil maps. Prediction accuracy of the developed model is significantly influenced by the number of classes chosen to be predicted, the proportional area of each class, and the accuracy and scale of contributing environmental variables. Map units or taxonomic classes having a large areal extent

and/or are well identified by the environmental data incorporated in the model, were predicted with high accuracies. On the contrary, map units representing smaller areas or that were poorly identified from the data, were predicted with lower accuracies or entirely unpredicted.

Selecting the appropriate model to predict soil classes depends on the details that one wants to be retrieved from the generated soil map. Soil taxa maps (i.e., order, suborder, and great groups) would be useful in developing the preliminary soil maps at small scales, where detailed information is not necessarily required. On the other hand, predicting soil subgroups and soil map units could be very valuable at larger scale where detailed knowledge about soil characteristics is required. Also, selecting the environmental variables to be used in producing predictive soil maps varies from one environment to another depending on the prevalent environmental conditions under each environment. The decision tree analysis (DTA) approach shows high flexibility under a variety of environmental conditions.

Using DTA in developing predictive soil maps from field data and soil-surveyor knowledge provides significant results compared to extrapolating retrieved soil-landscapes models from reference areas. These results could be helpful in reducing the great amount of field data required in conventional soil mapping methods. DTA could significantly reduce the large amount of time consumed in developing soil maps in the traditional ways and facilitating their updates. Moreover, the DTA approach can help in achieving the most important goals in predictive soil mapping which are producing soil maps in more effective, objective, and less-expensive ways and providing information about accuracy of the developed maps which are not available in current soil maps.

Results revealed that certain points should be considered to enhance the outputs of using DTA in predictive soil mapping. First, assuming that available soil survey maps and environmental data of reference areas are accurate could be untrue for some areas. Soil maps and all other data integrated in predictive soil mapping techniques should be evaluated in advance for their accuracy. Second, extrapolating retrieved soil-landscape models is spatially limited to neighboring areas that have been developed under the same environmental conditions and have similar topography.

As a demonstration of the usefulness of the DTA approach to soil inventory, an issue common to semi and arid soilscapes was investigated. Field observations in the study area of Malheur County revealed a potential vulnerability of this soilscape to a severe salinity problem. Further study is recommended to determine if introduced irrigation would result in soil-salinity problems.

Use of conventional remote-sensing (RS) methods in mapping soil salinity in the study area only distinguished between severely salt-affected soils, indicated by poor and sparse vegetation and high spectral reflectance, and non-saline soils, indicated by healthy vegetation. Furthermore, salt-affected areas were overestimated when mapped using only spectral signatures of surface features. The obtained results from mapping soil-salinity using RS data are insufficient for modern soil salinity management with its five classes of salinity. This provides a good test of using decision tree analysis (DTA) in producing soil salinity maps. Resulting salinity prediction maps by DTA successfully and accurately distinguished between the five broadly used classes of salinity in the study area. DTA proved to be an efficient, useful approach for mapping soil salinity over large areas compared to traditional RS data, where several environmental variables

having significant impact on the development of soil salinity can be incorporated in the analysis and not only the spectral properties of the soil surface.

Bibliography

Bourgault, G., Journel, A. G., Rhoades, J. D., Corwin, D. L., Lesch, S. M. 1997. Geostatistical analysis of a soil data set. Adv. Agron., 58:241-292.

Brannon, G. R., Hajek, B. F. 2000. Update and recorrelation of soil surveys using GIS and statistical analysis. Soil Science Society of America Journal, 64(2):679–680.

Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J. 1984. Decision and regression trees. Wadsworth Belmont, CA.

Bui, E. N., Loughhead, A., Corner, R. 1999. Extracting soil-landscape rules from previous soil surveys. Aust. J. Soil Res., 37:495-508.

Burrough, P.A. 1986. Principles of geographical information systems for land resources assessment. New York: Oxford Univ. Press, p. 193.

Burrough, P.A. 1993. Soil variability revisited. Soils Fert., 56(5):529-562.

Burrough, P. A. 1997. Environmental modeling with geographic information systems. *In:* Innovations in GIS 4, Kemp, Z. (ed) London: Taylor & Francis, pp. 143-153.

Burrough, P. A., Beckett, P. H. T., Jarvis, M. G. 1971. The relation between cost and utility in soil survey. J. Soil Sci., 22:368-81.

Burrough, P. A., MacMillian, R. A., van Deusen, W. 1992. Fuzzy classification methods for determining land suitability from soil profile observations and topography. J. Soil Sci., 43:193-210.

Burrough, P. A., McDonnell, R. A. 1998. Principles of geographical information systems. Oxford Univ. Press, Oxford.

Clark, L. A., Pregibon, D. 1992. Tree-based models. *In*: Chambers, J. M., Hastie, T. J. (Eds.), Statistical Models. S. Wadsworth and Brooks, California, USA, p. 377-420.

Clarke, S. E., Bryce, S. A. 1997. Hierarchical subdivisions of the Columbia Plateau and Blue Mountains Ecoregions, Oregon and Washington. Portland: U.S. Department of Agriculture-Forest Service General Technical Report PNW-GTR-395, p. 114.

Cohen, J. 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20:37-46.

Congalton, R. G. 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of the Environment. 37:35-46.

Cook, S. E., Corner, R. J., Groves, P. R., Grealish, G. J. 1996. Use of gamma radiometric data for soil mapping. Aust. J. Soil Res., 34:183-194.

Csillag, F., Pasztor, L., Biehl, L. L. 1993. Spectral band selection for the characterization of salinity status of soils. Remote Sens. Environ., 43:231-242.

De Gruitjer, J.J., 2000. Sampling for spatial inventory and monitoring of natural resources. Technical report, Alterrrarapport 070, Wageningen, Alterra, Green World Research.

De Gruijter, J. J., Walvoort, D. J. J., van Gaans, P. F. M. 1997. Continuous soil maps - a fuzzy set approach to bridge the gap between aggregation levels of process and distribution models. Geoderma, 77:169-195.

Dobos, E., Carre, F., Hengl, T., Reuter, H. I., Toth, G. 2006. Digital soil mapping as a support to prediction of functional map. Digital Soil Mapping Working Group of the European Bureau Network. University of Miskolc, Miskolc-Egyetemváros, Hungary

Dwivedi, R. S., Sreenivas, K. 1998. Image transforms as a tool for the study of soil salinity and alkalinity dynamics. Int. J. Remote Sensing, 19(4):605-619.

Earth Satellite Corporation 2003. CART software user's guide. U.S. Geological Survey – National Land Cover Database (NLCD).

Ehlschlaeger, C. R., Goodchild, M. F. 1994. Dealing with uncertainty in categorical coverage maps: Defining, visualizing, and managing errors. In: Proceedings of the Workshop on Geographical Information Systems at the Conference on Information and Knowledge Management, Gaithersburg, Maryland: 86–91.

Eklund, D P. W., Kirkby, S. D., Salim, A. 1998. Data mining and soil salinity analysis. Int. j. geographical information science, 12(3):247-268.

Elnaggar, A. A., Noller, J. S. 2007. Assessing the consistency of conventional soil survey data: Switching from Conventional to Digital Soil Mapping Techniques. In press.

Everitt, J., Escobar, D., Gerbermann, A. L., Alaniz, M. 1988. Detecting saline soils with video imagery. Photogrammetric Engineering and Remote Sensing, 54:1283–1287.

Franklin, J. 1995: Predictive vegetation mapping: geographic modeling of biospatial patterns in relation to environmental gradients. Progress in Physical Geography, 19:474–490.

Freund, Y., Schapire, R. 1996. Experiments with a new boosting algorithm. *In:* Machine Learning: Proceedings of the Thirteenth International Conference, July, 1996. San Mateo, California: Morgan Kaufmann.

Friedl, M. A., Brodley, C. E. 1997. Decision tree classification of land cover from remotely sensed data. Remote Sensing of the Environment. 61:399-409.

Friedl, M. A., Brodley, C. E., Strahler, A. H. 1999. Maximizing land cover classification accuracies produced by decision trees at continental to global scales. IEEE Transactions on Geoscience and Remote Sensing. 37(2):969-977.

Fu, P., Rich, P. M. 1999. The solar analyst 1.0 user manual. Helios Environmental Modeling Institute, LLC. http://www.hemisoft.com.

Furby, S., Flavel, R., Sherrah, M., McFarlane, J. 1998. Mapping salinity in the upper south east catchment in South Australia. A report from the LWWRDC project, Mapping dryland salinity (CDM2). CSIRO Mathematical and Information Sciences Primary Industries South Australia. CMIS 98/104.

Gessler, P. E., Moore, I. D., McKenzie, N. J., Ryan, P. J. 1995. Soil landscape modelling and spatial prediction of soil attributes. Int. J. Geogr. Info. Syst., 9:421–432.

Goodchild, M. F. 1992. Geographical data modeling. Computers and Geosciences, 18:401-408.

Hansen, M., Dubayah, R., DeFries, R. 1996. Decision trees: an alternative to traditional land cover classifiers. Int. J. Remote Sensing, 17(5):1075-1081.

Haugerud, R. A., Harding, D. J., Johnson, S. Y., Harless, J. L., Weaver, C. S., Sherrod, B. L. 2003. High-resolution LiDAR topography of the Puget Lowland, Washington-A bonanza for Earth Science. GSA Today, 13(6):4-10.

Henderson, B. L., Bui, E. N., Moran, C. J., Simon, D. A. P. 2005. Australia-wide predictions of soil properties using decision trees. Geoderma, 124: 383-398.

Herold, N. D., Koeln, G., Cunnigham, D. 2003. Mapping impervious surfaces and forest canopy using classification and regression tree (CART) analysis. ASPRS 2003 Annual Conference Proceedings. Anchorage, Alaska.

Huang, C., Davis, L. S., Townshend, J. R. G. 2002. An assessment of support vector machines for land cover classification. Int. J. Remote Sensing, 23(4): 725-749.

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

Huete, A. 1988. A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment, 25:295-309.

International Atomic Energy Agency (IAEA) 1991. Airborne gamma-ray spectrometer surveying. Technical Series 323, IAEA.

Jenness, J. 2005. Topographic position index (tpi_jen.avx) extension for ArcView 3.x. Jenness Enterprises. http://www.jennessent.com/arcview/tpi.htm

Jenny, H., 1941, Factors in soil formation: New York, McGraw-Hill, 281 p.

Jensen, J. R. 1996. Introductory digital image processing: A remote sensing perspective. 2nd Ed., Prentice Hall, Inc., Upper Saddle River, New Jersey, USA.

Joshi, M. D., Sahai, B. 1993. Mapping salt-affected land in Saurashtra coast using Landsat satellite data. Int. J. Remote Sensing, 14(10):1919-1029.

Kagan, J., Caicco, S. 1992. Manual of Oregon actual vegetation. Idaho Cooperative Fish and Wildlife Research Unit, University of Idaho.

Kauth, R. J., Thomas, G. S. 1976. The tasseled cap: a graphic description of the spectral-temporal development of agricultural crops as seen by LANDSAT. Proceedings of the Symposium on Machine Processing of Remotely Sensed Data, Purdue University of West Lafayette, Indiana, p. 4B-41 to 4B-51.

Khan, N. M., Rastoskuev, V. V., Shalina, E. V., Sato, Y. 2001. Mapping salt-affected soils using remote sensing indicators - a simple approach with the use of GIS-IDRSI. Paper presented at the 22nd Asian Conference of Remote Sensing. November 5-9, Singapore.

MacMillan, R. A., Pettapiece, W. W., Brierley, J. A. 2005. An expert system for allocating soil to landforms through the application of soil survey tacit knowledge. Canadian Journal of Soil Science, 85(1):103-112.

Manchanda, M. L. 1984. Use of remote sensing techniques in the study of distribution of salt-affected soils in north-west India. Indian Soc. Soil Sci., 32:701-706.

Mark, D. M., Csillag, F. 1989. The nature of boundaries on area-class maps. Cartographica, 21:65-78.

Mather, P. M. 2004. Computer processing of remotely-sensed images – an introduction. 3rd Ed., John Wiley and Sons Ltd., Chichester, England.

McBratney, A. B., Lagacherie, P. 2004. Global workshop on digital soil mapping. Montpellier, France.

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

McBratney, A. B., Odeh, I. O. A. 1997. Applications of fuzzy sets in soil science: fuzzy logic, fuzzy measurements and fuzzy decisions. Geoderma, 77:85-113.

McBratney, A. B., Odeh, I. O. A., Bishop, T. F. A., Dunbar, M. S., Shatar, T. M. 2000. An overview of pedometric techniques for use in soil survey. Geoderma, 87:293-327.

Mckenzie, N. J., Ryan, P. J. 1999. Spatial prediction of soil properties using environmental correlation. Geoderma, 89:67-94.

McKenzie, N. J., Gessler, P. E., Ryan, P. J., O'Connel, D. 2000. The role of terrain analysis in soil mapping. *In:* Wilson, J.P., Gallant, J.C. (Eds.), Terrain Analysis-Principles and Applications. Wiley, New York, p. 245-265.

Metternicht, G. I. 1998. Fuzzy classification of JERS-1 SAR data: an evaluation of its performance for soil salinity mapping. Ecological Modeling, 111:61-74.

Michie, D., Spiegelhalter, D. J., Taylor, C. C. 1994. Machine learning, neural and statistical classification. New York: Ellis Horwood, p. 289.

Minasny, B., McBratney, A. B. 2002. The neuro-m method for fitting neural network parametric pedotransfer functions. Soil Sci. Soc. Am. J., 66:352-361.

Moore, I. D., Gessler, P. E., Nielsen, G. A., Peterson, G. A. 1993. Soil attribute prediction using terrain analysis. Soil Sci. Soc. Am. J., 57:443-452.

Moran, C. J., Bui, E. N. 2002. Spatial data mining for enhanced soil map modeling. Int. J. Geographical Information Science, 16(6):533-549.

Moreau, S.S. 1996. Application of remote sensing and GIS to the mapping of saline/sodic soils and evaluation of sodification risks in the province of Villarroel, Central Altiplano, Bolivia. Paper presented at the 4th International Symposium on High Mountain Remote Sensing Cartography, Karlstad - Kiruna - Troms, August 19-29.

Mosaic Mapping Systems Inc., 2001. A white paper on LiDAR mapping. Ottawa, ON, Canada. ftp://ftp-fc.sc.egov.usda.gov/NCGC/products/elevation/lidar-applications-whitepaper.pdf

Narayanan, R. M., Hirsave, P. P. 2001: Soil moisture estimation models using SIR-C SAR data: a case study in New Hampshire, USA. Remote Sens. Environ., 75:385–96.

NRCS 2007. InterMap IFSAR elevation data, Malheur County, Oregon. National Cartography Geospatial Center, Fort Worth, Texas.

Odeh, I. O. A., McBratney, A. B., Chittleborough, D. J., 1992. Fuzzy-c-means and kriging for mapping soil as a continuous system. Soil Sci. Soc. Am. J., 56:1848-1854

Odeh, I. O. A., McBratney, A. B., Chittleborough, D. J. 1995. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. Geoderma, 67:215-225.

Olea, R.A. 1999. Geostatistics for engineers and Earth scientists. Kluwer Academic Publishers, Boston.

Pozdnyakova, L., Zhang, R. 1999. Geostatistical analyses of soil salinity in a large field. Precision Agriculture, 1:153-65.

Prima, O. D. A., Echigo, A., Yokoyama, R., Yoshida, T. 2006. Supervised landform classification of Northeast Honshu from DEM-derived thematic maps. Geomorphology, (78):373–386.

Qi, F., Zhu, A. X. 2003. Knowledge discovery from soil maps using inductive learning. Int. J. Geographical Information Science, 17(8):771–795.

Quinlan, J. R. 1990. Learning Logical Definitions from Relations. Machine Learning, 5(3): 239-266.

Quinlan, J. R. 1993. C4.5: Programs for machine learning. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Quinlan, J. R. 2001. See5: An Informal Tutorial. http://www.rulequest.com.

Riaza, A., Mediavilla, R., Santistieban, J. I. 2000. Mapping geological stages of climate-dependent iron and clay weathering alteration on lithologically uniform sedimentary units using Thematic Mapper imagery (Tertiary Duero Basin, Spain). Int. J. Remote Sensing, 21(5):937-950.

Richards, L. A. 1954. Diagnosis and improvement of saline and alkali soils. U.S. Dep. Agric. Handb. No. 60.

Rossiter, D. G., 2001. Assessing thematic accuracy of area-class maps. Soil Science Division, ITC. Enschede, Netherlands.

Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W. 1973. Monitoring vegetation systems in the Great Plains with ERTS, Third ERTS Symposium, NASA SP-351, 1:309-317.

Rowan, L. C., Goetz, A. F. H., Ashley, R. P. 1977. Discrimination of hydrothermally altered and unaltered rocks in the visible and the near infrared multispectral images. Geophysics, 42:522-535.

Roy, G., Vallee, G., Jean, M. 1993. Lidar-inversion technique based on total integrated backscatter calibrated curves. Applied Optics, 32(33):6754-6763.

Scull, P., Franklin, J., Chadwick, O. A. 2005. The application of decision tree analysis to soil type prediction in a desert landscape. Ecological Modeling, 181:1–15.

Scull, P., Franklin, J., Chadwick, O. A., McArthur, D. 2003. Predictive soil mapping: a review. Progress in Physical Geography, 27(2):171–197.

Sethi, M., Dasog, G. S., Lieshout, A. V., Salimath, S. B. 2006. Salinity appraisal using IRS images in Shorapur Taluka, Upper Krishna Irrigation Project, Phase I, Gulbarga District, Karnataka, India. Int. J. Remote Sensing, 27(14):2917–2926.

Sharma, R. C., Bhargawa, G. P. 1988. Landsat imagery for mapping saline soils and wetlands in north-west India. Int. J. Remote Sensing, 9:69-84.

Singh, A. N., Kristof, S. J., Baumgardner, M. F. 1977. Delineating salt-affected soils in the Ganges Plain, India by digital analysis of Landsat data. The Laboratory of Applications of Remote Sensing. Purdue University, West Lafayette, Indiana.

Skidmore, A. K., Watford, F., Luckananurug, P., Ryan, P. J. 1996. An operational GIS expert system for mapping forest soils. Photogrammetric Engineering and Remote Sensing, 62:501-511.

Soil Survey Division Staff. 1993. Soil survey manual. Soil Conservation Service. U.S. Department of Agriculture Handbook 18.

Spies, B., Woodgate, P. 2005. Salinity mapping methods in the Australian context. Published by the Department of the Environment and Heritage, and Agriculture, Fisheries and Forestry.

Sukchan, S., Yamamoto, Y. 2002. Classification of salt affected areas using remote sensing and GIS. JIRCAS Working Report No. 30. http://www.jircas.affrc.go.jp/english/publication/working/30/30-01-02.pdf

Sunila, R., Laine, E., Kremenova, O. 2004. Fuzzy model and kriging for imprecise soil polygon boundaries. Proc. 12th Int. Conf. on Geoinformatics − Geospatial Information Research: Bridging the Pacific and Atlantic, June 7-9, University of Gävle, Sweden.

Szabolcs, I. 1987. The global problems of salt-affected soils. Acta Agronomica Hungarica, 36(1-2):159-172.

Tobalske, C. 2002. Map of historic vegetation for the State of Oregon. Oregon Natural Heritage Program. Portland, Oregon.

USDA. 2006. Land Resource Regions and Major Land Resource Areas of the United States, the Caribbean and the Pacific Basin. Handbook 296.

USDA-NRCS. 1999. Oregon annual precipitation. National Cartography and Geospatial Center. Fort Worth, TX.

USDA-NRCS. 2002. Field book for describing and sampling soils. Version 2. National Soil Survey Center - Natural Resources Conservation Service - U.S. Department of Agriculture.

USDA-NRCS. 2004a. Soil survey geographic (SSURGO) database for Benton County, Oregon. U.S. Department of Agriculture, Natural Resources Conservation Service, Fort Worth, Texas.

USDA-NRCS. 2004b. Soil survey geographic (SSURGO) database for Harney County area, Oregon. U.S. Department of Agriculture, Natural Resources Conservation Service, Fort Worth, Texas.

USDA-NRCS. 2007. Assessment and implementation of digital soil mapping in the national cooperative soil survey: a challenge dialogue with the soil survey community. Workshop held January 9-11, USDA Lyng Service Center, Davis, California.

USFS. 2006. Landfire existing vegetation cover. USDA forest Service. Missoula, Montana.

USGS. 2000-2001. Digital Orthophoto Quadrangles. U.S. Geological Survey, Reston, VA.

USGS. 2003. Spatial digital database for geologic map of Oregon. U.S. Department of Interior, U.S. Geological Survey.

USGS. 2005. LANDSAT TM – Path: 42 Row: 30 for Scene: 5042030000519810 and Path: 42 Row: 31 for Scene: 5042031000519810. U.S. Geological Survey Center for Earth Resources Observation and Science (EROS). Sioux Falls, SD.

USGS-EROS Data Center. 1999. Oregon 10m DEM. U.S. Geological Survey, Sioux Falls, SD.

USGS-EROS Data Center. 2000. Landsat ETM – Path: 46 Row: 29. U.S. Geological Survey and EROS Data Center, Sioux Falls, SD.

Venables, W. N., Ripley, B. D. 1994. Modern applied statistics with S-PLUS. Springer-Verlag, New York.

Walker, G.W., MacLeod, N.S., Miller, R.J., Raines, G.L., Connors, K.A., 2003, Spatial digital database for the geologic map of Oregon: U.S. Geological Survey, Open-File Report 03-67, ver. 2.0, 22 p.

Webster, R. 1994. The development of pedometrics. Geoderma, 62:1-15.

Webster, R., Oliver, M. A., 1990. Statistical methods in soil and land resource survey. Oxford University Press, Oxford.

Weiss, A. 2001. Topographic position and landforms analysis. Poster Presentation, ESRI User Conference, San Diego, CA.

Wilding, L. P. 1985. Spatial variability: its documentation, accommodation and implication to soil surveys. *In:* Nielsen, D.R., Bouma, J. (Eds.), Soil Spatial Variability. Proceedings of a Workshop of ISSS and the SSSA, Les Vegas USA. Nov. 30-Dec. 1, 1984.

Xian, G., Zhu, Z., Hoppus, M., Fleming, M. 2002. Applications of decision-tree techniques to forest group and basal area mapping using satellite imagery and forest inventory data. Pecora 15/Land Satellite Information IV/ISPRS Commission I/FIEOS Conference Proceedings.

Zadeh, L. A. 1965. Fuzzy sets. Information and Control, 8:338-53.

Zhou, B., Zhang, X., Wang, R. 2004. Automated soil resources mapping based on decision tree and Bayesian predictive modeling. J. Zhejiang Univ. Sci., 5(7):782-795.

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

Zhu, A. X. 2000. Mapping soil landscape as spatial continua: the neural network approach. Water Resources Research, 36:663-677.

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

Zhu, A. X., Hudson, B., Burt, J., Lubich, K., Simonson, D. 2001. Soil mapping using GIS, expert knowledge, and fuzzy logic. Soil Sci. Soc. Am. J., 65:1463-1472.

Zylman, J., Weindorf, D. C., Wittie, R., McFarland, A., Butler, T. 2005. Field-truthing of USDA-Natural Resources Conservation Service soil survey geographic data on Hunewell Ranch, Erath County, Texas. Soil Survey Horizons, 46(4):135-145.

Appendix 1

Table A.1.1. Map symbols and names of SMUs in the study area of Benton County, Oregon and their relative areas as percentage.

Map Symbol	Map Unit Name	%
1	Abiqua silty clay loam, 0 to 3 percent slopes	0.10
8	Amity silt loam, 0 to 3 percent slopes	4.24
9	Apt-McDuff complex, 5 to 30 percent slopes	1.92
10	Apt-McDuff complex, 30 to 50 percent slopes	0.29
12	Awbrig silty clay loam, 0 to 2 percent slopes	2.81
13	Bashaw clay, 3 to 12 percent slopes	0.05
17	Bellpine-Jory complex, 2 to 12 percent slopes	4.53
18	Bellpine-Jory complex, 12 to 20 percent slopes	6.83
19	Bellpine-Jory complex, 20 to 30 percent slopes	7.68
20	Bellpine-Jory complex, 30 to 60 percent slopes	10.79
21	Blachly-Kilowan complex, 5 to 30 percent slopes	0.42
22	Blachly-Kilowan complex, 30 to 60 percent slopes	0.24
23	Bohannon-Preacher complex, 30 to 60 percent slopes	3.96
24	Bohannon-Preacher complex, 60 to 90 percent slopes	1.56
27	Burntwoods-Oldblue complex, 30 to 60 percent slopes	0.81
28	Camas gravelly sandy loam, 0 to 3 percent slopes	0.35
29	Camas gravelly sandy loam, relict bar, 0 to 3 percent slopes	0.94
30	Caterl-Laderly-Romanose complex, 30 to 60 percent slopes	0.48
32	Caterl-Murtip-Giveout complex, 30 to 60 percent slopes	0.73
33	Caterl-Murtip-Laderly complex, 30 to 60 percent slopes	0.36
36	Chehalem silty clay loam, 0 to 3 percent slopes	0.14
37	Chehalem silty clay loam, 3 to 12 percent slopes	0.22
38	Chehalis silt loam, 0 to 3 percent slopes	0.06
40	Chehalis silty clay loam, 0 to 3 percent slopes	4.90
46	Cloquato silt loam, 0 to 3 percent slopes	0.26
48	Coburg complex, rarely and occasionally flooded, 0 to 3 percent	1.00
49	Coburg silty clay loam, 0 to 3 percent slopes	4.14
50	Coburg silty clay loam, rarely flooded, 0 to 3 percent slopes	1.11
51	Concord silt loam, 0 to 2 percent slopes	0.00
52	Conser silty clay loam, 0 to 3 percent slopes	4.90
53	Dayton silt loam, 0 to 2 percent slopes	22.27
54	Dayton silt loam, clay substratum, 0 to 2 percent slopes	0.77
55	Digger-Bohannon complex, 5 to 30 percent slopes	0.75
56	Digger-Remote-Umpcoos complex, 30 to 60 percent slopes	2.62

Table A.1.1. Map symbols and names of SMUs in the study area of Benton County, Oregon and their relative areas as percentage (Continued).

Map Symbol	Map Unit Name	%
57	Digger-Umpcoos-Remote complex, 60 to 90 percent slopes	1.00
58	Dixonville-Gellatly complex, 12 to 30 percent slopes	0.46
59	Dixonville-Gellatly complex, 30 to 60 percent slopes	0.33
60	Dixonville-Gellatly-Witham complex, 2 to 12 percent slopes	0.15
61	Dupee silt loam, 3 to 12 percent slopes	2.84
65	Fiverivers-Grassmountain-Chintimini complex, 30 to 60 percent slopes	0.63
68	Formader-Hemcross complex, 3 to 35 percent slopes	0.98
69	Formader-Hemcross complex, 35 to 60 percent slopes	0.10
70	Formader-Klistan-Hemcross complex, 60 to 80 percent slopes	0.16
75	Harslow-Kilchis-Rock outcrop complex, 60 to 90 percent slopes	0.30
83	Hemcross-Klistan complex, 5 to 30 percent slopes	0.89
84	Hemcross-Klistan complex, 30 to 60 percent slopes	0.73
85	Holcomb silt loam, 0 to 3 percent slopes	0.59
86	Honeygrove-Peavine complex, 3 to 30 percent slopes	11.34
87	Honeygrove-Peavine complex, 30 to 60 percent slopes	1.18
88	Honeygrove-Peavine complex, 3 to 30 percent slopes, basalts	0.90
89	Honeygrove-Peavine complex, 30 to 60 percent slopes, basalts	0.47
90	Honeygrove-Shivigny complex, 3 to 30 percent slopes	4.92
91	Jory silty clay loam, 2 to 12 percent slopes	2.84
94	Jory silty clay loam, sediments, 2 to 12 percent slopes	7.44
95	Jory silty clay loam, sediments, 12 to 20 percent slopes	8.52
96	Jory silty clay loam, sediments, 20 to 30 percent slopes	5.08
97	Jory-Dupee complex, 2 to 12 percent slopes	2.20
98	Jory-Gelderman complex, 12 to 30 percent slopes	5.88
102	Klistan-Harslow complex, 30 to 60 percent slopes	1.83
104	Laderly-Murtip-Giveout complex, 5 to 30 percent slopes	2.99
106	Linslaw loam, 3 to 8 percent slopes	0.00
109	MacDunn-Price-Ritner complex, 60 to 90 percent slopes	0.06
110	Malabon silty clay loam, 0 to 3 percent slopes	5.62
113	McAlpin silty clay loam, 0 to 3 percent slopes	1.40
114	McAlpin silty clay loam, 3 to 6 percent slopes	0.04
117	McAlpin silty clay loam, rarely flooded, 0 to 3 percent slopes	0.72
118	McBee silty clay loam, 0 to 3 percent slopes	1.21
119	McBee silty clay loam, nonflooded, 0 to 3 percent slopes	0.04

Table A.1.1. Map symbols and names of SMUs in the study area of Benton County, Oregon and their relative areas as percentage (Continued).

Map Symbol	Map Unit Name	%
120	Meda-Treharne-Wasson complex, 2 to 20 percent slopes	1.84
123	Murtip-Giveout-Laderly complex, 5 to 30 percent slopes	3.39
125	Newberg fine sandy loam, 0 to 3 percent slopes	0.52
127	Newberg loam, 0 to 3 percent slopes	0.77
128	Oldblue-Burntwoods complex, 5 to 30 percent slopes	1.02
130	Pengra silt loam, 2 to 12 percent slopes	1.29
133	Pits silty clay, 0 to 5 percent slopes	0.16
134	Preacher-Blachly-Bohannon complex, 5 to 30 percent slopes	1.31
135	Preacher-Bohannon complex, 5 to 35 percent slopes	0.32
136	Preacher-Bohannon-Slickrock complex, 35 to 60 percent slopes	4.77
137	Price-MacDunn-Ritner complex, 30 to 60 percent slopes	1.00
139	Salem gravelly silt loam, 0 to 3 percent slopes	0.14
140	Santiam silt loam, 2 to 8 percent slopes	6.83
141	Santiam silt loam, 8 to 20 percent slopes	1.59
145	Shivigny-Honeygrove complex, 30 to 60 percent slopes	2.29
146	Slickrock gravelly medial loam, 3 to 25 percent slopes	2.07
147	Steiwer-Chehulpum complex, 3 to 12 percent slopes	0.13
148	Steiwer-Chehulpum complex, 12 to 30 percent slopes	0.17
149	Steiwer-Chehulpum complex, 30 to 60 percent slopes	0.04
150	Treharne-Eilertsen-Zyzzug complex, 0 to 7 percent slopes	0.28
154	Verboort silty clay loam, 0 to 3 percent slopes	1.00
155	Waldo silty clay loam, 0 to 3 percent slopes	10.50
157	Wapato silty clay loam, 0 to 3 percent slopes	1.30
159	Water	0.70
163	Willakenzie loam, 2 to 12 percent slopes	0.35
164	Willakenzie loam, 12 to 20 percent slopes	0.30
169	Willamette silt loam, 0 to 3 percent slopes	0.93
170	Willamette silt loam, 3 to 12 percent slopes	0.93
174	Witzel-Ritner complex, 3 to 12 percent slopes	0.05
177	Woodburn silt loam, 0 to 3 percent slopes	7.63
178	Woodburn silt loam, 3 to 12 percent slopes	0.30
180	Woodburn silt loam, 20 to 55 percent slopes	0.18

Table A.1.2. Soil taxonomic classification of soil series in the study area.

Soil Name	Taxonomic Classification
Awbrig	Fine, smectitic, mesic Vertic Albaqualfs
Dayton	Fine, smectitic, mesic Vertic Albaqualfs
Concord	Fine, smectitic, mesic Typic Endoaqualfs
Eilertsen	Fine-silty, isotic, mesic Ultic Hapludalfs
Treharne	Fine-silty, isotic, mesic Aquultic Hapludalfs
Dupee	Fine, mixed, superactive, mesic Aquultic Haploxeralfs
Linslaw	Fine, mixed, superactive, mesic Aquultic Haploxeralfs
Santiam	Fine, mixed, superactive, mesic Aquultic Haploxeralfs
Willakenzie	Fine-loamy, mixed, active, mesic Ultic Haploxeralfs
Burntwoods	Medial-skeletal over loamy-skeletal, mixed over isotic, frigid Typic Fulvudands
Caterl	Medial-skeletal, ferrihydritic, frigid Alic Hapludands
Formader	Medial over loamy, ferrihydritic over isotic, mesic Alic Hapludands
Giveout	Medial, ferrihydritic, frigid Alic Hapludands
Harslow	Medial-skeletal, ferrihydritic, mesic Alic Hapludands
Hemcross	Medial, ferrihydritic, mesic Alic Hapludands
Klistan	Medial-skeletal, ferrihydritic, mesic Alic Hapludands
Laderly	Medial-skeletal, ferrihydritic, frigid Alic Hapludands
Murtip	Medial, ferrihydritic, frigid Alic Hapludands
Slickrock	Medial over loamy, ferrihydritic over isotic, mesic Alic Hapludands
Romanose	Medial-skeletal, ferrihydritic, frigid Lithic Hapludands
Wasson	Coarse-loamy, mixed, superactive, nonacid, mesic Fluvaquentic Humaquepts
Zyzzug	Fine-silty, isotic, acid, mesic Typic Humaquepts
Blachly	Fine, isotic, mesic Typic Dystrudepts
Bohannon	Fine-loamy, isotic, mesic Andic Dystrudepts
Chintimini	Loamy-skeletal, isotic, frigid Andic Dystrudepts
Fiverivers	Fine-loamy, isotic, frigid Andic Dystrudepts
Grassmountain	Fine-loamy, isotic, frigid Andic Dystrudepts
Oldblue	Fine-loamy, isotic, frigid Andic Dystrudepts
Preacher	Fine-loamy, isotic, mesic Andic Dystrudepts
Meda	Fine-loamy, isotic, mesic Humic Dystrudepts
Kilchis	Loamy-skeletal, isotic, mesic Humic Lithic Dystrudepts
Kilowan	Fine, isotic, mesic Typic Dystrudepts
Digger	Loamy-skeletal, isotic, mesic Dystric Eutrudepts
Umpcoos	Loamy-skeletal, isotic, mesic Lithic Eutrudepts
Remote	Loamy-skeletal, isotic, mesic Typic Eutrudepts
MacDunn	Clayey-skeletal, mixed, superactive, mesic Typic Haploxerepts
Price	Fine, mixed, superactive, mesic Typic Haploxerepts
Ritner	Clayey-skeletal, mixed, superactive, mesic Typic Haploxerepts

Table A.1.2. Soil taxonomic classification of soil series in the study area (Continued).

Soil Name	Taxonomic Classification
Amity	Fine-silty, mixed, superactive, mesic Argiaquic Xeric Argialbolls
Verboort	Fine, mixed, superactive, mesic Xerertic Argialbolls
Holcomb	Fine, smectitic, mesic Typic Argialbolls
Conser	Fine, mixed, superactive, mesic Vertic Argiaquolls
Chehalem	Fine, smectitic, mesic Cumulic Vertic Endoaquolls
Waldo	Fine, smectitic, mesic Fluvaquentic Vertic Endoaquolls
Wapato	Fine-silty, mixed, superactive, mesic Fluvaquentic Endoaquolls
Pengra	Fine-silty over clayey, mixed, superactive, mesic Vertic Epiaquolls
Woodburn	Fine-silty, mixed, superactive, mesic Aquultic Argixerolls
Coburg	Fine, mixed, superactive, mesic Oxyaquic Argixerolls
Gellatly	Fine, mixed, superactive, mesic Pachic Argixerolls
Dixonville	Fine, mixed, superactive, mesic Pachic Ultic Argixerolls
Malabon	Fine, mixed, superactive, mesic Pachic Ultic Argixerolls
Salem	Fine-loamy over sandy or sandy-skeletal, mixed, superactive, mesic Pachic Ultic Argixerolls
Willamette	Fine-silty, mixed, superactive, mesic Pachic Ultic Argixerolls
McAlpin	Fine, mixed, superactive, mesic Aquic Cumulic Haploxerolls
McBee	Fine-silty, mixed, superactive, mesic Aquic Cumulic Haploxerolls
Abiqua	Fine, mixed, superactive, mesic Cumulic Ultic Haploxerolls
Chehalis	Fine-silty, mixed, superactive, mesic Cumulic Ultic Haploxerolls
Cloquato	Coarse-silty, mixed, superactive, mesic Cumulic Ultic Haploxerolls
Camas	Sandy-skeletal, mixed, mesic Fluventic Haploxerolls
Newberg	Coarse-loamy, mixed, superactive, mesic Fluventic Haploxerolls
Witzel	Loamy-skeletal, mixed, active, mesic Lithic Ultic Haploxerolls
Chehulpum	Loamy, mixed, superactive, mesic, shallow Ultic Haploxerolls
Steiwer	Fine-loamy, mixed, superactive, mesic Ultic Haploxerolls
Witham	Fine, smectitic, mesic Vertic Haploxerolls
Apt	Fine, isotic, mesic Typic Haplohumults
McDuff	Fine, isotic, mesic Typic Haplohumults
Peavine	Fine, mixed, active, mesic Typic Haplohumults
Bellpine	Fine, mixed, active, mesic Xeric Haplohumults
Gelderman	Fine, mixed, active, mesic Xeric Haplohumults
Honeygrove	Fine, mixed, active, mesic Typic Palehumults
Shivigny	Clayey-skeletal, mixed, active, mesic Typic Palehumults
Jory	Fine, mixed, active, mesic Xeric Palehumults
Bashaw	Very-fine, smectitic, mesic Xeric Endoaquerts
Pits	Fine, smectitic, mesic Xeric Endoaquerts
Water	Water

Appendix 2

Table A.2.1. Soil orders, suborders, great groups in the study area and their codes.

Order	Code	Order name
1	1	Alfisols
2	2	Andisols
3	3	Inceptisols
4	4	Mollisols
5	5	Ultisols
6	6	Vertsols
7	7	Water
1 and 3	8	
2 and 3	9	
3 and 4	10	
1 and 5	11	

Suborder	Code	Sub-order name
1	1	aqualfs
2	2	udalfs
3	3	xeralfs
4	4	udands
5	5	aquepts
6	6	udepts
7	7	xerepts
8	8	albolls
9	9	aquolls
10	10	xerolls
11	11	humults
12	12	aquerts
13	13	Water
2 and 5	14	
4 and 6	15	
2, 5, and 6	16	
7 and 10	17	
3 and 11	18	

Great group	Code	Great group name
1	1	Albaqualfs
2	2	Endoaqualfs
3	3	Hapludalfs
4	4	Haploxeralfs
5	5	Fulvudands
6	6	Hapludands
7	7	Humaquepts
8	8	Dystrudepts
9	9	Eutrudepts
10	10	Haploxerepts
11	11	Argialbolls
12	12	Argiaquolls
13	13	Endoaquolls
14	14	Epiaquolls
15	15	Argixerolls
16	16	Haploxerolls
17	17	Haplohumults
18	18	Palehumults
19	19	Endoaquerts
20	20	Water
3 and 7	21	
5 and 8	22	
3, 7, and 8	23	
6 and 8	24	
8 and 9	25	
15 and 16	26	
10 and 16	27	
17 and 18	28	
5 and 18	29	

Table A.2.2. Soil subgroups in the study area and their codes.

SG	Code	Subgroup name	SG	Code	Subroup name
1	1	Vertic Albaqualfs	31	31	Pachic Ultic Argixerolls
2	2	Typic Endoaqualfs	32	32	Aquic Cumulic Haploxerolls
3	3	Ultic Hapludalfs	33	33	Cumulic Ultic Haploxerolls
4	4	Aquultic Hapludalfs	34	34	Fluventic Haploxerolls
5	5	Aquultic Haploxeralfs	35	35	Lithic Ultic Haploxerolls
6	6	Ultic Haploxeralfs	36	36	Ultic Haploxerolls
7	7	Typic Fulvudands	37	37	Vertic Haploxerolls
8	8	Alic Hapludands	38	38	Typic Haplohumults
9	9	Lithic Hapludands	39	39	Xeric Haplohumults
10	10	Fluvaquentic Humaquepts	40	40	Typic Palehumults
11	11	Typic Humaquepts	41	41	Xeric Palehumults
12	12	Typic Dystrudepts	42	42	Xeric Endoaquerts
13	13	Andic Dystrudepts	43	43	Water
14	14	Humic Dystrudepts	4 and 11	44	
15	15	Humic Lithic Dystrudepts	7 and 13	45	
16	16	Dystric Eutrudepts	8 and 9	46	
17	17	Lithic Eutrudepts	12 and 15	47	
18	18	Typic Eutrudepts	12 and 13	48	
19	19	Typic Haploxerepts	8, 12, and 13	49	
20	20	Argiaquic Xeric Argialbolls	4, 10, and 13	50	
21	21	Xerertic Argialbolls	16, 17, and 18	51	
22	22	Typic Argialbolls	30 and 31	52	
23	23	Vertic Argiaquolls	30, 31, and 37	53	
24	24	Cumulic Vertic Endoaquolls	19 and 35	54	
25	25	Fluvaquentic Vertic Endoaquolls	39 and 41	55	
26	26	Fluvaquentic Endoaquolls	39 and 41	55	
27	27	Vertic Epiaquolls	38 and 40	56	
28	28	Aquultic Argixerolls	5 and 41	57	
29	29	Oxyaquic Argixerolls	12 and 16	58	
30	30	Pachic Argixerolls			

Table A.2.3. Soil orders, suborders, great groups, and subgroups in the study area and their proportional areas (%).

Order		Great Group		Subgroup			
Order	%	GG	%	SG	%	SG	%
1	17.13	1	11.74	1	11.74	47	0.30
2	6.69	4	5.40	5	5.10	48	3.26
3	6.32	6	9.49	6	0.30	49	2.16
4	26.66	8	3.85	8	6.60	50	0.84
5	37.67	9	1.64	13	0.29	51	1.64
6	0.09	10	0.48	19	0.48	52	0.36
7	0.32	11	2.65	20	1.93	53	0.07
8	0.97	12	2.22	21	0.45	54	0.02
9	3.13	13	5.53	22	0.27	55	16.20
10	0.02	14	0.58	23	2.22	56	6.32
11	1.00	15	10.33	24	0.17	57	1.00
Suborder		16	5.27	25	4.78	58	0.35
Suborder	%	17	1.01	26	0.59		
1	11.74	18	14.14	27	0.58		
3	5.40	19	0.09	28	3.68		
4	6.69	20	0.32	29	2.84		
6	5.84	21	0.13	31	3.46		
7	0.48	22	0.83	32	1.56		
8	2.65	23	0.84	33	2.39		
9	8.33	24	2.16	34	1.17		
10	15.67	25	0.35	36	0.16		
11	37.67	26	0.07	38	1.01		
12	0.09	27	0.02	40	3.28		
13	0.32	28	19.85	41	10.87		
14	0.13	29	1.00	42	0.09		
15	3.13			43	0.32		
16	0.84			44	0.13		
17	0.02			45	0.83		
18	1.00			46	0.22		

Table A.2.4. User accuracy of soil subgroups and major soil map units in the study area.

Subgroups (SG)				Major soil map units			
SG	User accuracy	SG	User accuracy	SMU	User accuracy	SMU	User accuracy
1	92.40	43	54.44	8	42.30	96	70.85
5	85.87	44	72.22	9	75.67	97	81.73
6	71.60	45	86.38	12	44.25	98	81.73
8	87.99	46	60.32	17	84.33	102	85.66
13	63.89	47	44.05	18	76.10	104	86.82
19	48.89	48	83.48	19	73.61	110	75.46
20	42.90	49	65.96	20	83.43	113	73.33
21	31.75	50	71.37	23	71.01	118	76.54
22	49.37	51	78.17	24	73.95	120	81.89
23	53.71	52	94.74	40	91.15	123	93.26
24	60.87	53	22.22	48	54.23	128	82.01
25	84.22	54	62.50	49	50.35	130	82.04
26	28.11	55	91.94	50	40.74	134	86.89
27	69.01	56	93.38	52	60.23	136	72.58
28	39.41	57	74.11	53	95.09	137	68.87
29	46.42	58	48.67	56	82.04	140	88.53
31	71.62			57	77.33	141	65.99
32	73.05			61	73.43	145	79.57
33	80.75			68	77.08	146	74.37
34	43.11			86	94.32	154	17.22
36	69.57			87	51.72	155	86.73
38	67.83			90	93.06	157	33.51
40	92.69			91	83.55	177	42.61
41	81.18			94	77.86		
42	58.62			95	74.68		
Total			79.43				77.71

Table A.2.5. User accuracy of soil great groups and all soil map units in the study area.

Great groups (GG)		All Soil Map Units (SMUs)					
GG	User accuracy	SMU	User accuracy	SMU	User accuracy	SMU	User accuracy
1	88.37	1	45.45	57	78.86	120	77.73
4	85.81	8	40.24	58	96.88	123	90.35
6	84.36	9	67.87	59	81.25	125	61.73
8	84.30	10	31.25	60	35.00	127	37.61
9	74.07	12	54.97	61	69.17	128	84.00
10	56.34	13	0.00	65	81.93	130	77.51
11	37.47	17	81.48	68	71.54	133	55.56
12	50.16	18	68.03	69	54.55	134	87.43
13	75.37	19	71.68	70	50.00	135	59.38
14	75.31	20	83.48	75	94.74	136	73.04
15	68.38	21	36.54	83	42.98	137	64.89
16	91.88	22	61.76	84	50.00	139	0.00
17	62.12	23	66.42	85	64.94	140	87.43
18	83.61	24	67.91	86	93.72	141	60.29
19	25.00	27	79.66	87	49.68	145	78.55
20	63.54	28	55.81	88	55.65	146	70.98
21	67.50	29	25.20	89	50.82	147	73.33
22	81.28	30	69.09	90	94.45	148	90.48
23	72.31	32	68.89	91	77.57	149	20.00
24	64.64	33	53.66	94	82.11	150	85.11
25	50.00	36	58.82	95	73.41	154	21.14
26	19.05	37	81.25	96	66.67	155	86.54
27	0.00	38	30.00	97	78.98	157	30.07
28	92.86	40	85.20	98	80.31	159	47.87
29	75.87	46	58.33	102	80.50	163	82.22
		48	46.15	104	82.85	164	45.45
		49	44.32	109	100.00	169	67.42
		50	36.24	110	76.13	170	41.07
		52	57.74	113	67.05	174	62.50
		53	93.52	114	57.14	177	39.40
		54	71.03	117	82.80	178	33.33
		55	55.56	118	73.41	180	42.86
		56	71.43	119	60.00		
Total	81.22						74.17

Appendix 3

Table A.3.1. Soil samples and their XY coordinates, saturation percentage (SP), field capacity (FC), PH and EC values.

Sample No	Latitude (X)	Longitude (Y)	SP (%)	FC (%)	PH	EC (uS/m)	EC (dS/m)
1	427377.94	4732683.5	23	11.5	8.61	520	0.52
2	427408.01	4732738.4	22	11	7.95	440	0.44
3	425123.02	4733171.7	23	11.5	7.7	420	0.42
4	426066.03	4729825.9	22	11	6.5	430	0.43
5	425888.5	4729396.5	20	10	7.28	490	0.49
6	426422.18	4727613.1	17	8.5	7.38	360	0.36
8	423415.2	4726609.6	20	10	7.3	580	0.58
9	426927.68	4702650.1	14.6	7.3	7.29	630	0.63
10	426253.98	4702676.5	21.78	10.89	7.04	700	0.7
11	426017.04	4702615.3	20.76	10.38	6.68	290	0.29
12	424288.7	4702239.8	30.08	15.04	8.24	670	0.67
13	423567	4702049.6	27.44	13.72	8.22	750	0.75
14	421671.78	4700790.2	19.34	9.67	7.24	370	0.37
15	419418.24	4699755.5	21.51	10.75	6.72	260	0.26
16	418614.62	4699396.5	25	12.5	6.5	590	0.59
17	416709.02	4699157.9	24.8	12.4	6.65	180	0.18
18	415714.33	4699196.4	31.13	15.56	7.64	590	0.59
19	414747.16	4701045	24	12	7.32	460	0.46
20	414546.51	4701464.7	24	12	7.32	490	0.49
21	415354.86	4698688.2	23	11.5	7.5	450	0.45
22	414068.77	4698174.9	23	11.5	7.37	530	0.53
23	412616.4	4697856.8	24	12	7.47	290	0.29
24	410844.16	4697052.9	23	11.5	7.61	1170	1.17
25	410501.52	4697923.7	21	10.5	7.44	480	0.48
26	410719.77	4697799	26	13	7.34	310	0.31
27	409901.46	4696267.5	21.5	10.75	8.01	1030	1.03
28	406761.63	4694464.1	21	10.5	7.44	480	0.48
29	406892.03	4694315.7	20	10	7.05	310	0.31
30	414572.08	4702216.2	32.5	16.25	7.28	680	0.68
31	414849.78	4702808.5	17.5	8.75	7.53	370	0.37

Table A.3.1. Soil samples and their XY coordinates, saturation percentage (SP), field capacity (FC), PH and EC values (Continued).

Sample No	Latitude (X)	Longitude (Y)	SP (%)	FC (%)	PH	EC (uS/m)	EC (dS/m)
32	415031.27	4702509.5	19	9.5	6.86	340	0.34
33	414944.99	4702179.6	18	9	7.53	380	0.38
34	414987.25	4702322.7	36	18	8.16	10460	10.46
35	413758.88	4702718.2	23	11.5	7.52	280	0.28
36	412967.26	4704219.4	17	8.5	7.44	1200	1.2
37	413296.51	4705308	17.5	8.75	6.91	340	0.34
38	414475.75	4707121.7	18	9	7.35	610	0.61
39	414792.83	4707805.3	22	11	7.53	330	0.33
40	412294.93	4704190.7	23	11.5	8.29	750	0.75
41	412123.3	4704070	22	11	8.15	570	0.57
42	410889.01	4703133.1	21	10.5	8.15	580	0.58
43	409854.15	4701926.1	16	8	8.11	640	0.64
44	410118.53	4700490.7	25	12.5	7.47	240	0.24
45	410487.47	4699346.5	26	13	8.44	220	0.22
46	410777.35	4698820.7	39	19.5	8.4	520	0.52
50	422172.62	4727424.8	23	11.5	7.81	860	0.86
51	422212.53	4727397.2	24	12	7.81	870	0.87
52	418912.71	4729983.2	19	9.5	8.01	520	0.52
53	418594.51	4730554.4	21	10.5	7.06	310	0.31
54	418514.02	4731016.1	38	19	8.83	2740	2.74
55	417476.8	4733657.3	34	17	6.91	390	0.39
56	418193.83	4730349.4	24	12	7.35	570	0.57
57	415416.28	4731009.9	17	8.5	7.67	390	0.39
58	412253.86	4731769.9	28	14	7.96	410	0.41
59	410686.92	4732067.5	20	10	7.87	390	0.39
60	408363.35	4733590.3	22	11	7.59	250	0.25
61	406542.5	4732975	29	14.5	7.6	390	0.39
62	406331.06	4731990.5	24	12	8.15	860	0.86

Table A.3.1. Soil samples and their XY coordinates, saturation percentage (SP), field capacity (FC), PH and EC values (Continued).

Sample No	Latitude (X)	Longitude (Y)	SP (%)	FC (%)	PH	EC (uS/m)	EC (dS/m)
63	406156.09	4730817.4	28	14	7.13	280	0.28
64	405694.8	4729525.3	21	10.5	7.15	400	0.4
65	404604.87	4727921.7	21	10.5	7.87	470	0.47
66	403608.37	4725797.8	23	11.5	7.79	890	0.89
67	403430.18	4723891.5	19	9.5	7.78	420	0.42
68	404401.12	4721619.5	23	11.5	7.15	300	0.3
69	405882.81	4719256.3	21	10.5	8.05	510	0.51
70	406868.4	4717727	24	12	8.04	1030	1.03
71	408388.72	4707558.1	26	13	7.66	620	0.62
72	408495.26	4708655.8	30	15	7.26	170	0.17
73	409219.82	4709510.1	32	16	7.51	290	0.29
74	408962.63	4709787.6	21	10.5	8.45	1010	1.01
75	409722.16	4710948.8	29	14.5	7.42	600	0.6
76	410581.4	4712056.5	23	11.5	7.92	290	0.29
77	425193.15	4725245.9	28	14	8.3	440	0.44
78	425025.44	4724556.7	26	13	9.88	1120	1.12
79	424811.89	4724361.9	32	16	9.96	24400	24.4
80	424355.15	4723633.5	24	12	8.09	370	0.37
81	424380.79	4722649	19	9.5	8.29	430	0.43
82	425197.7	4721033	24	12	7.84	430	0.43
83	425699.29	4717899.2	19	9.5	7.76	390	0.39
84	425871.59	4716142.9	23	11.5	7.74	460	0.46
85	425960.85	4714926.9	24	12	8.43	380	0.38
86	426037.33	4713176.8	23	11.5	8.28	790	0.79
87	427108.89	4713900.8	21	10.5	7.54	350	0.35
88	426086.03	4712601.7	25	12.5	8.44	310	0.31
89	426142.93	4711680.6	28	14	8.53	320	0.32
90	425617.18	4710505.5	27	13.5	7.44	760	0.76
91	425862.26	4709560.5	28	14	7.67	510	0.51
92	426553.44	4709005.8	22	11	7.76	580	0.58
93	426866.19	4708889.3	24	12	7.94	470	0.47

Table A.3.1. Soil samples and their XY coordinates, saturation percentage (SP), field capacity (FC), PH and EC values (Continued).

Sample No	Latitude (X)	Longitude (Y)	SP (%)	FC (%)	PH	EC (uS/m)	EC (dS/m)
94	425925.04	4708450.5	20	10	7.46	310	0.31
95	426054.15	4707450.4	25	12.5	7.36	380	0.38
96	426035.03	4706655.1	23	11.5	7.37	250	0.25
97	426244.79	4705883.6	25	12.5	7.15	330	0.33
98	426776.98	4704959.9	26	13	7.38	320	0.32
100	411721.7	4730550.2	23	11.5	7.85	580	0.58
101	411380.19	4728612.9	19	9.5	7.54	610	0.61
102	411389	4726929	20	10	7.53	330	0.33
103	411774.97	4724880.3	23	11.5	7.63	390	0.39
104	411958.85	4721803.1	22	11	7.61	370	0.37
105	411607.96	4720336.8	22	11	7.25	370	0.37
106	410443	4719557.4	20	10	7.04	290	0.29
107	409301.52	4719135.5	22	11	7.33	350	0.35
108	407367.34	4719432	28	14	7.81	410	0.41
109	423857	4723402.9	18	9	8.28	380	0.38
110	423125.47	4722782.8	23	11.5	7.47	640	0.64
111	422452.4	4722349.4	20	10	8.31	360	0.36
112	421205.9	4721734.9	24	12	7.21	490	0.49
113	419635.81	4720876.2	28	14	7.62	350	0.35
114	418378.39	4720229.8	21	10.5	6.92	270	0.27
115	417502.42	4718077.2	23	11.5	7.68	190	0.19
116	416984.91	4715862.5	24	12	7.27	170	0.17
117	409519.47	4702445.9	30	15	8.43	530	0.53
118	409344.89	4702486.3	26	13	7.61	230	0.23
119	409069.46	4703190.4	26	13	7.92	370	0.37
120	408710.97	4704169.8	24	12	8.26	270	0.27
121	408327.55	4705144.5	22	11	7.45	200	0.2
122	408258.71	4706210.8	20	10	7.65	350	0.35
123	407613.56	4707253.9	23	11.5	7.59	290	0.29
124	406082.2	4719961.2	24	12	8.03	460	0.46

Table A.3.1. Soil samples and their XY coordinates, saturation percentage (SP), field capacity (FC), PH and EC values (Continued).

Sample No	Latitude (X)	Longitude (Y)	SP (%)	FC (%)	PH	EC (uS/m)	EC (dS/m)
125	401181.16	4720263.5	19	9.5	8.11	650	0.65
126	401891.45	4720889.9	21	10.5	7.54	540	0.54
127	402931.29	4721663.3	27	13.5	7.91	380	0.38
128	402937.38	4721702.1	24	12	8.43	610	0.61
129	403470.33	4722067.7	20	10	7.53	370	0.37
130	403489.42	4722037.1	18	9	8.06	850	0.85
131	402883.27	4724779.1	21	10.5	7.21	440	0.44
132	416657.91	4727823.7	17	8.5	7.46	510	0.51
133	417245.27	4728204.2	19	9.5	7.08	770	0.77
134	418062.92	4729002.8	17	8.5	7.5	330	0.33
135	418810	4729336.4	21	10.5	7.16	430	0.43
136	419185.89	4729536.1	20	10	7.98	460	0.46
137	420717.46	4724326	19	9.5	7.33	220	0.22
138	420965.41	4724276.8	21	10.5	8.04	620	0.62
139	421327.8	4724224.7	23	11.5	7.44	390	0.39
140	422147.46	4724465.7	17	8.5	7.41	190	0.19
141	423049.65	4725197.1	24	12	7.83	3720	3.72
142	427381.18	4732731.7	23	11.5	8.61	530	0.53
143	426658.09	4733220.6	21	10.5	7.95	420	0.42
144	425412.24	4732829.9	24	12	7.7	410	0.41
145	425662.89	4731654.6	19	9.5	8.07	570	0.57
146	425683.09	4731683.9	19	9.5	7.3	230	0.23
147	426042.92	4729849.2	21	10.5	6.5	420	0.42
148	425968.68	4728624.1	21	10.5	7.12	560	0.56
149	426084.54	4728686	25	12.5	7	300	0.3
150	426482.65	4727604.4	17	8.5	7.38	360	0.36
151	426387.26	4727194.6	24	12	7.21	600	0.6
152	424742.04	4726287.9	18	9	6.4	250	0.25
153	423634.52	4725882.5	19	9.5	7.3	530	0.53
154	424152.64	4726384.3	17	8.5	0	310	0.31
155	428193.07	4715631.7	17	8.5	0	290	0.29

Table A.3.1. Soil samples and their XY coordinates, saturation percentage (SP), field capacity (FC), PH and EC values (Continued).

Sample No	Latitude (X)	Longitude (Y)	SP (%)	FC (%)	PH	EC (uS/m)	EC (dS/m)
156	427625.09	4715915.1	19	9.5	7.2	520	0.52
157	427973.69	4715474.2	17	8.5	7.1	350	0.35
158	426133.77	4712108.2	15	7.5	0	550	0.55
159	425623.06	4711016.5	19	9.5	0	250	0.25
160	425780.62	4710137.8	18	9	0	340	0.34
161	425953.47	4708881.6	17	8.5	0	740	0.74
162	425899.97	4708882.5	18	9	0	340	0.34
163	426015.92	4707870.6	22	11	6.8	360	0.36
164	426455.35	4705477.9	18	9	0	230	0.23
165	426896.44	4702641.3	19	9.5	6.5	370	0.37
166	407057.59	4707440.4	19	9.5	0	670	0.67
167	405338.26	4708371.3	22	11	0	440	0.44
168	407099.17	4706095.7	31	15.5	0	500	0.5
169	406782.44	4706500.2	30	15	0	530	0.53
170	408105.4	4704253.8	19	9.5	0	240	0.24
171	407871.3	4703503.9	20	10	0	260	0.26
172	407486.02	4701886.9	21	10.5	0	270	0.27
173	407224.03	4700263.2	18	9	0	430	0.43
174	406892.73	4698757.6	20	10	0	540	0.54
175	407353.1	4696580.2	19	9.5	8	760	0.76
176	408929.56	4695946.1	20	10	7.1	760	0.76
177	408152.77	4696268	21	10.5	0	420	0.42
178	406318.14	4697027.9	22	11	6.2	360	0.36
179	405332.87	4697421.3	21	10.5	0	210	0.21
180	404896.24	4697608.8	28	14	0	4760	4.76
181	405785.21	4696543.5	21	10.5	0	240	0.24
182	405764.47	4696574.4	17	8.5	8.2	480	0.48
183	405399.65	4696082.4	23	11.5	6.4	310	0.31
184	404927.86	4695668.7	23	11.5	6	190	0.19
185	404459.85	4695017.9	33	16.5	7.7	920	0.92
186	403694.64	4694173.1	19	9.5	7.5	1880	1.88

Table A.3.1. Soil samples and their XY coordinates, saturation percentage (SP), field capacity (FC), PH and EC values (Continued).

Sample No	Latitude (X)	Longitude (Y)	SP (%)	FC (%)	PH	EC (uS/m)	EC (dS/m)
187	403227.03	4692649.6	20	10	7.3	450	0.45
188	403211.77	4691923.2	22	11	7.2	350	0.35
189	403954.43	4691811	21	10.5	8.1	500	0.5
190	403943.17	4691854.1	19	9.5	8	690	0.69
191	405440.78	4692872.6	38	19	7.8	890	0.89
192	406271.42	4693505.5	22	11	7.3	450	0.45
193	407801.61	4694698.7	20	10	7.3	390	0.39
194	417839.79	4699197.6	26	13	6.5	590	0.59
195	420708.95	4700330.8	18	9	7.3	660	0.66
196	422735.11	4701180.6	19	9.5	7.5	460	0.46
197	425314.18	4726403.4	26	13	8.6	12460	12.46
198	425197.87	4725922.8	23	11.5	8.6	82800	82.8
199	424462.08	4724346.2	25	12.5	8.6	27500	27.5
200	421690.61	4721690.2	26	13	8.5	420	0.42
201	416253.58	4715294.6	25	12.5	6.3	410	0.41
202	415306.92	4714382.1	35	17.5	7.7	550	0.55
203	413780.43	4713957.4	22	11	6.8	830	0.83
204	413477.89	4713701.7	26	13	6.2	180	0.18
205	412733.85	4713277.4	22	11	6.8	830	0.83
206	407566.18	4715316.7	33	16.5	8.6	10500	10.5
207	407816.6	4716032.2	31	15.5	8.43	10530	10.53
208	408710.94	4715853.3	36	18	8.16	11270	11.27
209	409641.06	4714440.3	31	15.5	8.44	10480	10.48
210	408432.14	4714335.7	34	17	8.63	11070	11.07
211	409891.48	4713367	36	18	8.57	10860	10.86
212	408979.25	4712472.7	34	17	8.73	11070	11.07
213	408192.23	4712723.1	30	15	8.81	10660	10.66
214	415939.02	4709556.7	22	11	7.12	790	0.79
215	416740.94	4712447.6	20	10	7.1	830	0.83