uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Geneviève Ouellet
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.Sc. (Geography)
GRADE / DEGREE

Department of Geography
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Towards Regionally-Based Indicators of Soil Infiltrability: Eastern Ontario, Canada
TITRE DE LA THÈSE / TITLE OF THESIS

Michael Sawada
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

David Lapen
CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

K. Gajewski

A. Lewkowicz

Gary W. Slater
Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

Towards Regionally-Based Indicators of Soil Infiltrability: Eastern Ontario, Canada

By
Geneviève Ouellet

Thesis submitted to
The Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements for the degree
Master of Science, Geography

Department of Geography
University of Ottawa
Ottawa, Ontario
May 2007

Your file Votre référence
ISBN: 978-0-494-34098-1
Our file Notre référence
ISBN: 978-0-494-34098-1

Abstract

Macropore flow is a gravity-driven phenomenon that exhibits high contamination risk for surface and shallow groundwater. In order to address the necessity of having reliable and tangible indicators of macropore flow incidence, this research is an extensive data exploration that uncovers potential regional macropore flow incidence indicators that are synthesized as hypotheses for future research. CART-based regression trees are used to uncover significant relationships among soil physical properties, management regimes, field saturated hydraulic conductivity, saturated soil air-entry tensions and earthworm biomass. All of the modeling is based on field-measured data. Prediction scenarios are established in order to develop potential regional indicators of field saturated hydraulic conductivity (K_{fs}), soil saturated air-entry tension (AEV), and earthworm biomass. The magnitude, range, and pattern of variability of soil water infiltrations are controlled primarily by the well-developed and stable soil structure at the field scale, rather than by texture, organic carbon or surface topography, as classical pedotransfer functions suggest. Tillage is influential for the estimation of all targeted variables. Conventional tillage is associated with rapid shallow infiltration rates and lower air entry tensions both in the cultivation layer and below. Absence of tillage is found to be beneficial for earthworm communities as sampled biomass is greater on no-till field sites. The strong correlations between K_{fs} and AEV demonstrate good prediction potential but because of their intrinsic spatial and temporal variability, their utility as regional macropore flow indicators is questionable.

Remerciements

This research was supported by an OCPA (Ontario Corn Growers Association) research grant to D.R. Lapen, and by the Agriculture Policy Framework's GAPs program research grant to D.R. Lapen. Financial support was also offered by Agriculture and Agri-Food Canada Matching Investment Initiative in cooperation with Ontario Federation of Agriculture, Ontario Ministry of Agriculture, Food and Rural Affairs, and Ontario Ministry of Environment. Furthermore, this research was supported by infrastructure grants from the Canadian Foundation for Innovation and Ontario Innovation Trust to Dr. M. Sawada of the Laboratory for Applied Geomatics and GIS Science at the University of Ottawa.

Special thanks to David, for your support and your contagious motivation for soils science. Thanks to Mike for his patience and his understanding. Thanks to the Geography department.

I wish to acknowledge the great help of Mark Edwards, from Agriculture and Agri-Food Canada for his advice and for helping me become more perfectionistic in field site preparation and general sampling.

Merci à Jean Bjornson pour son amitié, les nombreuses discussions et les petites questions qui m'ont souvent remise sur la bonne voie.

Merci à mes colocs, d'aujourd'hui et d'hier, pour tous les bons moments, les encouragements et les plaisirs de la vie. Merci à ma famille pour avoir cru en moi.

Thanks to everyone who helped me grow up... For all challenges and frustrations.

Table of contents

List of tables

List of figures

Chapter 1

Towards Regionally-Based Indicators of Soil Infiltrability: Eastern Ontario Canada

Understanding contaminant transport in the vadose, or unsaturated zone, is required to develop and implement many agriculturally-related best-management practices. Often, regional and/or landscape indicators of contamination risk are required to inform the policy-making processes. Agriculture and Agri-Food Canada (AAFC) are developing a suite of agri-environmental indicators to assess how well agricultural systems manage and conserve natural resources and how compatible they are with the natural systems and processes in the broader environment (AAFC, 2005). The present research is embedded within this scheme of thought. The importance of valid and reliable indicators of macropore flow occurrence, and therefore surface and groundwater contamination potential is well recognized. The overall purpose of this study is to propose potential regional indicators of macropore flow to be used as a tool to assess the state of impacts of any agricultural practice. Figure 1.1 is a graphic representation of the conceptual model underlying this research project.

The central research question of this work is: what relationships exist among soil crop/management and physical/structural variables and soil hydraulic properties? The answer to this question is fundamental for predictive models of contaminant transport, for best-management practices and agricultural policy formulation in Canada.

1.1 INTRODUCTION

Once known as a renewable resource, water is now facing scarcity and severe contamination threats. Internationally, there is a growing awareness concerned with the Earth's water situation (UN-Water, 2006). For example, 2003 was decreed the

International year of Freshwater by the United Nations. Kofi Annan launched the "Decade of water for life" in New York, on April 21, 2005 stating that water is crucial for sustainable development, including the preservation of our natural environment and the alleviation of poverty and hunger (UN-Water, 2006). In the Canadian context, the tragedy of Walkerton in 2000 resulted in growing concerns about soil/water contamination from agricultural activities (AAFC, 2003). Since then, there has been political resolve to better inform policy-making regarding surface and groundwater protection and preservation (AAFC, 2003).

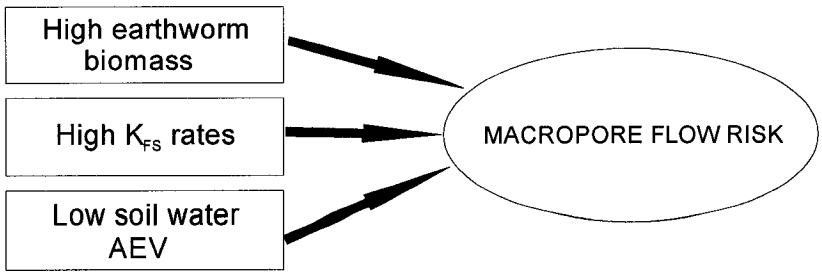


Figure 1.1 Macropore flow risk incidence indicators conceptual model

Hydraulic properties of soils are highly variable and difficult to comprehend and interpret for non-experts (Arya et al., 1999; Papchepsky et al., 2006). Therefore, interpretable and easy-to-understand indicators of high infiltrability and contamination risks would be of considerable use in order to assess the potential environmental impacts of agricultural activities (van Genuchten, 1980; Heppell et al., 2002).

As part of this initiative, earthworms are considered to be important macropore architects. Therefore, being able to estimate their biomass based on easily-interpreted criteria and spatially continuous parameters would be beneficial to contamination predictions. First, high earthworm biomass is associated with a greater tendency for vertical gravity flow since the diameter of worm burrows are too wide to allow capillarity action and subsequent flow. Second, rapid infiltration represents the phenomenon of macropore flow and associated conditions, and third, low soil saturated air-entry tensions represent 'larger' diameter continuous pore networks, allowing water to move freely and rapidly.

The first part of this thesis consists of an exploration of the relationships and trends among *in situ* soil physical, crop, management properties, hydraulic properties and earthworm biomass based on Classification and Regression Tree (CART) methodology. Modeling efforts are based on field collected data. The results are compared with currently available indicators (generally digital records and survey data).

1.2 SOIL HYDROLOGY

Water and soil are interrelated: soil hydraulic properties are of critical importance for the design and monitoring of irrigation and drainage systems, manure impoundments, septic tanks, canals and reservoirs, sanitary landfills, and many other agricultural, industrial and environmental installations (Bouwer, 1978; McKeague et al., 1982; Reynolds, 1993;

Reynolds and Zebchuk, 1996; Reynolds et al., 2000). Field measurements of such properties are also useful for predicting rainfall infiltration and runoff, and for characterizing changes in soil macrostructure due to changing land management practices (Puckett et al., 1985; Gregorich et al., 1993; Reynolds, 1993; Reynolds and Zebchuk, 1996; Reynolds et al., 2000; Heppell et al., 2002; Bagarello and Sgroi, 2004; Papchepsky et al., 2006).

Before the 1980s, water movement in soils was strictly calculated by models driven and constrained by soil moisture tension (Edwards et al., 1993). Beven and Germann (1982) showed, however, that both local and field scale heterogeneities must be accounted for in hydrologic processes. Continuous macropores in unsaturated soil allows water to move freely and rapidly, causing short-circuiting, a phenomenon also referred to as macropore or preferential flow (Bouma et al., 1981). Macropore flow is gravity-driven. Therefore, it represents the highest risk for surface and groundwater contamination. Macropore flow is recognized to be a well-established phenomenon for a wide range of soil types (Beven and Germann, 1982; Wilson and Luxmoore, 1988; Heppell et al., 2002) adding complexity to our understanding of soil-water processes. Many factors have been identified as important macropore flow precursors. One of the most influential variables is the presence of vertical earthworm (nightcrawlers in particular) burrows (Beven and Germann, 1982; Edwards et al., 1988; Shipitalo and Gibbs, 2000; Shipitalo et al., 1997, 2000, 2004). There is increasing evidence that the transport of agrochemicals, such as fertilizers and herbicides, through such pathways are an important mechanism by which diffuse pollution of surface and groundwater can occur (Bouma et al., 1981; Kladivko et al, 1991; Edwards et al., 1993; Heppell et al., 2002).

Due to the presence of spatial heterogeneities, soil hydraulic properties are highly spatially variable (van Genuchten, 1980; Arya et al., 1999; Topp et al., 1980; Bagarello and Sgroi, 2004), sensitive to sample size and sampling method (Topp et al., 1980; Reynolds et al., 2000), costly and time consuming to measure (van Genuchten, 1980; Topp et al., 1980; Arya et al., 1999; Lin et al., 1999; Papchepsky et al., 2006). For all of these reasons, measuring soil hydraulic properties over large areas is often considered impractical. The willingness to go out in the field and rigorously measure such properties is usually overcome by financial and managerial limitations. Therefore, more convenient approaches to acquire information about soil hydraulic properties are highly desirable. Consequently, considerable efforts have been devoted to the indirect estimation of soil hydrology and hydraulic behaviour (Bouma, 2006).

The indirect estimation of soil hydraulic characteristics involves the use of pedotransfer functions (PTFs). Hydraulic properties may encompass many possible variables but pedotransfer functions are primarily used for estimating water retention curves and saturated hydraulic conductivity (Papchepsky et al., 2006). The hydraulic properties for which indicators will be derived in this research project are: (1) field saturated hydraulic conductivity (K_{fs}), a common target variable estimated using PTFs; and (2) soil saturated air-entry tensions (AEV), a parameter that is related to the soil water desorption curve that has not yet been exhaustively studied. K_{fs} represents *in situ* measurement of the steady-state infiltration rate (recharge) required to maintain a steady depth of water

within the porous medium (soil) (Reynolds, 1993). K_{fs} measurements also provide information about the matric flux potential (Φ_m), which is related to sorptivity. Sorptivity is an indicator of the capillarity of an unsaturated medium; i.e., its ability to absorb water (Reynolds, 1993). Generally speaking, the magnitude of Φ_m depends on the texture and structure of the porous medium and on the amount of water already present within it at the time of measurement (Reynolds, 1993). An 'α' value is used to represent the texture/structure categories (Reynolds and Elrick, 1990; Reynolds, 1993) and can be derived from K_{fs}.

The soil air-entry tension (AEV) is evaluated as an indicator of continuous pore network based on tension (cm water) and translated to minimum continuous pore size (Fallow and Elrick, 1996). AEV is considered valuable for assessing macropore flow incidence since soil structure is a crucial factor in estimating soil water infiltration rates in macropore flow regions (Lin et al., 1999).

Direct measurements of spatially continuous soil properties such as soil texture, bulk density and porosity can aid in building PTF functions to estimate the aforementioned soil hydraulic properties. Additionally, distinct class variables like soil structure and residue coverage should also provide valuable estimates of soil hydraulic properties.

The utility of pedotransfer functions originates in the multiple uses of hydraulic properties (Papchepsky et al., 2006) as previously mentioned. An example of such use is predicting water and contaminant movement through the vadose zone. Pedotransfer functions empirically predict hydraulic properties by relating simple soil characteristics that are found in soil surveys like particle size and bulk density, to more complex parameters that are used in modeling and that are relatively difficult to measure (Bouma and van Lanen, 1987; Bouma, 1989; van Genuchten and Leji, 1992; Arya et al., 1999; Lin et al., 1999; Bouma, 2006). More recently, hydropedology research has focused on the validity of PTFs in macropore flow regions (Lin et al., 1999). This new trend brought segregation between classical PTF and structural PTF. Structural PTFs consider soil structural information to be crucial in characterizing hydraulic behavior in macropore flow regions (Lin et al., 1999; Papchepsky et al., 2006). Since soil structure is greatly affected by burrowing earthworms (Shaw and Pawluck, 1986; Shipitalo and Protz, 1987; Kladivko, 2001; Tomlin and Fox, 2003), this thesis explores the possibility of including earthworm biomass numbers in PTFs.

1.3 EXPLORING THE RELATIONSHIPS AMONG SOIL STRUCTURE AND HYDRAULIC PROPERTIES

The methods utilized in this thesis to explore the relations among soil physical/crop/management parameters and hydraulic properties derive from the field of 'data mining'. Relationships uncovered by data mining techniques could serve as hypotheses in order to develop regional indicators of high infiltrability and associated soil properties. Data mining is loosely defined as a mathematical or statistical analysis of data that aims to uncover relations between variables that provide previously unknown but

useful information (Frawley et al., 1992; Hand et al., 2001). Data mining is heuristic, as opposed to the hypothesis-based initiative of variable selection. However, the nature of the relations uncovered are data-driven, and while often following a scientific hypothesis the relations uncovered may be hypothesis-forming or call for modification of the interrogator's world view.

The data mining method used to indirectly estimate soil hydraulic properties and earthworm biomass is a technique called Classification and Regression Trees (CART). Since numerical data were gathered in the field, regression trees were designed to identify existing relationships among all measured variables.

A decision tree is a predictive model derived from the data used to encode the distribution of the target variable (e.g. earthworm biomass) in terms of the predictor attributes (independent variables: tillage, etc.). A decision tree describes a tree structure wherein 'leaves' (nodes) represent regressions and 'branches' represent conjunctions of features (splitting criteria) that lead to those regressions (Menzies and Hu, 2003).

The tree is 'grown' by splitting the feature space (dataset) into homogeneous subsets according to *If-Then* conditions based on the independent variables' value (Hastie et al., 2001). The dataset is split into homogeneous subsets (nodes) according to a specific splitting criterion (independent variable's threshold value), minimizing both variance and model complexity (Breiman et al., 1984; Steinberg and Colla, 1997). The binary splitting process is repeated on each derived subset (node) in a recursive manner (Breiman et al., 1984; Steinberg and Colla, 1987). The top node represents all of the dataset (feature space). The regression tree algorithm finds the best way to explain the dependent variable's (e.g. earthworm biomass) variations with independent data (e.g. tillage) (Hastie et al., 2001). Right and left 'child nodes' are then derived from this splitting process. To follow the given example, all samples characterized by conventional tillage would merge to the right node and the remaining ones (no-till field sites) to the left node.

CART is one amongst many data mining techniques available to uncover relationships within large and complex datasets. CART was judged a pertinent method to discover relationships among soil physical/crop/management properties, earthworm biomass and soil hydraulic properties because results are simple to understand and interpret. No data transformations are needed, no data distribution is assumed and CART is able to use nominal and categorical data (Hastie et al., 2001).

Better understanding of field scale macropore flow is needed in order to address agricultural best-management practices and also to inform policy and decision makers. In order to do so, intelligible and tangible indicators are required. This project explores existing relationships among soil physical properties, agricultural management parameters, soil hydraulic properties and earthworm biomass. This data exploration was done according to the CART methodology based on field-sampled data. Most of the aforementioned PTF studies used existing data without going back to the field for further verification of the models. In our quest for potential indicators of preferential flow mechanisms even the target variables (K_{fs}, AEV, and earthworm biomass) were site-

specifically measured and not computationally inferred. As such, the central research question in this thesis is: what valid/significant relationships exist among soil crop/management and physical/structural variables and soil hydraulic properties?

1.4 THESIS STRUCTURE

The second chapter of this work is entitled 'Methodology and Data Presentation'. It includes a description of the studied field sites and further details concerning field sampling and laboratory analyses. Chapters 3 and 4 are prepared in manuscript format for scientific publication. Both of these chapters are co-authored with my supervisors and research affiliates. For these potential publications, I undertook the field work, collated data and undertook the laboratory and statistical analyses. Furthermore, I wrote these manuscripts with input and editing and from the co-authors. The first of these manuscripts, Chapter 3 presents indirect estimations of the field saturated hydraulic conductivity (K_{fs}) and saturated soil air-entry values (AEV) based on field-measured soil physical properties and management parameters. The first objective of Chapter 3 is to explore and uncover important relationships among soil physical properties, management parameters and hydraulic properties using CART. CART is used to generate hypotheses regarding those factors that are the most influential on regional macropore flow. The ultimate goal of Chapter 3 is to identify potential regional indicators of soil hydraulic properties, namely K_{fs} and AEV, in order to be able to assess the impacts of agricultural activities. Chapter 3 will be submitted to a peer-reviewed journal soon. In its current stage it does not meet publication requirements due to length, but offers interesting analyses of *in situ* variable interactions. A slightly modified version of Chapter 4 has been accepted pending revisions in *Applied Soil Ecology*. The paper proposes a model to indirectly predict earthworm biomass according to field sampled soil physical properties, management parameters, K_{fs} and AEV. CART was used in order to identify robust regional indicators of macropore flow incidence based on strong relationships found between soil physical and hydraulic properties, and earthworm biomass. The meaningful relationships uncovered could ultimately be transformed into potential regional indicators of earthworm biomass, and therefore, of macropore flow incidence. Regional indicators of macropore flow incidence require further analysis if they are to become a tangible tool for policy makers but this is a first step in this enterprise. The last chapter summarizes the findings of the thesis.

1.5 REFERENCES

AAFC 2003: Canadian Adaptation and Rural Fund; Canada – Ontario Research Development Program, Performance Report, 2002-2003. http://www.agr.gc.ca/acaaf/card/pdf/councils.aac.cordp.03.pdf

AAFC 2005: Environmental sustainability of Canadian Agriculture: Agri-Environmental Indicator Report Series – report #2 http://www.agr.gc.ca/env/naharp-pnarsa/index_e.php?page=aei

Arya, L.M., Leji, F.J., Shouse, P.J. and van Genuchten, M., Th. 1999. Relationship between the hydraulic conductivity function and the particle-size distribution. Soil Science Society of America Journal 63: 1063-1070

Bagarello, V. and Sgroi, A. 2004. Using the single-ring infiltrometer method to detect temporal changes in surface soil field-saturated hydraulic conductivity. Soil and Tillage Research 76: 13-24

Beven, K. and Germann, P. 1982. Macropores and water flows in soils. Water Resources Research 18: 1311-1325

Bouma, J., Dekker, L.W., and Muilwijk, C.J. 1981. A field method for measuring short-circuiting in clay soils. Journal of Hydrology 52: 347-354

Bouma, J. and van Lanen, J.A.J. 1987. Transfer functions and threshold values: from soil to land qualities. Pp. 106-110. In: Beek, K.J. et al. (Ed.) Quantified land evaluation. International Institute Aerospace Survey and Earth Science. ITC Publication 6: 106-110

Bouma, J., 1989. Using soil survey data for quantitative land evaluation. In: Stewart, B.A. (Ed.), Advances in soil science, vol. 9, Springer Verlag, New York, pp. 177-213.

Bouma, J. 2006. Hydropedology as a powerful tool for environmental policy research. Geoderma 131: 275-286

Bouwer, H. 1978. Groundwater hydrology. McGrawHill, Toronto, pp. 123-124.

Breiman, L., Freidman, J., Olshen, R., and Stone, C. Classification and Regression trees. Pacific Grove: Wadsworth, 1984

Edwards, W.M., Shipitalo, M.J., and Norton, L.D. 1988. Contributions of macroporosity to infiltration into a continuous corn no-tilled watershed: implications for contaminant movement. Journal of Contaminant Hydrology 3: 193-205

Edwards, W.M., Shipitalo, M.J., and Owens, L.B. 1993. Gas, water and solute transport in soils macropores: a review of methodology. Geoderma 57: 31-49

Fallow, D.J., and Elrick, D.E. 1996. Field measurement of air-entry and water-entry soil water pressure heads. Soil Society of America Journal 60: 1036-1039

Frawley, W., Piatetsky-Shapiro, G., and, Matheus, C. 1992. Knowledge Discovery in Databases: An Overview. AI Magazine, Fall 1992, pp. 213-228

Gregorich, E.G, Reynolds, W.D., McGovern, M.A., Culley, J.L.B., and Curnoe, W.E. 1993. Changes in some physical properties with depth in a structurally degraded soil under no-tillage. Soil and Tillage Research 26: 289-299

Hand, D., Mannila, H., Smyth, P. 2001. Principles of Data Mining. MIT Press, Cambridge, MA, 2001.

Hastie, T., Tibshirani, R., and Friedman, J. 2001. The elements of statistical learning: Data mining, inference and prediction, Springer Series in Statistics, New York, pp. 266-278.

Heppell, C.M., Worrall, F., Burt, T.P., and Williams, R.J. 2002. A classification of drainage and macropore flow in an agricultural catchement. Hydrological Processes 16: 27-46

Kladivko, E.J., Van Scoyoc, G.E., Monke, E.J., Oates, K.M., and Pask, W. 1991. Pesticide and nutrient movement into subsurface tile drains on a silt loam soil in Indiana. Journal of Environmental Quality 20(1): 264-270

Kladivko, E.J. 2001. Tillage systems and soil ecology. Soil & Tillage Research 61:61-76

Lin, H.S., McInnes, K.J., Wilding, L.P., and Hallmark, C.T. 1999. Effects of soil morphology on hydraulic properties II. Hydraulic pedotransfer functions. Soil Science Society of America Journal 63: 955-961

McKeague, J.A., Wang, C., and Topp, G.C. 1982. Estimating saturated hydraulic conductivity from soil morphology. Soil Science Society of America Journal 46: 1239-1244

Menzies, T. and Hu, Y. 2003. Data mining for very busy people. IEEE Computer, October 2003, pp. 18-25

Puckett, W.E., Dane, J.H., and Hajek, B.F. 1985. Physical and mineralogical data to determine soil hydraulic properties. Soil Science Society of America Journal 49: 831-836

Papchepsky, Y.A., Rawls, W.J., and Lin, H.S. 2006. Hydropedology and pedotransfer functions, Geoderma 131:308-316

Reynolds, W.D. 1993. Saturated hydraulic conductivity: field measurements; chapter 56 in: Carter, M.R. (Ed.) Soil Sampling and Methods of Analysis, Lewis Publishers, Boca Raton, Florida, 823 pages. Saturated Hydraulic conductivity – Field measurements, pp: 599-613

Reynolds, W.D., and Zebchuk, W.D. 1996. Hydraulic conductivity in a clay soil: two measurement techniques and spatial characterization. Soil Science Society of America Journal 60: 1679-1685

Reynolds, W.D., Bowman, B.T., Brunke, R.R., Drury, C.F., and Tan, C.S. 2000. Comparison of tension infiltrometer, pressure infiltrometer, and soil core estimates of saturated hydraulic conductivity. Soil Science Society of America Journal 64: 478-484

Shaw, C. and Pawluk, S. 1986. The development of soil structure by *Octolasion tyrtaeum, Aporrectodea turgida* and *Lumbricus terrestris* in parent materials belonging to different textural classes. Pedobiologia 29: 327-339

Shipitalo, J.M., and Protz, R. 1987. Comparison of morphology and porosity of a soil under conventional and zero tillage. Canadian Journal of Soil Science 67: 445-456

Shipitalo, M.J., Edwards, W.M., and Owens, L.B. 1997. Herbicide losses in runoff from conservation-tilled watersheds in a corn-soybean rotation. Soil Science Society of America Journal 61: 267-272

Shipitalo, M.J., Dick, W.A., and Edwards, W.M. 2000. Conservation tillage and macropore factors that affect water movement and the fate of chemicals. Soil and Tillage Research 53: 167-183

Shipitalo, M.J., and Gibbs, F. 2000. Potential of Earthworm burrows to transmit injected animal wastes to tile drains. Soil Science Society of America Journal 64: 2103-2109

Shipitalo, J.M., Nuutinen, V., and Butt, K.R. 2004. Interaction of earthworm burrows and cracks in a clayey, subsurface-drained, soil. Applied Soil Ecology 26: 209-217

Steinberg, D., and Colla, P. CART – Classification and Regression Trees. Sand Diego, CA. Salford Systems, 1997

Tomlin A.D. and Fox C.A. 2003. Earthworms and agricultural systems: Status of knowledge and research in Canada. Canadian Journal of Soil Sciences 83:265-278

Topp, G.C., Zebchuk, W.D., and Dumanski, J. 1980. The variation of in-situ measured soil water properties within soil map units. Canadian Journal of Soil Science 60: 497-509

United Nations (UN) Portal for Water. Coping with water scarcity: A strategic issue and priority for system-wide action. UN-Water Thematic Initiatives, August 2006 ftp://ftp.fao.org/agl/aglw/docs/waterscarcity.pdf

van Genuchten, M.Th. 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Science Society of America Journal 44: 892-898

van Genuchten, M.Th., and Leji, F. 1992. On estimating the hydraulic properties of unsaturated soils. pp. 1-14 In M.Th. In: van Genuchten et al. (Ed.) Proc. Int. workshop on indirect methods of estimating the hydraulic properties of unsaturated soils. 11-13 Oct., 1989. U.S. Salinity laboratory and Department of Soil and Environmental Science, University of California, Rivers.

Wilson, G.V., and Luxmoore, R.J., 1988. Infiltration, macroporosity, and mesoporosity distributions on two forested watersheds. Soil Science Society of America Journal 52(2): 329-335

Chapter 2

Methodology and Data Presentation

2.1 STUDY AREAS

Sampling was performed in the region extending from Kemptville, ON. to St-Isidore, ON. Some sampling also occurred in southern Ontario, in Woodstock, at the University of Guelph's Research Station. The region of study was chosen because it experiences extensive agriculture. Ontario accounts for over 25% of Canada's agricultural production (Statistics Canada). In Canada, 36 395 150 hectares of land are in crops of which, 3 656 705 hectares are located in Ontario (Statistics Canada).

2.2 MANAGEMENT

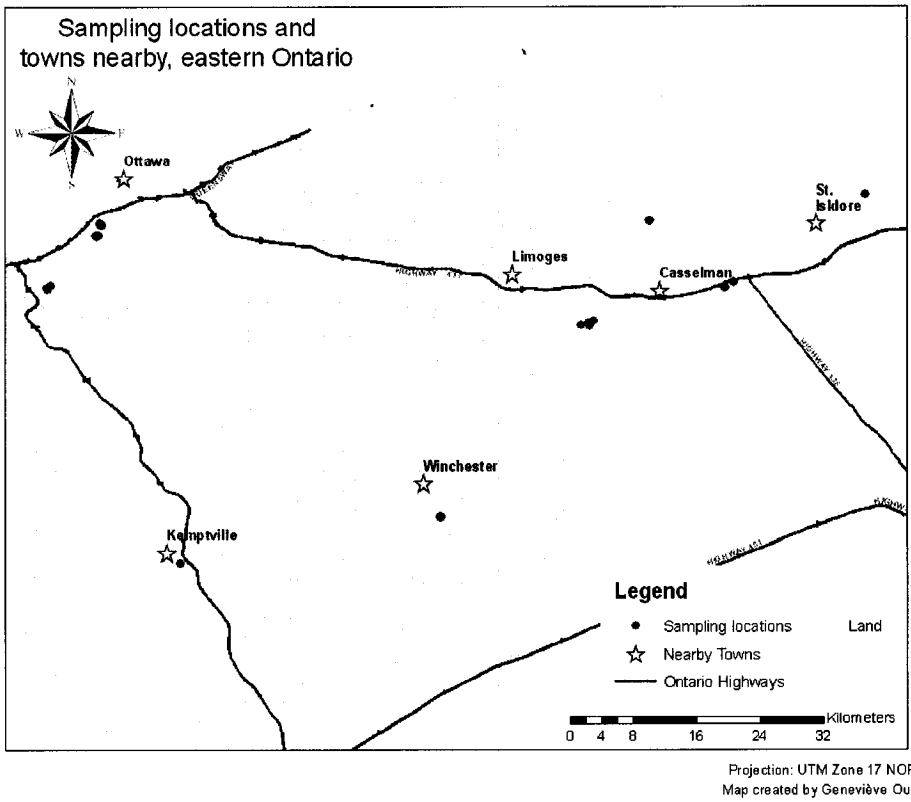


Figure 2.1 Regional map

Seventy-eight (78) field sites were analysed in this study. See Figure 2.1 for the regional map. Fifty-nine (59) of them are conventionally tilled and 19 underwent long-term-no-till. Long-term-no-till means that the field in question has not undergone tillage of the soil matrix for at least the past 5 years. Conventional tillage implicates fall season ploughing and springtime secondary tillage. "Tillage" is treated as a binary variable; no-till field sites are coded "0" and tilled sites are coded "1".

Crop type is also included as a variable in the analysis. The 4 field sites that had soy beans planted on at the time of sampling were coded "0". Standing corn (60 field sites) was coded "1" and bare fields (14) were coded "2". The cropping system's codes are nominal differences of kind and therefore, crop codes were not included in the Spearman's Rank correlation analyses.

Some fields, mostly the ones on the Central Experimental Farm, exhibited different management practices and interesting texture gradients. Every field plot was therefore considered as a distinctive field site.

2.3 FIELD AND LABORATORY METHODOLOGY

2.3.1 Sampling locations

The visited field sites were chosen for their proximity. Property access was a limiting factor for choosing sampling locations since access to the land was denied many times. Additionally some theoretical assumptions mandate minimal soil disturbance for valid infiltration parameters to be measured. This project aims at up-scaling soil macroporosity and the related soil's infiltrability in Eastern Ontario. The authors are aware that the sampling was not totally random and that the spatial representivity of this study may be impaired due to the sampling strategy.

2.3.2 Texture

The sampling strategy was designed to include soils with a wide range of texture. Heavier soils had clay contents up to 36%. Such soils were located on the Central Experimental farm in Ottawa, ON. Lighter soils, on the other hand, were sand dominated up to 95% and were located on a field owned by the Canadian Food Inspection Agency, called Field 8.

Grab surface soil samples (15 cm diameter by 15 cm depth), were collected in the center of each sampling area (usually 15 m by 10 m) for particle size distribution and organic matter analyses. Two aliquots from the grab sample served for texture determination. First, the organic matter was removed from both aliquots with hydrogen peroxide (H_2O_2), according to Sheldrick and Wang (1993). Particle fractions greater than 63 µm were then determined by wet sieving on one sub-sample and from the second sub-sample, particle fractions smaller than 63 µm were determined using a Lecotrac LT-100 Particle Size Analyzer (PSA) (Sedimentology Laboratory, G.S.C., Ottawa, ON). Both particle size analysis methods were combined (as percentages) to generate total textural distribution of soils. When wet-sieving, all the fine particles (silts and clays) were washed out. This "lost weight" provides the proportional weight occupied by particles smaller or equal to 63 µm. This proportional weight was translated into a percentage of the initial, complete sample and partitioned using the PSA results generated at the Sedimentology Laboratory (GSC) in order to calculate the exact silt and clay proportions. Here is a concrete example:

> Let's consider sample 114 that have a dry pre-sieving weight of 90.48 g. After wet-sieving, the remaining dried sediment (sands and gravels) had a weight of 21.59 g and 0.06 g respectively. By a simple subtraction, we can conclude that this sample contained 68.83 g of fine particles (silts and clays). Since we were not to include gravels in the analysis, their weight was subtracted from the initial mass. Dividing (90.48 g – 0.06 g) by the fines' weight (68.83 g) we obtained 76.12%. The same operation indicated 23.87% sand.
>
> The PSA identified 41.47% of clay and 58.53% of silt. The PSA analysis excluded all particles with a diameter greater than 2000 µm. To know which fraction of the complete sample was represented by clay, for example, we multiplied 41.47% by the factor representing the fines (76%), ending up with 31.55% of clay, 44.58% of silt and 23.87% of sand.

2.3.3 Organic matter

Two more aliquots were sub-sampled from the aforementioned grab sample for organic matter content determination. The soil organic matter content was estimated using loss-on-ignition (LOI) (Heiri et al., 2001). A constant volume of wet sample was taken using a small glass vial. A standard soil sample (ECSS2) accompanied the samples to the muffle furnace.

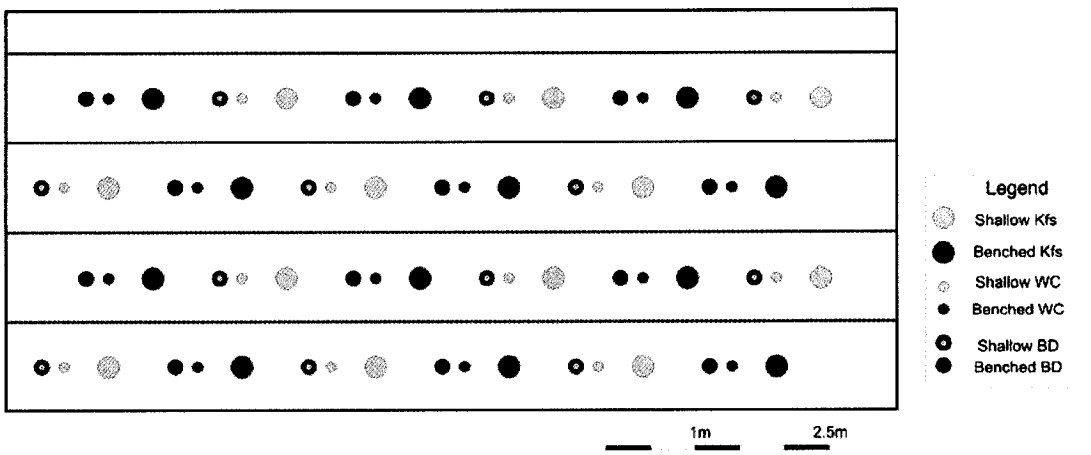


Figure 2.2 Typical sampling scheme

2.3.4 Bulk density

Soil cores (5.5 cm diameter and 10 cm length) were used to measure soil bulk density, as prescribed by Culley (1993). Bulk density samples were taken at two different depths: on the soil surface (5 cm) and at 15 cm depth. Bulk densities were sampled near hydraulic conductivity measurement sites as shown in Figure 2.2.

2.3.5 Water content

Soil moisture (%) range was measured with time-domain reflectometry (TDR) technology following Topp's (1993) comments on the subject. Soil water content was not included as an independent predictor of either hydraulic properties (K_{fs} and AEV) nor earthworm biomass because of its transient nature. We are trying to identify macropore flow precursors that are spatially and temporally stable. The correlation coefficients between water content and hydraulic properties (K_{fs} and AEV) are negligible (<0.01). Soil moisture was measured only because it is a necessary parameter for the calculation of the field saturated hydraulic conductivity (K_{fs}).

2.3.6 Field saturated hydraulic conductivity (K_{fs})

K_{fs} was measured for two depths: shallow measurements (5 cm) and benched measurements (15 cm) with a constant head pressure (single-ring 10.5 cm diameter by 6 cm height) infiltrometer, commonly called a pressure infiltrometer (PI). The PI is an apparatus providing *in situ* determinations of field-saturated hydraulic conductivity (K_{fs}) and matric flux potential (Φ_m) in the vadose zone (Reynolds, 1993).

Twelve replicates were realised for each depth, that is to say that 24 K_{fs} measurements were performed at every field site. All K_{fs} measurements were realized within a 1.5 m

distance of the next one in order to minimize the impact of neighbouring wetting fronts on subsequent measurements. As shown on Figure 2.2, the sampling was organized into four rows, with three shallow replicates alternating with three benched measurements.

The first step for every K_{fs} measurement was the site preparation. The soil was dug to the prescribed depth (5 cm or 15 cm) and carefully levelled with a trowel in order to avoid smearing and/or compaction of the soil. Soil smearing and compaction may lead to biased K_{fs} rates. The second step was the ring insertion. This is a crucial step for assessing K_{fs} reading's validity, as short-circuits of soil drainage can be created by inclined ring insertion, for example. In order to maximize consistency, sites were always prepared, and the rings inserted, by the same person. The third and last step before starting the readings was the ponding of the measurement site. Ponding is realized by gently opening the reservoir to allow water to slowly saturate the studied area. Rapid ponding may disrupt initial soil structure and incomplete ponding allows air to enter into the system.

In order to calculate K_{fs} rates a steady-state is assumed; that is to say that in the field, one must complete five consecutive and identical infiltration readings. Multiple head analyses were performed using 10 cm and 40 cm heads. Multiple head analysis provides information on the degree of gravity vs. capillary flow as expressed via the capillary length parameter (α) (Reynolds and Elrick, 1990; Reynolds, 1993).

2.3.7 Air-entry tension (AEV)

The soil saturated air-entry tension was measured directly after reaching steady-state infiltration rates with an auxiliary device attached to the PI called a tensimeter (Fallow and Elrick, 1996; Ball-Coelho et al., 2007). Air-entry tensions can be used as an indicator of continuous pore networks or pore sizes that allows water to enter the saturated soil. The greater the tension, the smaller the continuous pore network radii (Fallow and Elrick, 1996).

Twelve AEV measurements were realized per depth, summing to 24 tension readings per field. The AEV readings were included in the analysis only if all theoretical assumptions were met.

2.3.8 Earthworm biomass

Earthworms were sampled in October 2005 on 43 field sites using the "hot-mustard" method (http://sci.agr.ca/london/faq/worm-ver_e.htm). This method is known to work better for worms that have burrows opening to the soil surface (http://sci.agr.ca/london/faq/worm-ver_e.htm). Metal frames (20 cm large x 40 cm long x 20 cm high) were inserted into the soil to a depth of 15 cm. 5L of hot mustard/water solution, consisting of 17.5g mustard to 1L water (http://sci.agr.ca/london/faq/worm-ver_e.htm) was then poured over the delimited area. After approximately 10 minutes post application, the worms that emerged to the surface were hand collected. The worms were subsequently stored in glass jars containing 30% ethanol until they were cleaned and weighed to determine their wet biomass. For every field, 6 (randomly placed) replicate measurements were taken over a 15 m by 10 m area, representing surface residue coverage ranging from no cover (0%) to full cover (100%).

2.4 PRESENTATION OF DATASETS

2.4.1 KFS & AEV

The first part of the data presentation and description concerns the data analysed in Chapter 3. Table 2.1 describes all of these variables.

Table 2.1 Variables description and units included into KFS and AEV modeling

Variables description and units (codes are categorical values used in statistical analysis)		
Variable	**Units**	**Description**
BD	g cm^{-3}	Soil bulk density
ORG	%	Soil organic matter content
CLAY	%	Soil clay content
SILT	%	Soil silt content
SAND	%	Soil sand content
CROP	#	Type of crop when sampling. Bare field or no new crops (coded 2), corn (coded 1) and soy beans (coded 0)
TILL	#	Conventional tillage (coded 1), no-tillage (coded 0)
STRUCTURE	#	Visual description of soil structure. Homogeneous, cohesive soil (coded 0), blocky, chunky soil (coded 1), weak apparent porosity (coded 2), medium apparent porosity (coded 3), high apparent porosity (coded 4) and unconsolidated structure or structural collapse (coded 5)
VEG	#	Visual description of the vegetal coverage found at measuring sites. Bare soil, no residue (coded 0), roots (coded 1), a little amount of residue such as soy bean leaves and weeds (coded 2) and a thick layer of vegetal residue such as decaying corn (coded 3)
AEV	cmwater	Soil water air-entry values
KFS	cm s^{-1}	Field saturated hydraulic conductivity

The summary statistics will be presented as two distinct datasets 'shallow' and 'benched' since the data was segregated according to depth in the regression tree analysis process. Table 2.2 summarises the descriptive statistics:

Table 2.2 Summary statistics of non-categorical variables involved in prediction modeling of shallow hydraulic properties

Summary statistics of non-categorical variables involved in prediction modelling of shallow hydraulic properties (K_{fs} and AEV)

Variable	N	MIN	1st quartile	Median	Mean	3rd quartile	MAX	STDEV	NA
KFS ($cm\ s^{-1}$)									
Shallow	357	$1.07(10)^{-5}$	$1.95(10)^{-3}$	$7.72(10)^{-3}$	$2.15(10)^{-2}$	$2.32(10)^{-2}$	$2.64(10)^{-1}$	$3.47(10)^{-2}$	4
Benched	420	$2.09(10)^{-6}$	$1.39(10)^{-3}$	$5.10(10)^{-3}$	$1.79(10)^{-2}$	$1.53(10)^{-2}$	$5.27(10)^{-1}$	$4.20(10)^{-2}$	8
AEV (cmwater)									
Shallow	347	-58.12	-18.36	-12.24	-14.05	-8.16	0.0	8.42	14
Benched	406	-44.87	-20.39	-14.28	-15.26	-9.18	0.0	7.40	22
BD ($g\ cm^{-3}$)									
Shallow	353	0.96	1.25	1.38	1.38	1.51	1.80	0.17	8
Benched	420	1.02	1.34	1.46	1.46	1.58	1.86	0.17	8
ORG (%)	78	1.6	4.2	4.8	5.1	6.2	9.0	1.4	0
CLAY (%)	78	1	4	7	11	17	36	9	0
SILT (%)	78	4	43	58	54	67	87	18	0
SAND (%)	78	9	16	27	32	46	95	20	0

The variables ORG, CLAY, SILT and SAND are grab samples reaching depths of 15 cm from the surface; the same value was used for both scenarios, shallow and benched.

The median and the mean statistics describe the data location; that is to say that they measure the central tendency. The MIN, MAX, 1st and 3rd quartiles, along with the STDEV give an estimate of the data dispersion. Histograms (Figures 2.3, 2.4) were built to demonstrate the sampled target variables' distribution.

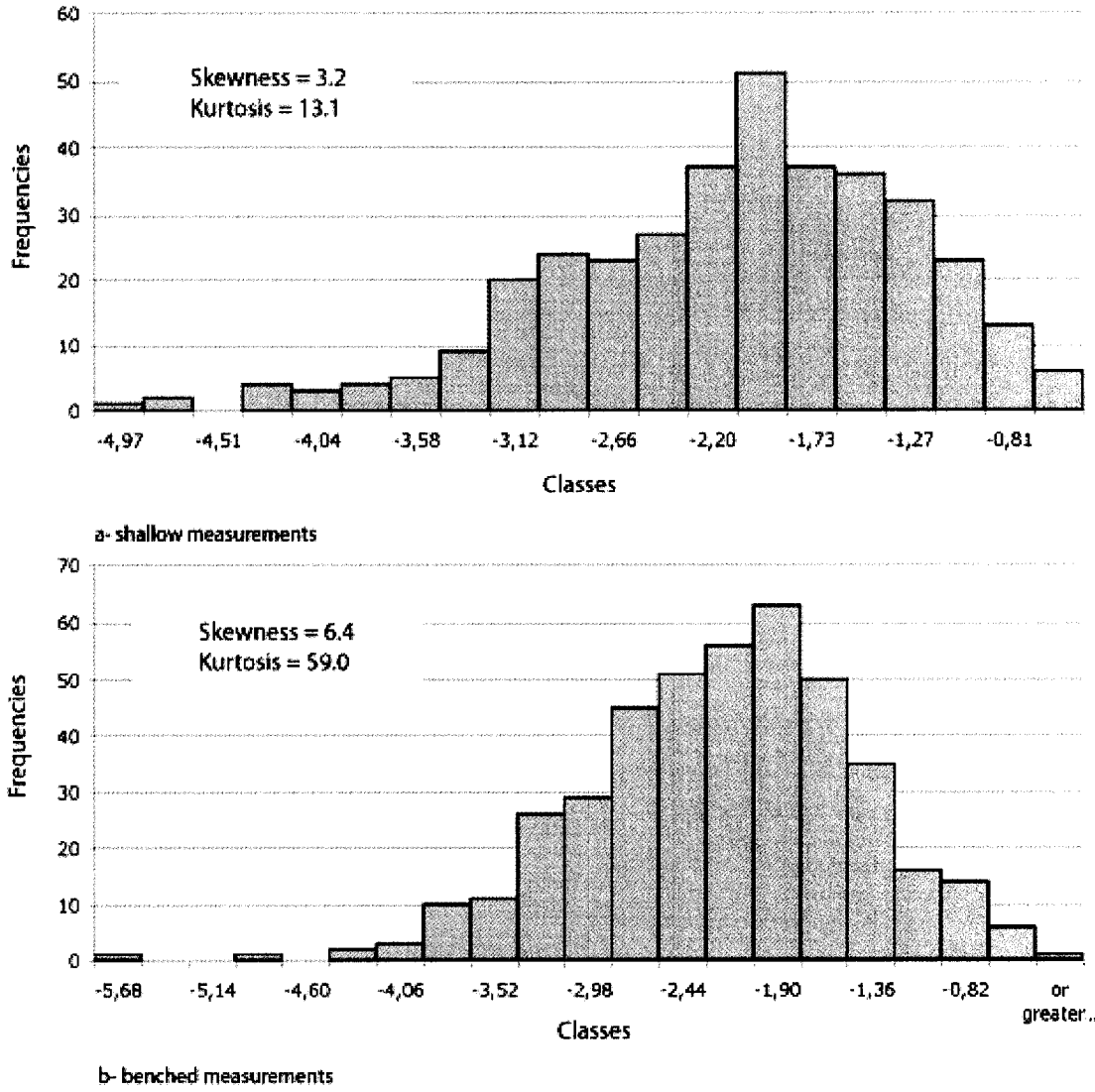


Figure 2.3 Shallow and benched KFS (cm s^{-1}) histograms

Benched KFS have a wider range than shallow KFS. Its standard deviation (STDEV) is also greater but its average is lower than for the shallow measurements. That means that benched infiltrations are usually slower. This can be confirmed by looking at the quartile data: at the 1st quartile limit, shallow KFS are 5 times greater than the benched ones, and at the 3rd quartile limit, they are 1.5 times greater. This is another way of saying that infiltration is usually slower at 15 cm depth than on the surface.

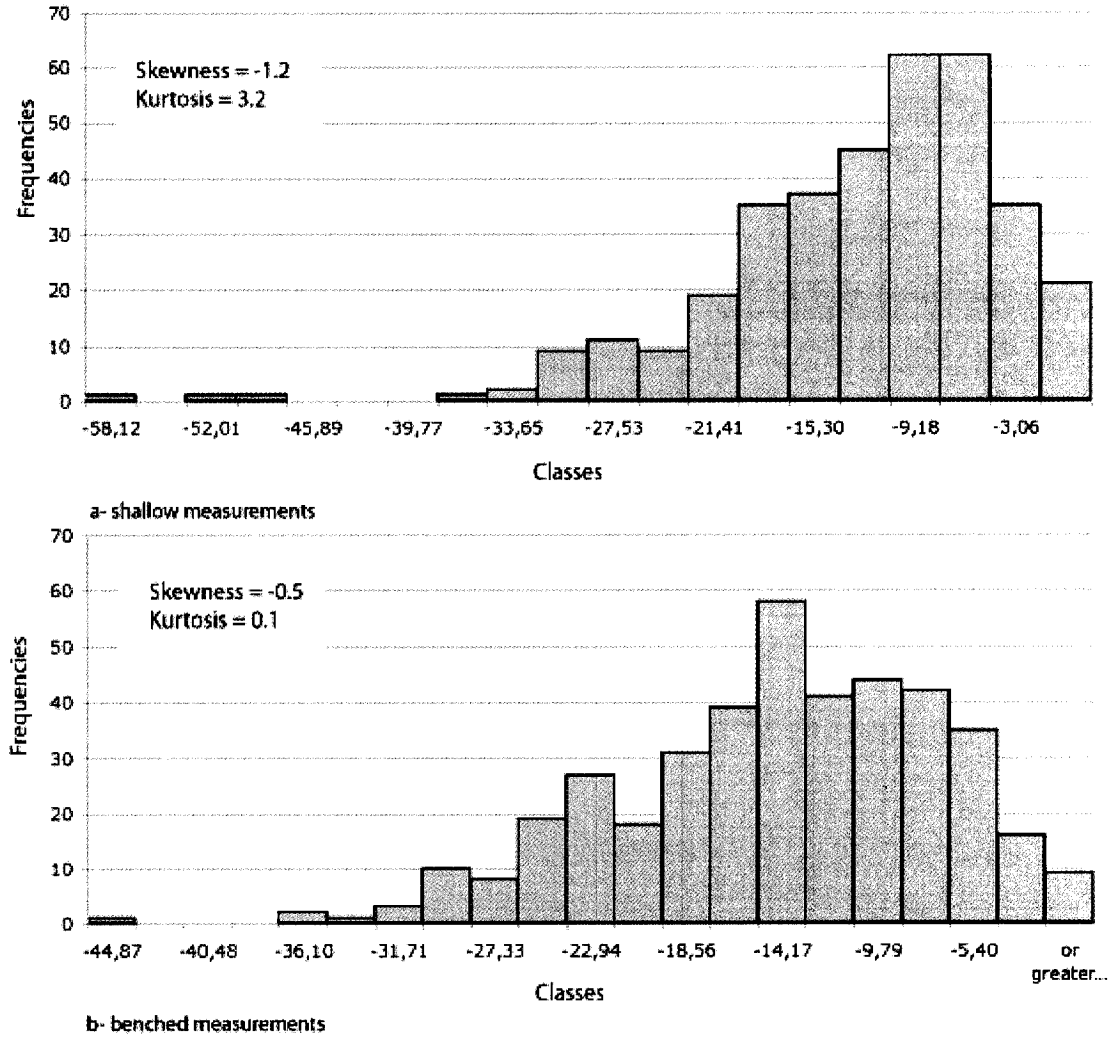


Figure 2.4 Shallow and benched AEV (cm water) histograms

For AEV there is the inverse trend; that is to say that surface measurements have a greater range and a higher standard deviation. This observation supports the hypothesis that surface soil is more heterogeneous and less consolidated than lower soil horizons.

Independent variables' distributions are illustrated with histograms (Figures 2.5, 2.6, 2.7, 2.8, and 2.9).

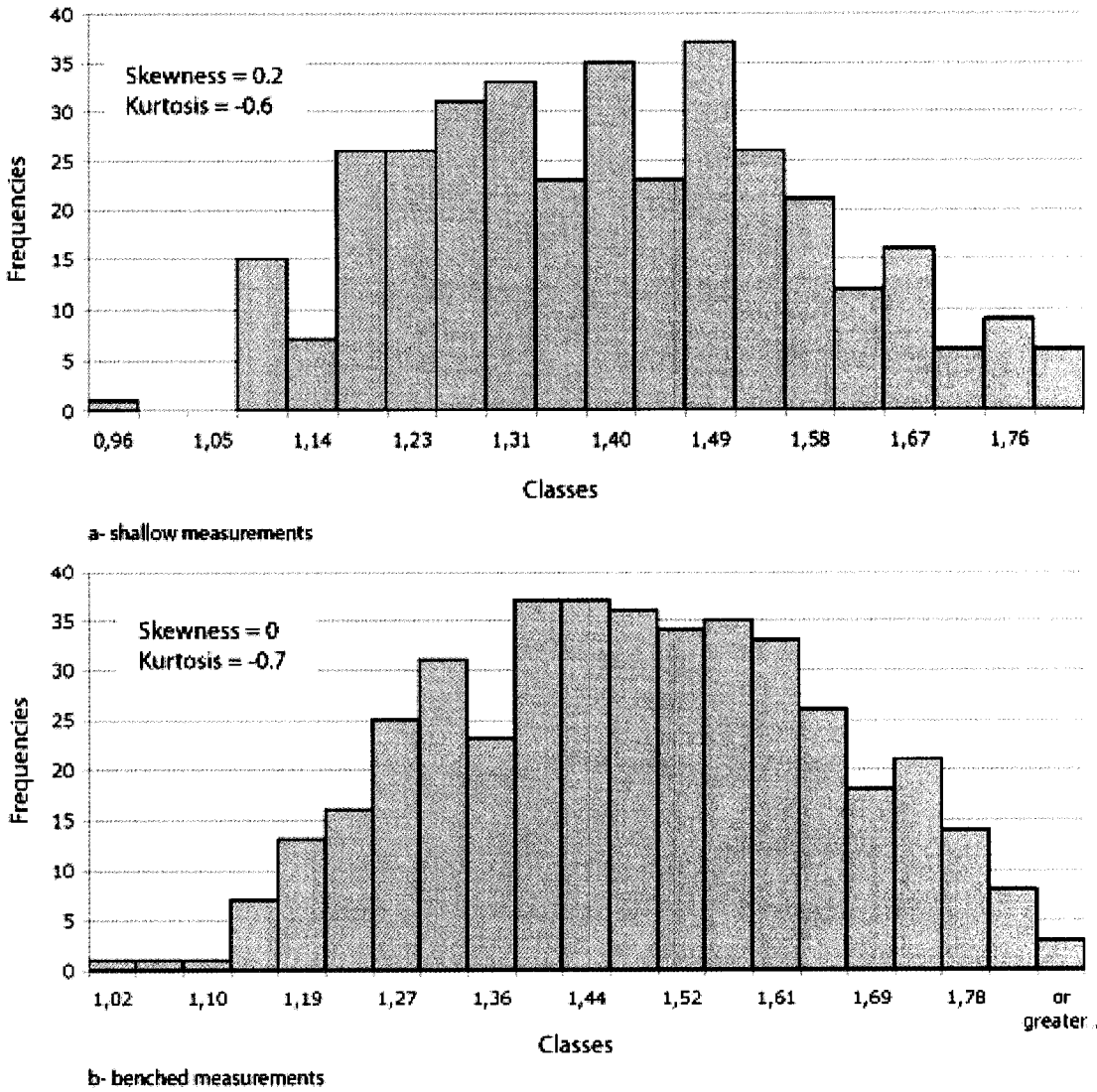


Figure 2.5 Shallow and benched bulk density (g cm^{-3}) histograms

BD dispersion patterns appear similar between shallow and benched measurements. The major distinction is the higher values for benched cores.

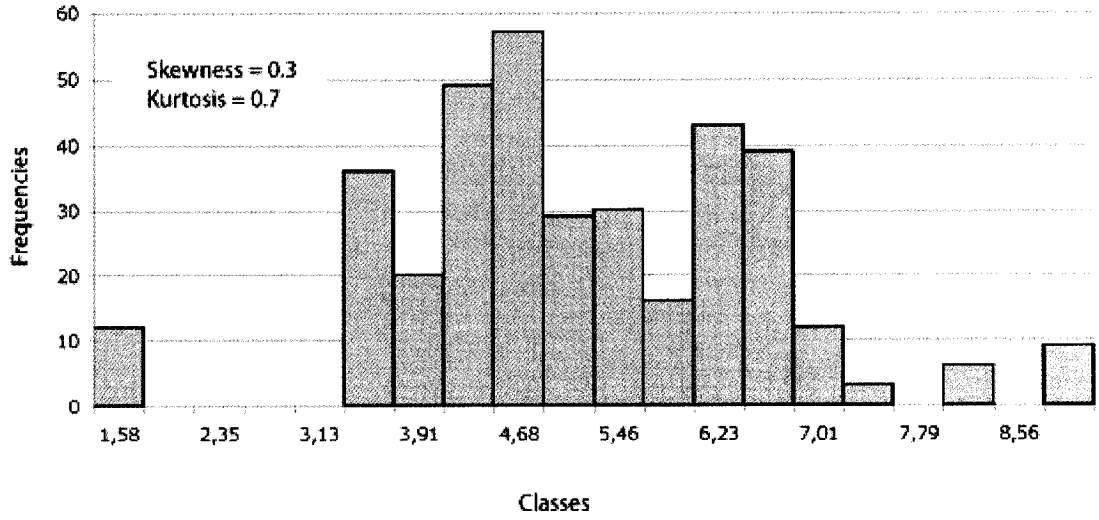


Figure 2.6 Organic Matter (%) histogram

Figure 2.7 Clay (%) histogram

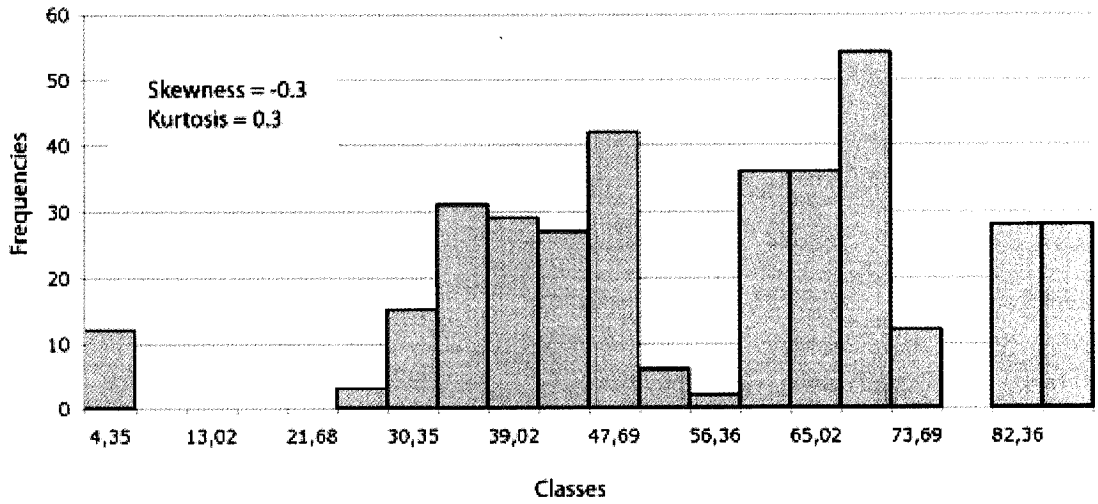


Figure 2.8 Silt (%) histogram

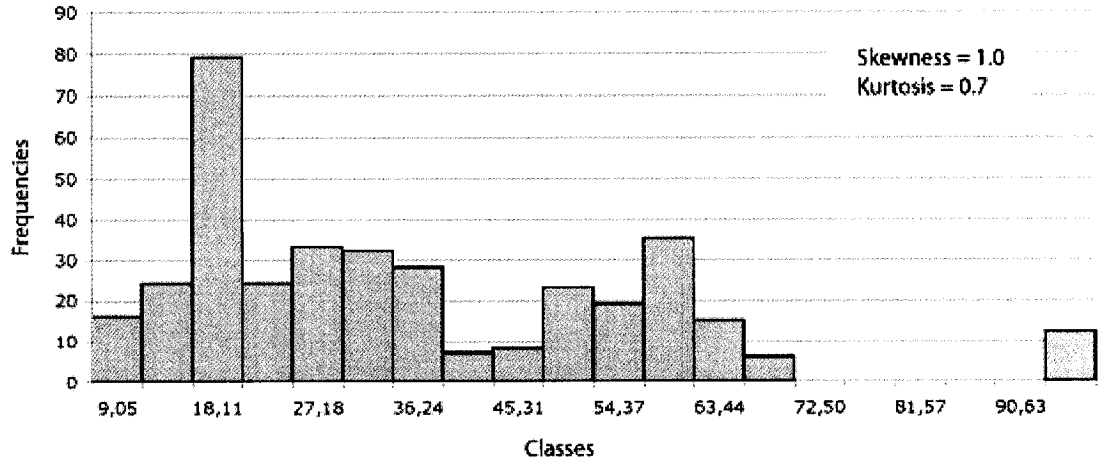
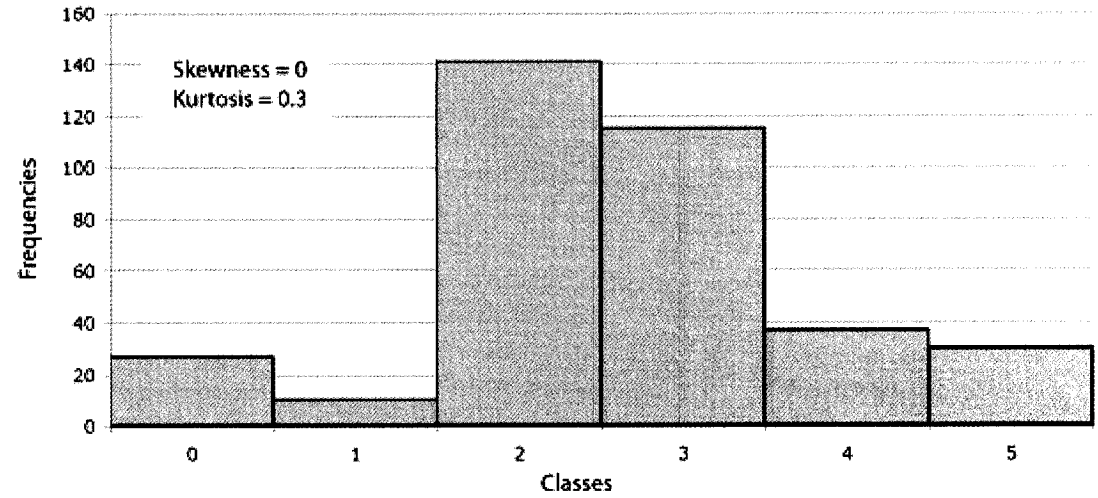


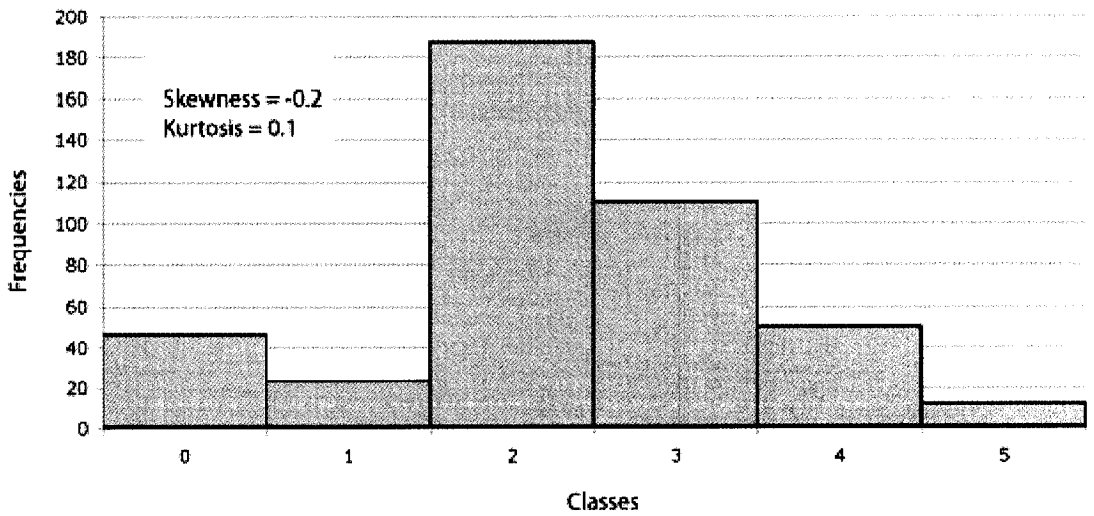
Figure 2.9 Sand (%) histogram

2.4.1.1 Coded variables

Some descriptive characteristics of the measurement sites have been translated into categorical variables, such as the soil's structure and the decaying vegetation embedded within the soil matrix. After PI readings, the 'ring' surface and bottom were scrutinized in order to observe local soil structure (macropores dimensions and number, for example). Comments were also made as to the presence and quantification of decaying organic matter. These observations were translated into ordinal codes (listed in Table 2.2). Figure 2.9 illustrate the frequency of every structure code for shallow and benched study sites respectively:



a- shallow measurements



b- benched measurements

Figure 2.9 Structure code frequencies for shallow and benched measurements.

The structure classification is very similar for both depths. The shape of the curves (for both structure categorical frequencies) exhibits some positive skewness.

Figure 2.10 represents the distribution of the vegetation codes assigned to every measurement site, at surface and at depth, respectively:

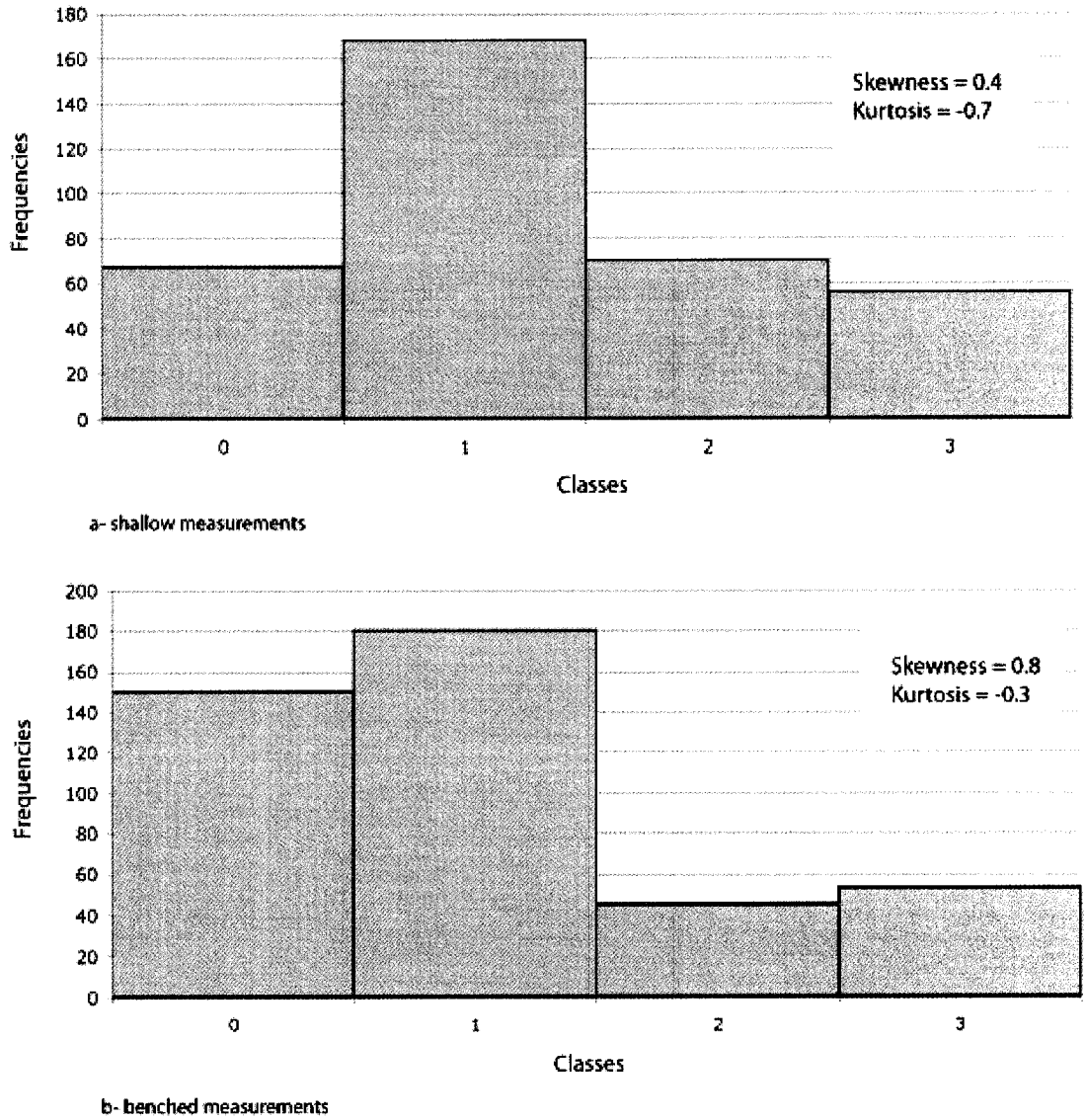


Figure 2.10 Vegetation code frequencies for shallow and benched measurements

In the case of vegetation (organic matter) categories, per depth, they appear to be more different than the structure codes for shallow and benched scenarios. This may be due to the natural decomposition of the organic matter, the work of burrowing worms and the effects of tillage.

Spearman's Rank correlations were performed in order to be able to include coded variables like tillage and structure in the analysis. Spearman's rank is a non-parametric statistic. It assesses how well a monotonic (can be said of functions between ordered sets if they preserve the given order) function describes the relationship between two variables; it can be used for variables measured at the ordinal level. The data are converted into rankings before calculating the coefficients.

2.4.2 Earthworm BIOMASS

All the variables included into the analysis of earthworm biomass modeling are described in the following tables. First, the variables and their units are listed and described in Table 2.3.

Table 2.3 Variable description and units included into BIOMASS modeling

Variable description and units (codes are categorical values used in statistical analysis)		
Variable	**Units**	**Description**
BIOMASS	g	Wet weight of earthworms
BD	g cm^{-3}	Soil bulk density
ORG	%	Soil organic matter content
CLAY	%	Soil clay content
SILT	%	Soil silt content
SAND	%	Soil sand content
CROP	#	Bare field or no new crop (coded 2), corn (coded 1) and soy beans (coded 0).
TILL	#	Conventional tillage (coded 1), no-tillage (coded 0)
RESIDUE	#	Amount of residue covering the soil: [0, 5%[(coded 0), [5, 25%[(coded 1), [25, 50%[(coded 2), [50, 75%[(coded 3), [75, 100%[(coded 4) And 100% (coded 5).
AEV	cmwater	Soil water air-entry values
KFS	cm s^{-1}	Field saturated hydraulic conductivity

Table 2.4 presents the summary statistics for non-categorical data.

Table 2.4 Summary statistics of non-categorical variables involved in earthworm biomass prediction modeling

Summary statistics of non-categorical variables involved in prediction modelling Of earthworm biomass

Variable	N	MIN	1st Quartile	Median	Mean	3rd Quartile	MAX	STDEV	NA
BIOMASS (g)	243	0.0	2.4	6.2	9.0	13.0	44.6	8.8	0
KFS (cm s^{-1})	231	$6.62(10)^{-5}$	$5.63(10)^{-3}$	$9.29(10)^{-3}$	$1.41(10)^{-2}$	$1.61(10)^{-2}$	$5.69(10)^{-2}$	0.01	12
AEV (cmwater)	243	-22.77	-17.71	-9.86	-11.86	-7.82	-3.74	5.19	0
BD (g cm^{-3})	243	1.12	1.31	1.41	1.43	1.51	1.71	0.14	0
ORG (%)	43	1.6	4.4	5.2	5.5	6.6	9.0	1.5	0
CLAY (%)	43	1	5	18	15	21	23	9	0
SILT (%)	43	4	40	49	54	67	87	16	0
SAND (%)	43	9	15	25	32	50	95	19	0

Observing the quartile limits for the BIOMASS variable one may note that there appears to be an excess of samples with low weights (Figure 2.11). Twenty-five percent of the samples weigh less than 2.4 g. There is only 2.4 g difference with the minimum biomass sampled. At the other extreme, the third quartile limit, 25% of the sampled population has a biomass greater than 13.0 g. The remaining 50% of the population has a biomass range between 2.4 g and 13.0 g.

Figure 2.11 Earthworm biomass (g) histogram

2.4.2.1 Coded variables

Codes were created to translate observations in order to include them in the numerical analysis. Tillage is one example for the worm biomass modeling effort. Conventionally tilled fields were coded 1 and field sites that underwent long-term-no-till were coded 0. Standing crops were also translated into codes as standing corn, soybeans and bare fields but without any physically meaningful order. The last coded variable concerning the modeling of worm biomass is the residue coverage. Residue cover was visually assessed in terms of covered percentages. As shown on Table 2.3, specific threshold values were identified as codes. See Figure 2.12 for an overview of residue code frequencies.

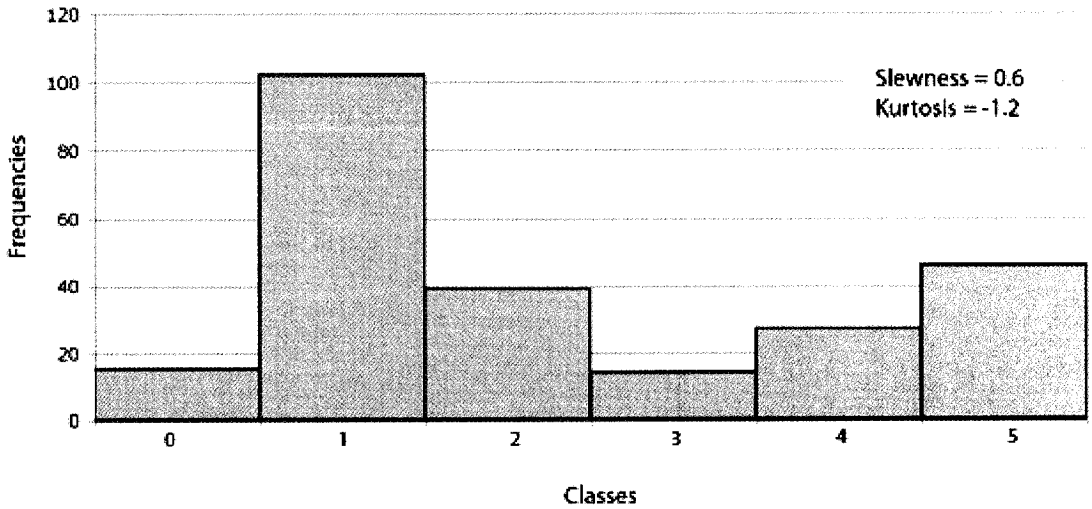


Figure 2.12 Residue code frequencies for earthworm biomass sampling replicates

Spearman's Rank correlations were performed in order to analyze coded variables. Spearman's rank is a non-parametric statistic. It assesses how well a monotonic (can be said of functions between ordered sets if they preserve the given order) function describes relationship between two variables; it can be used for variables measured at the ordinal level. The data are converted into rankings before calculating the coefficients. This statistic allows for the use of codes within the analyses.

2.5 REFERENCES

Ball Coehlo, B.R., Roy, R.C., Topp, E., and Lapen, D.R., 2007. Tile water quality following liquid swine manure application into standing corn. Journal of Environmental Quality 36: 580-587

Culley, J.L.B. 1993. Density and Compressibility in: Carter, M.R. (Ed.) Soil sampling and methods, Lewis Publishers, for the Canadian Society of Soil Science, 823pp. Density and Compressibility pp: 529-539

Fallow, D.J., and Elrick, D.E. 1996. Field measurement of air-entry and water-entry soil water pressure heads. Soil Science Society of America Journal 60: 1036-1039

Heiri, O., Lotter, A.F., and Lemcke, G. 2001. Loss on ignition as a method for organic and carbonate content in sediments: reproductibility and comparability of results. Journal of Paleolimnology 25: 101-110

National Science Programs, Agriculture and Agri-Food Canada
http://sci.agr.ca/london/faq/worm-ver_e.htm

Reynolds, W.D., Elrick, D.E., 1990. Ponded infiltration from a single ring: I. Analysis of steady flow. Soil Science Society of America Journal 54: 1233-1241

Reynolds, W.D. 1993. Saturated hydraulic conductivity: field measurements; chapter 56 in: Carter, M.R. (Ed.) Soil Sampling and Methods of Analysis, Lewis Publishers, Boca Raton, Florida, 823 pages. Saturated hydraulic conductivity – Field measurements pp: 599-613

Sheldrick, B.H., and Wang, C., 1993. Particle size distribution in: Carter, M.R. (Ed.) Soil sampling and methods, Lewis Publishers, for the Canadian Society of Soil Science, 823pp. Particle Size Distribution pp: 499-512

Statistics Canada:
http://www40.statcan.ca/l01/cst01/agrc05a.htm

Topp, G.C. 1993. Soil water content in: Carter, M.R. (Ed.) Soil sampling and methods, Lewis Publishers, for the Canadian Society of Soil Science, 823pp.

Chapter 3

Towards Development of Regional Indicators of Field-Saturated Hydraulic Conductivity and Soil Saturated Air-entry Tensions: Eastern Ontario, Canada

Ouellet, G[1]., Lapen, D.R[2]., Sawada, M[1]., Topp, E[3]., and Edwards, M[2].

1. University of Ottawa, Department of Geography, Ottawa, ON.
2. Agriculture and Agri-Food Canada, Ottawa, ON
3. Agriculture and Agri-Food Canada, London, ON

ABSTRACT

Preferential, or gravity flow, is a mechanism where water moves freely and rapidly through large pores (macropores and/or open burrows). A thorough understanding of regional macropore flow precursors would be beneficial to improve agricultural best-management practices and policy-making towards water and environmental preservation. Eight CART-based regression trees, modeling shallow (5 cm deep) and benched (15 cm deep) field-measured soil hydraulic properties (K_{fs} and AEV) were computed on the basis of *in situ* soil properties (clay, silt, sand, organic matter, bulk density, structure and vegetation), crop type, and tillage therapy. Potential robust indicators of regional macropore flow precursors were identified by analyzing relationships among physical/crop/management properties and soil hydraulic properties. Sandy (sand >45%) soils exhibit rapid infiltration, due to the porosity of the medium. Clays also exhibit high K_{fs} and low air-entry tensions, mostly when macropores are present in the matrix; the

structural organization of soils is as important as texture for estimating soils' infiltrability. Tilling activities modify soil's structure by loosening it, increasing the porosity and disrupting large channels and burrows. Regional macropore flow precursors are context dependent and dynamic. The conclusions of this chapter could establish the basis of further work on the topic. Soil structure (abundance of macropores, for example) was identified as an interesting indicator of preferential flow mechanisms and could potentially become instrumental in refining regional macroporosity indicators.

3.1 INTRODUCTION

The well-known tragedy of Walkerton resulted in growing public concern about soil/water contamination from agricultural activities (AAFC, 2003). Understanding contaminant transport in the vadose or unsaturated zone is required to develop and implement many agriculturally related best-management practices (AAFC, 2003). Often, regional and/or landscape indicators of contaminant risk are required to inform policy-making processes. Agriculture and Agri-Food Canada is developing a suite of agri-environmental indicators at these scales to assess how well agricultural systems manage and conserve natural resources and how compatible they are with natural systems and processes in the broader environment (AAFC, 2005).

Classification and Regression Tree (CART) analysis, a simple, yet powerful data mining technique, was performed on field-collected hydraulic conductivities and soil physical/structural data in order to evaluate the potential of soil physical properties to model field saturated hydraulic conductivity (K_{fs}) at field and/or regional scale. In order to do so, significant relationships uncovered by the predictive modeling scenario were translated into potential regional indicators of macropore flow incidence.

Hydraulic conductivity is a measure of the "ease" or "ability" of a saturated porous medium to transmit water (Reynolds, 1993). Consequently, field measurements of hydraulic conductivity are critically important in the design and monitoring of irrigation and drainage systems, manure impoundments, septic tanks, canals and reservoirs, sanitary landfills, and many other agricultural, industrial and environmental installations (Bouwer, 1978; McKeague et al., 1982; Reynolds, 1993; Reynolds and Zebchuk, 1996; Reynolds et al., 2000). Field measurements are useful for predicting rainfall infiltration and runoff, and for characterizing changes in soil macrostructure because of changing land management practices (Puckett et al., 1985; Gregorich et al., 1993; Reynolds, 1993; Reynolds and Zebchuk, 1996; Reynolds et al., 2000; Heppell et al., 2002; Bagarello and Sgroi, 2004; Papchepsky et al., 2006).

Before the eighties, water movement in soils was originally calculated by models driven and constrained by soil moisture tension, only (Edwards et al., 1993). Beven and Germann (1982) demonstrated, however, that both local and field scale heterogeneities must be considered when modeling hydrologic processes. Continuous macropores in unsaturated soil allow water to move freely and rapidly, causing short-circuiting, a phenomenon also called macropore, or preferential flow (Bouma et al., 1981).

Macropore flow is a well-established phenomenon in a wide range of soil types (Beven and Germann, 1982; Wilson and Luxmoore, 1988, Heppell et al., 2002) that adds complexity to our understanding of soil-water processes. There is increasing evidence that the transport of agrochemicals, such as fertilizers and herbicides, through such pathways is an important mechanism by which the diffuse pollution of surface and groundwater can occur (Bouma et al., 1981; Kladivko et al., 1991; Edwards et al., 1993; Heppell et al., 2002). In order to identify the threat to surface and groundwater posed by the use of agro-chemicals, a thorough understanding of the macropore flow under a variety of agricultural management regimes and soil types is needed (Heppell et al., 2002). As such, hydropedology becomes a powerful tool for environmental policy research (Bouma, 2006; Papchepsky et al., 2006). Determining temporal changes in field saturated hydraulic conductivity (K_{fs}) is important for understanding and modeling hydrological phenomena at the field scale (Bagarello and Sgroi, 2004).

Measurements of soil hydraulic properties are relatively time-consuming, costly, and become impractical when hydrologic estimates are needed for large areas (van Genuchten, 1980; Topp et al., 1980; Arya et al., 1999; Lin et al., 1999; Papchepsky et al., 2006). Direct measurements of hydraulic conductivity yield results that are variable, error-prone, and applicable to only a narrow range of saturation (Arya et al., 1999). Therefore, considerable efforts have been devoted to the indirect estimation of soil hydraulic properties (van Genuchten and Leji, 1992; Mualem, 1992; Arya et al., 1999).

Pedotransfer functions empirically predict hydraulic properties by relating simple soil characteristics like particle size and bulk density, found in soil surveys, to more complex parameters that are used in modeling and that are relatively difficult to measure (Bouma and van Lanen, 1987; Bouma, 1989; van Genuchten and Leji, 1992; Arya et al., 1999; Bouma, 2006). The utility of pedotransfer functions was immediately recognized because of the multiple uses of hydraulic properties (Papchepsky et al., 2006). Basic soil properties such as particle size distribution and bulk density are widely available for many soil types and can be accurately and routinely determined in laboratories around the world (Arya et al., 1999; Papchepsky et al., 2006). Hence, formulation of hydraulic properties entirely in terms of basic soil properties, using pedotransfer functions, should be of considerable benefit.

The intrinsic variability of hydraulic conductivity at field scale makes accurate estimates difficult to obtain (van Genuchten, 1980; Topp et al., 1980; Bagarello and Sgroi, 2004). Hydraulic conductivity is extremely sensitive to even relatively small differences in sample size, flow geometry and soil structure (Bouma et al., 1981; Bouma, 1983; Topp et al., 1980; Reynolds et al., 2000). These variations may be attributed to differences in particle size distribution, bulk density, mineralogy, microaggregation, and organic matter content within a textural class (Arya et al., 1999).

The primary objective of this study is to explore the use of CART to help uncover interactions among field-sampled soil physical characteristics, crop, and management parameters and soil hydraulic properties. Significant relationships could be translated into suitable indicators of macropore flow risk incidence. This chapter focuses on

relationships identified by the final decision tree models generated with CART. It is not intended to propose a final equation for soil hydraulic properties to be estimated from field-sampled data but rather to explore outstanding relationships that could lead to tangible indicators to address the impacts of agricultural activities on surface and groundwater resources.

3.2 METHODS AND MATERIALS

3.2.1 Regional settings

This study took place on agricultural fields located throughout eastern Ontario, Canada (Minimum latitude: 43.14°N, maximum latitude: 45.42°N; minimum longitude: 74.85°W and maximum longitude: 80.80°W) representing diverse soil textural classes, soil management, (eg., no-tillage, conventional tillage), and cropping systems (eg., bare fiels, corn, and soybean).

3.2.2 Field and laboratory measurements

Soil hydraulic properties, namely K_{fs} and AEV were measured between April and November 2004 and 2005. Twelve replicate infiltration rates and 12 replicate saturated air-entry tensions were measured at 2 depths (5 cm and 15 cm from the surface) for a total of 24 readings per field plot. *In situ* K_{fs} was measured using a single ring (10.5 cm diameter and 6 cm high) constant head pressure infiltrometer (Reynolds and Elrick, 1990; Reynolds, 1993). Multiple head measurements were performed using 10 cm and 40 cm heads. Multiple head analysis provides information on the degree of capillary flow vs. gravity flow with the capillary lenght parameter: 'α' (Reynolds, 1993). The measurements were spaced 1.5 m apart within a 15 m by 10 m area. AEV was measured directly after completing K_{fs} readings via a tensimeter (Soil Measurement Systems, Tucson, Arizona, USA) attached to the pressure infiltrometer according to Fallow and Elirck's (1996) and Ball Coelho et al.'s (2007) prescriptions. Saturated soil air-entry tensions can be used as an indicator of continuous pore networks and the minimum pore radius allowing air to enter the previously saturated medium (Fallow and Elrick, 1996). Infiltrometer measurements were made well into the crop growing season to allow for soil structural maturity, especially for the tilled sites (Lapen et al., 2001).

Bulk density, organic matter content and particle size distribution were also measured with soil hydraulic properties. A piston core (5.5 cm diameter by 10 cm length) was used to measure soil bulk density (Culley, 1993). Samples were taken at 2 depths, 5 and 15 cm within 1 m of the K_{fs} measurement sites. Texture and organic matter content were measured on soil grab samples (15 cm diameter by 15 cm deep) sampled in the center of the K_{fs} measurement area. For soil texture, organic matter was previously removed from the samples (Sheldrick and Wang, 1993). Particle fractions greater than 63 µm were determined using wet sieving and the particle fractions smaller than 63 µm were determined with a Lecotrac LT-100 particle size analyzer (PSA) (Sedimentology Laboratory, Earth Sciences Sector, Ottawa). The soil's organic matter content was determined by submitting soil sub-samples to loss-on-ignition (LOI) analysis (Heiri et al., 2001).

At every PI site, the soil structure was observed and described in terms of pore and crack numbers and sizes; a numerical 'structure' code was assigned to every site to be included in a subsequent analysis as an independent variable. Decaying organic matter and vegetal residue was also described and coded to be included in modeling activities. All variables included in this analysis are described in Table 3.1.

Table 3.1 Variables description and units

Variables description and units (codes are categorical values used in statistical analysis)		
Variable	**Units**	**Description**
BD	g cm^{-3}	Soil bulk density
ORG	%	Soil organic matter content
CLAY	%	Soil clay content
SILT	%	Soil silt content
SAND	%	Soil sand content
CROP	#	Type of crop when sampling. Bare field or no new crops (coded 2), corn (coded 1) and soy beans (coded 0)
TILL	#	Conventional tillage (coded 1), no-tillage (coded 0)
STRUCTURE	#	Visual description of soil structure. Homogeneous, cohesive soil (coded 0), blocky, chunky soil (coded 1), weak apparent porosity (coded 2), medium apparent porosity (coded 3), high apparent porosity (coded 4) and unconsolidated structure or structural collapse (coded 5)
VEG	#	Visual description of the vegetal coverage found at measuring sites. Bare soil, no residue (coded 0), roots (coded 1), a little amount of residue such as soy bean leaves and weeds (coded 2) and a thick layer of vegetal residue such as decaying corn (coded 3)
AEV	cmwater	Soil water air-entry values
KFS	cm s^{-1}	Field saturated hydraulic conductivity

General statistics on non-categorical variables are listed in Table 3.2. General statistics were computed for all the dataset, but the data were also stratified according to depth (surface and benched).

Table 3.2 Summary statistics of non-categorical variables

Summary statistics of non-categorical variables involved in prediction modelling of shallow hydraulic properties (K_{fs} and AEV)									
Variable	**N**	**MIN**	**1st quartile**	**Median**	**Mean**	**3rd quartile**	**MAX**	**STDEV**	**NA**
KFS ($cm\ s^{-1}$)									
Shallow	357	$1.07(10)^{-5}$	$1.95(10)^{-3}$	$7.72(10)^{-3}$	$2.15(10)^{-2}$	$2.32(10)^{-2}$	$2.64(10)^{-1}$	$3.47(10)^{-2}$	4
Benched	420	$2.09(10)^{-6}$	$1.39(10)^{-3}$	$5.10(10)^{-3}$	$1.79(10)^{-2}$	$1.53(10)^{-2}$	$5.27(10)^{-1}$	$4.20(10)^{-2}$	8
AEV (cmwater)									
Shallow	347	-58.12	-18.36	-12.24	-14.05	-8.16	0.0	8.42	14
Benched	406	-44.87	-20.39	-14.28	-15.26	-9.18	0.0	7.40	22
BD ($g\ cm^{-3}$)									
Shallow	353	0.96	1.25	1.38	1.38	1.51	1.80	0.17	8
Benched	420	1.02	1.34	1.46	1.46	1.58	1.86	0.17	8
ORG (%)	78	1.6	4.2	4.8	5.1	6.2	9.0	1.4	0
CLAY (%)	78	1	4	7	11	17	36	9	0
SILT (%)	78	4	43	58	54	67	87	18	0
SAND (%)	78	9	16	27	32	46	95	20	0

3.2.3 Statistical analysis

Summary statistics and non-parametric Spearman's Rank correlation (Table 3.3) analyses were conducted for all variables given in Table 3.1. Recursive, test-learn binary and non-parametric CART-based regression tree modeling (Breiman et al., 1984; Steinberg and Colla, 1997) was undertaken to uncover relationships and interactions between soil hydraulic properties (K_{fs} and AEV) and all of the other variables listed in Table 3.1. CART-based regression tree models are constructed by dividing dependent data groupings (target variables) recursively into sub-groups on the basis of independent variables (primary splitters criteria) using variance-minimizing algorithms expressed as improvement scores in CART (Hastie et al., 2001). The target variables (K_{fs}, for example) are partitioned into a series of descending 'child nodes'. A 'parent node' is partitioned into right and left child nodes according to a splitting criterion. The predicted value of the dependent variable for a specific node are presented here as the standard deviation (STD), the average value (AVG) and the number of node members (N). Tree size is limited by statistics maximizing predictive capacities (R^2) and minimizing model complexity and size (Hastie et al., 2001). CART performs cost-complexity pruning to determine the optimal number of terminal nodes (Steinberg and Colla, 1997; Hastie et al., 2001) to avoid model overfitting. An ancillary component to CART is characterizing the variance minimizing capacity of variables that mimic the primary splitting variable for a nodal split on a case by case basis. The variable that is most strongly associated (Breiman et al., 1984) with the primary splitting variable is deemed the strongest

surrogate splitting criteria; and surrogate information is useful for identifying potential substitutes for primary splitting variables as well as identifying potential multicollinearity. The competitor variable also provides some indication about multicollinearity since it is the second strongest variable in terms of variance minimizing for nodal splitting (Breiman et al., 1984). CART provides an indication of variable importance for a specific tree model. A variable's relative importance is calculated here as the sum of the primary split and surrogate improvement scores throughout the selected tree. The values are scaled relative to the best performing variable (score of 100) (Steinberg and Colla, 1997). For this study, CART software default parameters were used exclusively, as per Lapen et al. (1996, 2001). Eight CART-based regression tree models were computed for this study:

1. predict shallow AEV from soil physical (CLAY, SILT, SAND, ORG, BD, STRUCTURE and VEG), management (TILL) and crop (CROP) properties
2. predict shallow AEV from all variables listed in Table 3.1, including KFS as an independent predictor of AEV
3. predict benched AEV from soil physical (CLAY, SILT, SAND, ORG, BD, STRUCTURE and VEG), management (TILL) and crop (CROP) properties
4. predict benched AEV from all variables listed in Table 3.1, including KFS as an independent predictor of AEV
5. predict shallow KFS from soil physical (CLAY, SILT, SAND, ORG, BD, STRUCTURE and VEG), management (TILL) and crop (CROP) properties
6. predict shallow KFS from all variables listed in Table 3.1, including AEV as an independent predictor of KFS
7. predict benched KFS from soil physical (CLAY, SILT, SAND, ORG, BD, STRUCTURE and VEG), management (TILL) and crop (CROP) properties
8. predict benched KFS from all variables listed in Table 3.1, including AEV as an independent predictor of KFS.

3.3 RESULTS

The 78 field sites evaluated in this study represent a diverse suite of soil textures and soil tillage therapies (19 no-tills and 59 conventionally tilled). Surface crops consisted of bare fields (14 fields), standing corn (60 fields), and standing soybean (4 fields). Spearman's rank correlation coefficients (Table 3.3) were calculated for the whole dataset, first, to provide a global overlook of trends. Coefficients were also calculated for shallow measurements (5 cm), and benched ones (15 cm), in order to explore if there is some specificity, regarding depths, as to variables' interactions. Spearman's rank was chosen to calculate correlation coefficients since it allows such calculations for ordinal variables, likely DEPTH, TILL, CROP, STRUCTURE and VEG for this study. These variables are described here with ordinal values, assigned according to the 'key' found in Table 3.1.

Table 3.3 Spearman's Rank Correlation matrix for All data, shallow, and benched measurements

	DEPTH	KFS	AEV	BD	ORG	CLAY	SILT	SAND	TILL	CROP	STRUC-TURE	VEG
DEPTH												
All data	1	-0.13	-0.11	0.17	-0.01	-0.02	-0.02	-0.01	0.02	-0.04	-0.12	-0.12
Shallow	---	---	---	---	---	---	---	---	---	---	---	---
Benched	---	---	---	---	---	---	---	---	---	---	---	---
KFS												
All data	N=777	1	0.52*	-0.02	0.19*	0.42*	-0.15*	-0.02	0.28*	-0.14*	0.36*	0
Shallow	N=357	1	0.54*	-0.07	0.17*	0.38*	-0.17*	-0.06	0.31*	-0.15*	0.40*	-0.04
Benched	N=420	1	0.47*	0.09*	0.22*	0.45*	-0.14*	0.02	0.28*	-0.16*	0.32*	0
AEV												
All data	N=755	N=755	1	-0.09*	0.06	0.37*	0.12*	-0.19*	0.24*	0.02	0.26*	0.15*
Shallow	N=351	N=351	1	-0.09	0.09	0.37*	0.19*	-0.32*	0.28*	0.03	0.35*	0.07
Benched	N=404	N=404	1	-0.05	0.03	0.36*	0.05	-0.07	0.22*	0	0.15*	0.18*
BD												
All data	N=773	N=773	N=755	1	-0.22*	0.05	-0.42*	0.44*	0.23*	-0.37*	0.08*	-0.13*
Shallow	N=353	N=353	N=351	1	-0.16*	0.07	-0.42*	0.45*	0.27*	-0.33*	0.06	0
Benched	N=420	N=420	N=404	1	-0.28*	0.04	-0.46*	0.48*	0.20*	-0.38*	0.15*	-0.19*
ORG												
All data	N=789	N=777	N=755	N=773	1	0.52*	0.25*	-0.39*	-0.18*	0.22*	-0.05	-0.04
Shallow	N=361	N=357	N=351	N=353	1	0.50*	0.20*	-0.36*	-0.17*	0.16*	-0.03	0.11*
Benched	N=428	N=420	N=404	N=420	1	0.54*	0.30*	-0.41*	-0.19*	0.27*	0.07	-0.15*
CLAY												
All data	N=789	N=777	N=755	N=773	N=789	1	-0.05	-0.24*	0.22*	-0.20*	0.12*	0.04
Shallow	N=361	N=357	N=351	N=353	N=361	1	-0.13*	-0.24*	0.20*	-0.22*	0.17*	0.08
Benched	N=428	N=420	N=404	N=420	N=428	1	0.02	-0.25*	0.23*	-0.17*	0.06	0.01
SILT												
All data	N=789	N=777	N=755	N=773	N=789	N=789	1	-0.79*	-0.23*	0.41*	-0.09*	0.05
Shallow	N=361	N=357	N=351	N=353	N=361	N=361	1	-0.86*	-0.27*	0.46*	-0.12*	0.15*
Benched	N=428	N=420	N=404	N=420	N=428	N=428	1	-0.73*	-0.19*	0.38*	-0.06	-0.02
SAND												
All data	N=789	N=777	N=755	N=773	N=789	N=789	N=789	1	0.23*	-0.27*	0.10*	-0.04
Shallow	N=361	N=357	N=351	N=353	N=361	N=361	N=361	1	0.21*	-0.28*	0.07	0.12*
Benched	N=428	N=420	N=404	N=420	N=428	N=428	N=428	1	0.25*	-0.28*	0.12*	0.02
TILL												
All data	N=789	N=777	N=755	N=773	N=789	N=789	N=789	N=789	1	-0.23*	0.07	0.10*
Shallow	N=361	N=357	N=351	N=353	N=361	N=361	N=361	N=361	1	-0.20*	0.13*	0.12*
Benched	N=428	N=420	N=404	N=420	N=428	N=428	N=428	N=428	1	-0.25*	0.02	0.10*
CROP												
All data	N=789	N=777	N=755	N=773	N=789	N=789	N=789	N=789	N=789	1	-0.03	0.07*
Shallow	N=361	N=357	N=351	N=353	N=361	N=361	N=361	N=361	N=361	1	0.02	0.01
Benched	N=428	N=420	N=404	N=420	N=428	N=428	N=428	N=428	N=428	1	-0.11*	0.08
STRUCTURE												
All data	N=789	N=777	N=755	N=773	N=789	N=789	N=789	N=789	N=789	N=789	1	0.02
Shallow	N=361	N=357	N=351	N=353	N=361	N=361	N=361	N=361	N=361	N=361	1	0
Benched	N=428	N=420	N=404	N=420	N=428	N=428	N=428	N=428	N=428	N=428	1	0
VEG												
All data	N=789	N=777	N=755	N=773	N=789	N=789	N=789	N=789	N=789	N=789	N=789	1
Shallow	N=361	N=357	N=351	N=353	N=361	N=361	N=361	N=361	N=361	N=361	N=361	1
Benched	N=428	N=420	N=404	N=420	N=428	N=428	N=428	N=428	N=428	N=428	N=428	1

* Identifies statistically significant correlations (P-values < 0.05)

The stratification of the data according to sampling depth (5 cm or 15 cm) yielded interesting distinctions between correlation coefficients. On one hand, tillage and AEV generally have higher correlation coefficients for shallow measurements. On the other hand, physical characteristics such as CLAY, SILT and SAND, BD and ORG have stronger relationship coefficients for benched measurements.

The strongest correlation coefficient associated with KFS is for shallow measurement values and it is with AEV (0.54). The second strongest association is found between KFS and CLAY, for the benched scenario (0.45). The third strongest correlation coefficient implicating KFS is with TILL for shallow measurement values (0.31). AEV follows the same correlation pattern as the aforementioned for KFS; the second strongest correlation is found between AEV and CLAY (for benched measurement values) with an $R^2 = 0.37$ and the third strongest one being between AEV and TILL (for shallow measurement values) with an $R^2 = 0.28$. Other interesting correlations are found for benched measurement values between ORG and CLAY (0.54), and between BD and SAND (0.48).

3.3.1 Regression tree models organization

Eight CART-based regression trees were computed in an attempt to uncover potentially robust indicators of soil hydraulic properties namely field saturated hydraulic conductivity (KFS) and saturated soil air-entry tension (AEV). The entire dataset has been stratified into shallow measurements (5 cm from the surface), presented first, and benched ones (15 cm deep), presented second. Additionally, distinctions are made between modeling efforts based on soil physical and management properties only, and modeling including soil hydraulic properties as independent (predictor) variables with soil physical and management properties. The reason why half of the modeling scenarios exclude hydraulic properties is that these variables are known to be highly variable in both time and space (Lin et al., 1999 and Papchepsky et al., 2006). The regression trees presented here constitute an effort to provide information on robust relationships and interactions between soil hydraulic behavior and soil variables that are considered more temporally stable for spatial scaling activities (van Genuchten, 1980). Even though soil hydraulic properties, such as K_{fs} and AEV, are known to be transient, including them in half of the modeling scenarios allows for the exploration of relationships and interactions amongst them. Tables 3.4 and 3.5 summarize the most important variables' relative scores.

Table 3.4 Relative importance score for AEV tree modeling*

Relative importance score for AEV tree modeling			
Shallow AEV values predicted with soil physical and management properties		Shallow AEV values predicted with soil physical and management properties and hydraulic properties	
SAND	100	KFS	100
SILT	58.83	SILT	28.48
STRUCTURE	54.21	BD	16.48
CLAY	50.12	SAND	14.77
ORG	32.00	STRUCTURE	12.65
TILL	10.39	VEG	8.06
CROP	7.57	ORG	6.34
BD	4.49	CLAY	5.34
Benched AEV values predicted with soil physical and management properties		**Benched AEV values predicted with soil physical and management properties, and hydraulic properties**	
CLAY	100	KFS	100
SILT	53.48	CLAY	93.58
ORG	44.22	SILT	31.76
SAND	37.49	ORG	21.09
TILL	29.67	STRUCTURE	13.85
BD	27.82	SAND	8.86
CROP	22.22	BD	0.78
STRUCTURE	2.07		

* Two decimal places were kept to respect model integrity

Table 3.5 Relative importance scores for K_{fs} tree modeling*

Relative importance scores for K_{fs} tree modeling			
Shallow K_{fs} values predicted with soil physical and management properties		**Shallow K_{fs} values predicted with soil physical and management properties, and hydraulic properties**	
CLAY	100	AEV	100
SAND	57.27	CLAY	91.14
ORG	41.29	STRUCTURE	38.73
SILT	40.40	SILT	22.88
TILL	24.58	BD	0.51
STRUCTURE	22.37	SAND	0.02
CROP	16.55		
BD	2.16		
Benched K_{fs} values predicted with soil physical and management properties		**Benched K_{fs} values predicted with physical and management properties and hydraulic properties**	
CLAY	100	CLAY	100
ORG	15.02	ORG	15.02
STRUCTURE	2.75	STRUCTURE	2.75

* Two decimal places were kept to respect model integrity

3.3.2 CART-based Regression Trees description

3.3.2.1 Shallow AEV

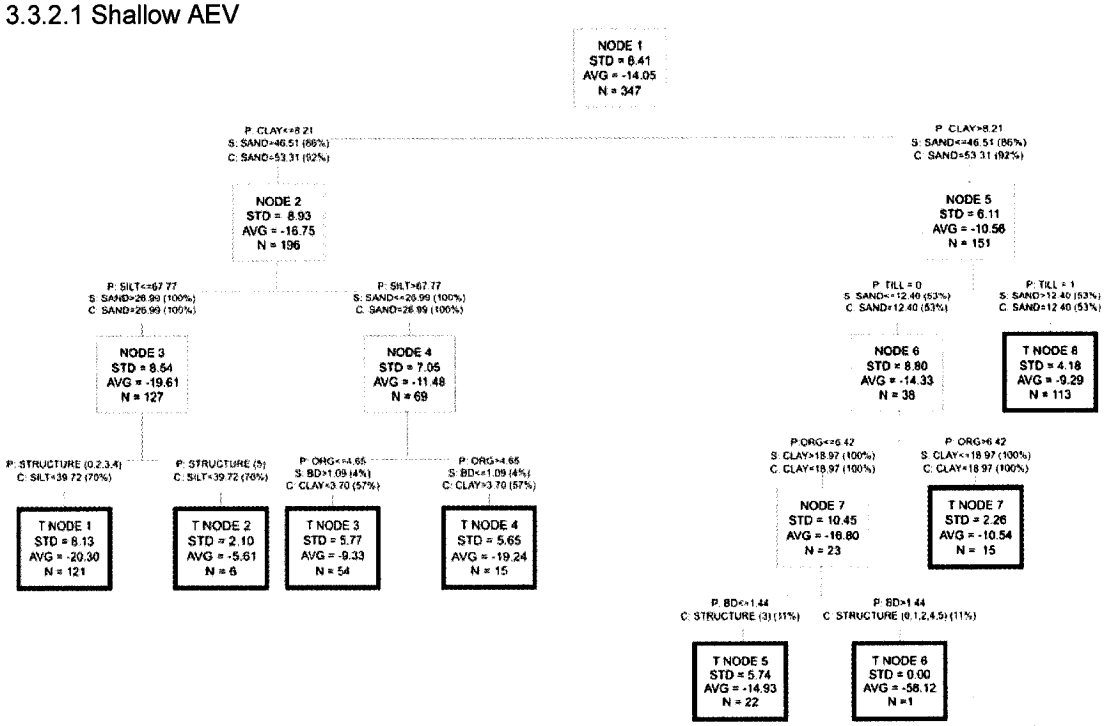


Figure 3.1 Regression tree modeling shallow AEV based on CLAY, SILT, SAND, ORG, BD, CROP, TILL, STRUCTURE and VEG

The automated cross-validated CART-based regression tree for modeling regional shallow AEV according to physical, crop and management properties (Table 3.1) is based on 347 field sites with an average AEV of -14.05±8.41 cm water (NODE 1). The model has 8 terminal nodes (Figure 3.1), an R^2 of 0.48 and a cross-validated relative error of 0.74. The most important variables, in terms of primary+surrogate improvement scores are SAND (100%), SILT (58.83%) and STRUCTURE (54.21%) for this modeling scenario (Table 3.4). Node 1 data stratifies based on CLAY content (8.21%) into low CLAY content soils with higher than the average AEV (NODE 2, average AEV=-16.75 ±8.93 cm water) and higher clay content soils with lower than the average AEV (NODE 5, average AEV=-10.56±6.11 cm water). SAND is identified as the first surrogate variable (46.51%) and the first competitor (53.31%) for that split; both improvement scores come within reach of the primary splitting variable's improvement score (86% and 92%, respectively). This can be interpreted as the possibility of using SAND to model shallow air-entry values at this stage of the modeling process. One hundred and ninety-six field sites with CLAY≤8.21% (NODE 2) further split on the basis of SILT (67.77% splitting criteria) into higher AEV for less silty soils (Node 3, average AEV=-19.61±8.54 cm water) and lower AEV for siltier soils (NODE 4, average AEV=-11.48±7.05 cm water). Surrogate and competitor splitting criteria are the same at this stage of the tree: SAND (26.99%) with an improvement score identical to the one of the primary splitting variable (SILT). The high correlation coefficients observed between SAND and SILT (see Table 3.3) reinforces the hypothesis that SAND could be used as a robust indicator of shallow AEV in the same degree that SILT is. 127 field sites with SILT≤67.77% are

regrouped into NODE 3 and rearranged following the structural description of the measurement sites (STRUCTURE). More consolidated soils (STRUCTURE = 0, 1, 2, 3 and 4) have higher AEV (T NODE 1, average AEV=-20.30±8.13 cm water) and looser soils (STRUCTURE = 5) have lower AEV (T NODE 2; average AEV=-5.61±2.10 cm water). There are no surrogate variables for that split and the competitor variable is SILT (40%) with an improvement score representing 70% of the primary splitter. This modest improvement score underlines the modest potential of SILT (40%) to mimic that split.

The 151 fields with CLAY>8.21% (NODE 5) are split into relatively higher AEV no-till sites (NODE 6, average AEV=-14.33±8.80 cm water) and relatively lower AEV tilled sites (T NODE 8, average AEV=-9.29±4.18 cm water). SAND (12.40%) is identified as the first surrogate and the first competitor variable for that split, representing 53% of the primary splitter's improvement score. The 8 no-till field sites (NODE 6) reorganize according to ORG (6.42%) into sites with organic matter content ≤6.42% and higher but more variable AEV (NODE 7, average AEV=-16.80±10.45 cm water) and sites with higher organic matter contents and lower and more stable AEV (T NODE 7, average AEV=-10.54±2.26 cm water). First surrogate and competitor splitting criteria for that split are embodied in CLAY (18.97%), representing 100% of the ORG (6.42%)'s improvement score. Shallow CLAY and shallow ORG interrelate strongly (see Table 3.3) according to the Spearman's Rank correlation matrix and this demonstrates in the good potential for CLAY to imitate the variance and complexity minimizing split of AEV according to ORG. Less organic soils (NODE 7) are segregated based on BD (1.44 g cm^{-3}) into near average AEV soils (T NODE 5, average AEV =-14.93±5.74 cm water) and extremely high AEV and dense soils (T NODE 6, AEV= -58.12 cm water). The competitor variable for that split is the soil's structure code but since its improvement score is low, it cannot be considered as a robust predictor of shallow AEV.

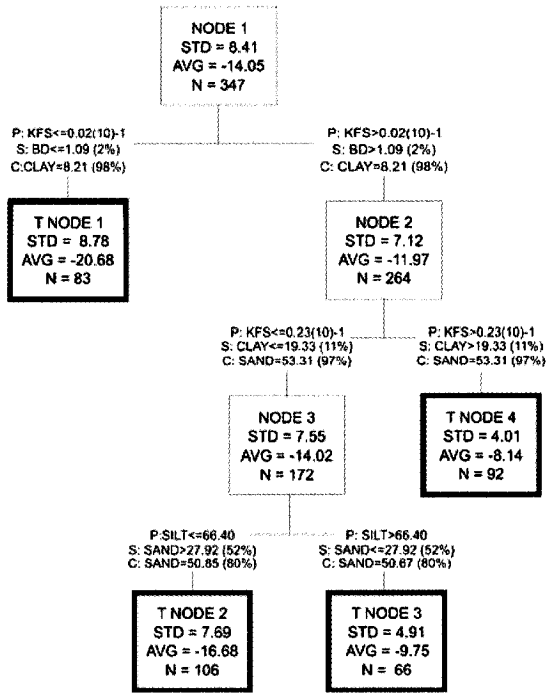


Figure 3.2 Regression tree modeling shallow AEV based on all variables listed in Table 3.1

When KFS was included as an independent variable to predict shallow AEV, a cross-validated CART based regression tree model (Figure 3.2) with 4 terminal nodes was produced. The R^2 of the model is 0.36, and the relative error is 0.79. The starting point of this regression tree is identical to the one on Figure 3.1. Table 3.4 lists KFS (100%), SILT (28.48%) and BD (16.48%) as the most important variables in terms of primary+surrogate improvement score for shallow AEV modeling. The first primary splitter for that tree is KFS (0.02 cm s^{-1}), segregating the data into slower infiltrating soils with higher tensions (T NODE 1 with average AEV = -20.68±8.78 cm water) and faster KFS with lower AEV (NODE 2, average AEV = -11.97±7.12 cm water). The surrogate splitting variable's improvement score is negligible but the competitor splitting variable has an improvement score of 98%, leading to the validity of using CLAY (8.21%) as a robust predictor of shallow AEV. The 264 members of NODE 2 split again on the basis of infiltration rates (0.23 cm s^{-1}) into higher air-entry tensions (NODE 3: average AEV = -14.02±7.55 cm water) and lower ones (T NODE 4: average AEV = -8.14±4.01 cm water). The surrogate splitter has a negligible improvement score and the competitor splitter (SAND = 53.31%) can be thought of as a potential regional indicator of regional shallow AEV, with an improvement score of 97%. SILT contents (66.40%) subdivide NODE 3 into less silty soils with higher AEV (T NODE 2: average AEV = -16.68±7.69 cm water) and siltier soils with lower air-entry tensions (T NODE 3: average AEV = -9.75±4.91 cm water). The surrogate and competitor splitting criteria's improvement scores are based on SAND contents and have a modest to good capacity for predicting shallow regional AEV.

3.3.2.2 Recapitulation for shallow AEV modeling

The general statistics on CART-based regression trees modeling shallow soil saturated air-entry values tend to demonstrate that soil physical/crop/management properties yield more significant robust regional indicators as the R^2 for that tree is higher than the one produced with the dataset including field saturated hydraulic conductivity (0.48>0.36). In addition, the cross-validated relative error of the first model is lower than the one computed for the second one (0.74<0.79).

Table 3.4 shows that there are numerous differences between the most important variables identified for the purpose of predicting regional shallow AEV according to the inclusion or exclusion of KFS as an independent predictor. The second tree (Figure 3.2), KFS is identified as the first most important variable (100%) far ahead of the second most important one: SILT (28.48%). Nevertheless some similar patterns between both scenarios can be acknowledged. In the first place, CLAY (8.21%) is the first splitter for the physical/crop/management regression tree (Figure 3.1) and it also is the first competitor splitter for the other tree, representing 92% of the primary splitter's improvement score. Second, SAND appears in both regression trees modeling regional shallow AEV as first competitor splitter with improvement scores higher or equal to 92%. Such improvement scores identify robust indicators of regional shallow AEV. These specifications allow us to conclude that soil textural properties are a valid alternative to K_{fs} in order to predict shallow AEV in terms of minimizing model variance and complexity.

3.3.2.3 Benched AEV

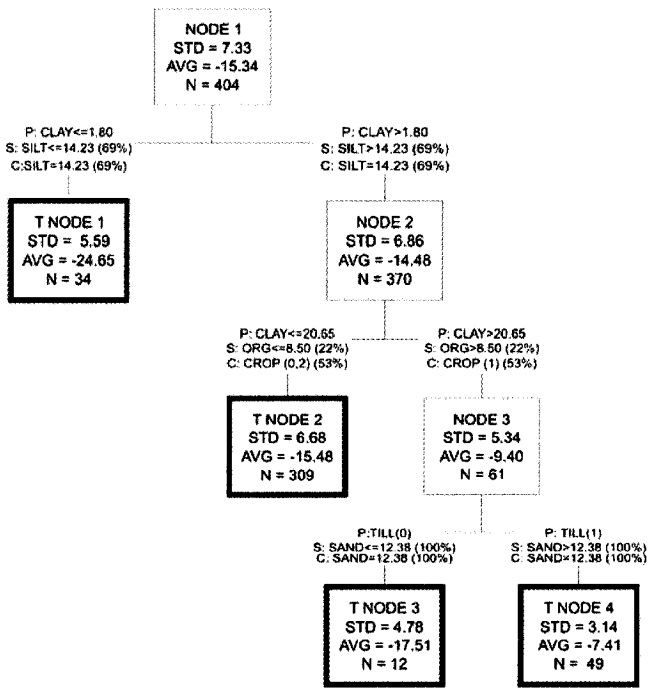


Figure 3.3 Regression tree modeling Benched AEV based on CLAY, SILT, SAND, ORG, BD, CROP, TILL, STRUCTURE and VEG

The cross-validated CART-based regression tree predicting benched AEV on the basis of CLAY, SILT, SAND, ORG, BD, VEG and STRUCTURE (Table 3.1) has an R^2 of 0.28 and a relative error of 0.75. Table 3.4 lists the most important variables for minimizing variance and complexity (based on simple least squares algorithms) of this model, namely CLAY (100%), SILT (53.48%) and ORG (44.22%) as the strongest ones. The 404 sites comprised by NODE 1 form the entire dataset for this model, with an AEV average of -15.34±7.33 cm water (See Figure 3.3). CLAY (1.80%) splits the dataset into higher air-entry tensions (T NODE 1, average AEV: -24.65±5.59 cm water) for soils with CLAY contents ≤ 1.80% and lower air-entry tensions for higher clay contents soils (NODE 2, average AEV: -14.68±6.86 cm water). The first surrogate splitting variable is the same as the first competitor: SILT (14.23%) and both variables have an improvement score of 69%. This importance score suggests that SILT could potentially, with some reserves, be used for modeling benched AEV. The 370 field sites with CLAY>1.80% (NODE 2) further split according to CLAY (20.65%), forming one subgroup with less clayey soils and tensions somewhat higher (T NODE 2, average AEV: -15.48±6.68 cm water) and another one, with more clayey soils and lower tensions (NODE 3; average AEV= -9.40±5.34 cm water). Surrogate and competitor splitting criteria for this split have low importance scores and therefore CLAY is maintained as a robust regional indicator of benched AEV. The 61 fields with CLAY>20.65% (NODE 3) are classified as higher air-entry tensions for long-term-no-till fields (T NODE 3; average AEV=-17.51±4.78 cm water) and lower ones for tilled soils (T NODE 4; average AEV=-7.41±3.14 cm water). The strongest surrogate variable, SAND (12.38%) is the same as the first competitor for that split, and both have the same improvement score as the primary splitter (TILL), representing their potential to mimic that split and their validity as a robust indicator of such benched air-entry values.

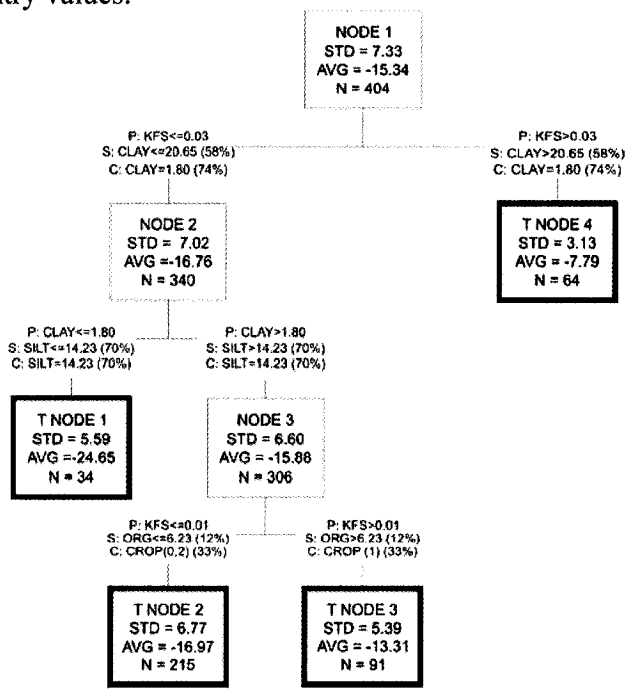


Figure 3.4 Regression tree modeling benched AEV based on all variables listed in Table 3.1

Including KFS into modeling activities as a potential predictor of benched AEV yielded a regression tree with 4 terminal nodes (Figure 3.5). The R^2 of the regression tree model is 0.35 and the model produced the lowest cross-validated relative error of 0.70. The most important variables in terms of primary+surrogate split improvement scores are KFS (100%), CLAY (93.58%) and SILT (31.76%) (Table 3.4). Node 1 includes 404 field sites with an AEV average of -15.34±7.33 cm water. The dataset first splits based on KFS, into soils with faster infiltrations and lower tensions (T NODE 4 with an AEV average of -7.79±3.13 cm water) and sites with slower infiltration rates and higher AEV (NODE 2 with an AEV average of -16.76±7.02 cm water). The surrogate (CLAY $\leq$ 20.65%) and competitor's (CLAY = 1.80%) improvement scores are relatively high, suggesting CLAY contents could be a robust indicator of AEV at this stage of the tree modeling. The 340 sites with KFS$\leq$0.03 cm s^{-1} (NODE 2) reorganize on the basis of CLAY contents (1.80%) into higher tension soils with low clay contents (T NODE 1 with an AEV average of -24.65±5.59 cm water) and lower tension soils with higher clay contents (NODE 3 with an AEV average of -15.88±6.60 cm water). The surrogate and competitor splitting criteria is the same, SILT=14.23%, with a great improvement score (representing 100% of the primary splitter's improvement score); this could be interpreted as the potential to use SILT=14.23% as a robust regional predictor for benched AEV. The 306 field sites with CLAY>1.80% (NODE 3) further reorganize on the basis of KFS (Threshold value = 0.01 cm s^{-1}) into terminal nodes 2 and 3. Greater infiltration rates are associated with relatively lower air-entry tensions (T NODE 3 with an average AEV of -9.75±4.91 cm water). The remaining 215 sites from T NODE 2, with KFS $\leq$0.01 cm s^{-1} have an average AEV of -16.97±6.77 cm water. The surrogate and competitor splitting criteria for this step of the regression tree are too low to include these variables as potential indicators of regional air-entry tensions.

3.3.2.4 Recapitulation for benched AEV modeling

The general statistics of the regression tree models indicate that including KFS as a predictor of benched AEV (Figure 3.4) appears useful since the R^2 corresponding to that model is higher than the one describing the regression tree without soil hydraulic properties (Figure 3.3) (0.35>0.28). The cross-validated variance and complexity minimizing relative error related to the inclusion of KFS into the model is lower than the one associated with the other regression tree modeling scenario (0.70<0.75). Even though the statistical model including KFS as a predictor leads to greater predictive potential (greater R^2), it is important to remember that hydraulic properties are highly variable both in time and space, and therefore impose a serious limitation to space scaling endeavours.

In spite of the superior R^2 of the regression tree model including KFS as an independent variable, texture properties such as CLAY and SILT appear as suitable alternative predictor variables of AEV to KFS. As an example, the first splitter of the regression tree model including hydraulic properties (Figure 3.4) is KFS (0.03cm s^{-1}) and interestingly, the first surrogate (CLAY=20.65%) and the first competitor (CLAY=1.80%) represent clay contents with relatively high improvement scores (representing 58% and 74% of KFS's, respectively). The second primary splitter for that tree model is CLAY (1.80%), the first primary splitting variable for the regression tree without KFS (Figure 3.3). SILT

and SAND are identified as potentially strong indicators of benched AEV in both regression tree models with improvement scores identical to their respective primary splitting variables.

3.3.2.5 Shallow KFS

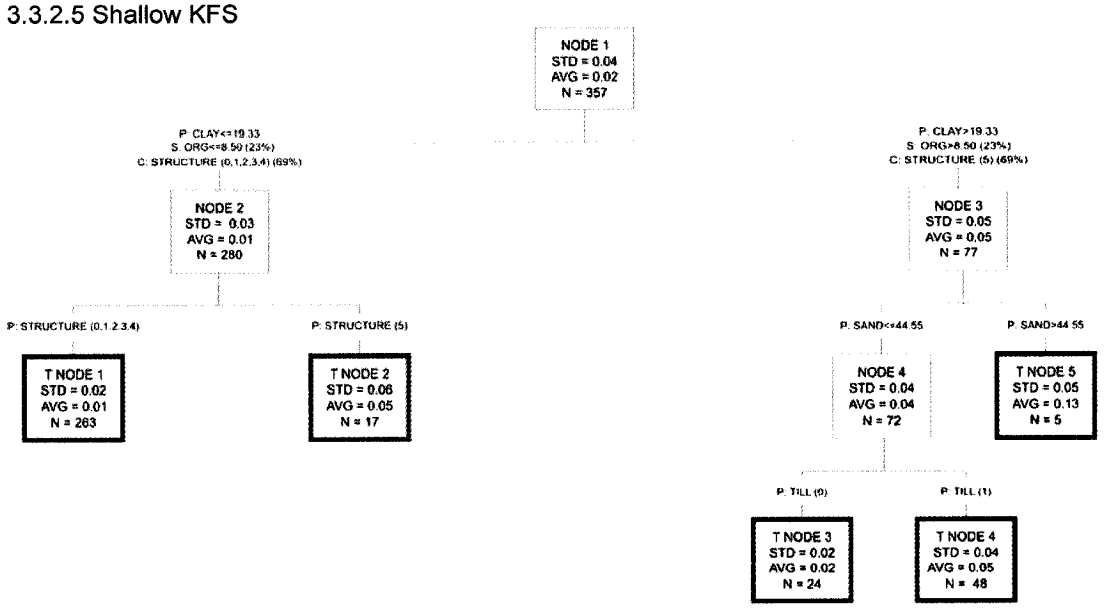


Figure 3.5 Regression tree modeling shallow KFS based on CLAY, SILT, SAND, ORG, BD, TILL, CROP, STRUCTURE and VEG

The cross-validated CART based regression tree for modeling shallow KFS based on soil physical, crop, and management properties (Table 3.1) has 5 terminal nodes, 357 field sites with an average of 0.02±0.04 cm s^{-1} (NODE 1), an R^2 of 0.44 and a relative error of 0.67 (Figure 3.5). The most important variables, in terms of primary+surrogate improvement score are CLAY (100%), SAND (57.27%) and ORG (41.29%) (Table 3.5). CLAY (19.33%) splits the dataset into less clayey soils with lower infiltration rates (NODE 2; average KFS=0.01±0.03 cm s^{-1}) and soils with clay contents > 19.33% and greater KFS (NODE 3; average KFS=0.05±0.05 cm s^{-1}). The first surrogate variable for this split is ORG (8.50%) but its improvement score is too low for ORG to be considered as a valid indicator of shallow KFS at this stage of the modeling process. The first competitor splitting criterion is STURCTURE, representing 69% of CLAY's improvement score. This is a modest representation of the actual split, suggesting that, with refinement, STRUCTURE could become a robust indicator of shallow infiltration rates. NODE 2, (soils with clay contents ≤ 19.33%) subdivides into more consolidated soil structure (STRUCTURE = 0, 1, 2, 3 and 4) and slower KFS (T NODE 1; average KFS=0.01±0.02 cm s^{-1}) and unconsolidated soils (STRUCTURE = 5) with more rapid infiltration rates (T NODE 2; average KFS = 0.05±0.06 cm s^{-1}).

Fields with clay contents > 19.33% (NODE 3) further split according to SAND (44.55%) into less sandy soils with lower KFS (NODE 4; average KFS=0.04±0.04 cm s^{-1}) and sandier soils with greater infiltration rates (T NODE 5; average KFS=0.13± 0.05cm s^{-1}). Less sandy sites (NODE 4) segregate into no-till field sites with lower KFS (T NODE 3; average KFS=0.02±0.02 cm s^{-1}) and tilled soils with faster KFS (T NODE 4; average KFS=0.05±0.04 cm s^{-1}).

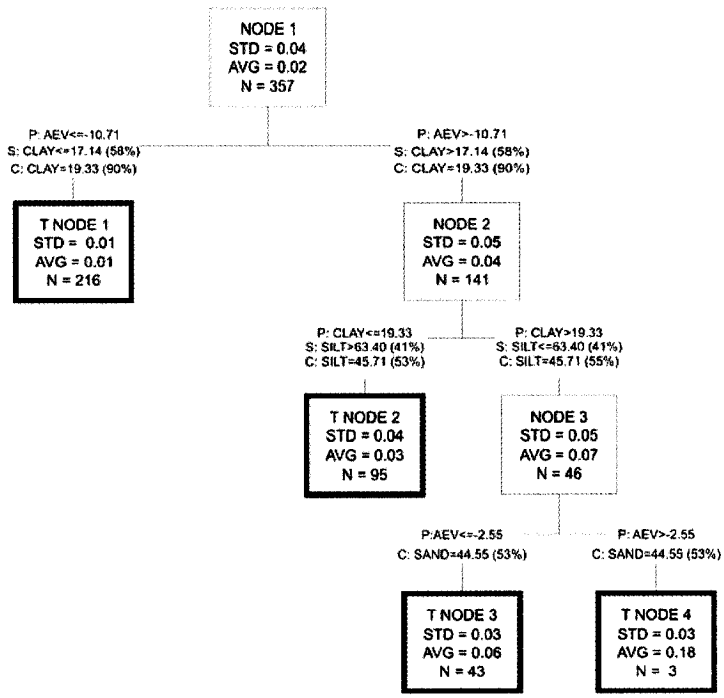


Figure 3.6 Regression tree modeling shallow KFS based on all variables listed in Table 3.1

When including AEV is included in the process of shallow KFS modeling, a cross-validated CART based regression tree with 4 terminal nodes based on 357 field sites with an average KFS of 0.02±0.04 cm s^{-1} was produced (Figure 3.6). This modeling tree is characterized by an R^2 of 0.43 and a relative error of 0.71. The first splitting criterion for this regression tree model is AEV (-10.71 cm water); splitting the data into sites with higher tensions and lower KFS (T NODE 1; average KFS=0.01±0.01 cm s^{-1}) and sites with lower AEV and greater infiltration rates (NODE 2; average KFS=0.04±0.05 cm s^{-1}). CLAY represents both the first surrogate (17.14%) and the first competitor (19.33%) variable for that split. The competitor splitting criteria represents 90% of the primary splitter's improvement score, indicating that CLAY (19.33%) could be a suitable indicator of shallow KFS. The 141 field sites with AEV>-10.71 cm water (NODE 2) rearrange into less clayey soils with lower KFS rates (T NODE 2; average KFS= 0.03±0.04 cm s^{-1}) and higher clay contents soils with faster KFS (NODE 3; average KFS=0.07±0.05 cm s^{-1}). Surrogate and competitor splitting criteria are based on SILT but both offer a weak representation of the primary splitter's improvement score, confirming that CLAY can be considered as a robust predictor of shallow KFS at this stage of the modeling process. The 43 sites with CLAY>19.33% further stratify into relatively higher AEV and relatively slower KFS (T NODE 3; average KFS=0.06±0.03 cm s^{-1}) and low AEV and fast infiltration rates (T NODE 4; average KFS=0.18±0.03 cm s^{-1}). There is no surrogate splitting criterion identified for that split and the competitor splitting criterion is SAND (44.55%) modestly representing the improvement score of the actual split.

3.3.2.6 Recapitulation for shallow KFS modeling

General statistics on the two CART-based regression trees modeling shallow KFS show better predictive potential for the predictive regression tree without AEV (Figure 3.5) since its R^2 is slightly higher (0.44>0.43) and its relative error lower (0.67<0.71).

Observing primary, surrogate and competitor splitting variables for both regression trees, one may note that in the tree including AEV (Figure 3.6), soil texture properties, such as CLAY and SAND, are identified as strong competitor splitting variables. More precisely, CLAY (19.33%) appears as the first competitor splitter to the first split primary criterion (AEV) with an improvement score representing 90% of AEV's. Therefore, CLAY (19.33%) can be thought of as a robust indicator of shallow KFS. Furthermore, CLAY (19.33%) is the first primary splitter for the soil physical/crop/management properties regression tree (Figure 3.5). NODE 3 data from that same tree (Figure 3.5), splits according to SAND (44.55%) without any strong surrogate or competitor variables, and this variable (SAND 44.55%) is identified as a potential alternative indicator of regional shallow KFS to AEV, since its improvement score, as a competitor variable to AEV represents 53% of AEV's improvement score.

3.3.2.7 Benched KFS

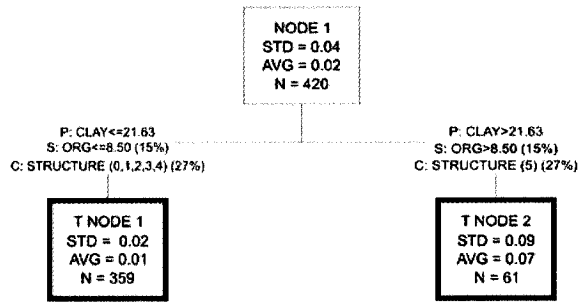


Figure 3.7 Regression tree modeling benched KFS based on CLAY, SILT, SAND, ORG, BD, TILL, CROP, STRUCTURE and VEG

The cross-validated CART-based regression tree modeling KFS at 15 cm depth on the basis of CLAY, SILT, SAND, ORG, BD, TILL, STRUCTURE AND VEGETATION (Table 3.1) comprises 420 sites with an average KFS of 0.02±0.04 cm s^{-1} (NODE 1), 2 terminal nodes (Figure 3.7) and has an R^2 of 0.28 and a relative error of 0.73. The only splitting criterion for this dataset is CLAY (21.63); organizing the data into lower clay contents soils with lower KFS (T NODE 1; average KFS =0.01±0.02 cm s^{-1}) and higher clay contents soils with greater infiltration rates (T NODE 2; average KFS=0.07±0.09 cm s^{-1}). Both the first surrogate and the first competitor splitting criteria are too low to be considered as valid robust regional indicators of benched KFS.

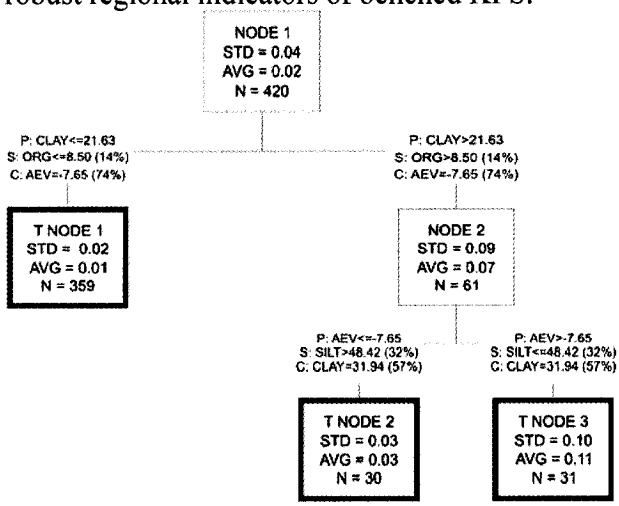


Figure 3.8 Regression tree modeling benched KFS based on all variables listed in Table 3.1

When AEV is used as a predictor, with all variables presented in Table 3.1, to model benched KFS, we obtain a CART-based regression tree including 420 field sites with an average KFS of 0.02±0.04 cm s^{-1} (NODE 1), 3 terminal nodes, and characterized by an R^2 of 0.43 and a relative error of 0.57 (Figure 3.8). The first splitting criterion is based on CLAY (21.63%) and the dataset reorganizes into low clay contents sites with lower infiltration rates (T NODE 1; average KFS = 0.01±0.02 cm s^{-1}) and higher clay contents soils with greater KFS (NODE 2; average KFS = 0.07±0.09 cm s^{-1}). The main surrogate variable has a low to negligible improvement score and the first competitor is AEV (-7.65

cm water), representing 74% of the primary splitter's improvement score. This modest improvement score suggests that AEV (-7.65 cm water) could be used as an indicator of benched KFS but the difficulties associated with AEV measurements and its spatial and temporal variability may relativize its effectiveness as a robust indicator of regional benched infiltration rates. Clayey soils (NODE 2) further organize into sites with higher air-entry tensions and slower KFS (T NODE 2; average KFS = 0.03±0.03 cm s^{-1}) and soils with lower air-entry values and faster infiltrations (T NODE 3; average KFS=0.11±0.10 cm s^{-1}). Both surrogate and competitor splitting criteria represent a weak proportion of the AEV's improvement score, stressing the validity of benched AEV as a robust indicator of benched KFS.

3.3.2.8 Recapitulation for benched KFS modeling

General statistics describing the two CART-based regression trees predicting benched KFS tend to identify the latter one (Figure 3.8) to have greater predictive potential since its R^2 is higher than the one for the regression tree based on soil physical/crop/management properties only (Figure 3.7) (0.43>0.28). Comparing the cross-validated model variance and complexity minimizing relative error of each regression tree model the one including all variables listed in Table 3.1 appears to be more favourable since its relative error is lower (0.57<0.73). The most important variables, in terms of primary+surrogate improvement scores, for both modeling scenarios identified the same variables with the same importance (CLAY 100%, ORG 14.02% and STRUCTURE 2.75%) (Table 3.5).

Both regression tree models have the same first primary splitter (CLAY 19.33%). In the case of benched KFS modeling based on variables excluding AEV (Figure 3.7), there are no surrogates or competitor splitters with strong enough improvement scores to be identified as robust indicators of benched KFS. For the other regression tree modeling benched KFS (Figure 3.8), the primary splitter (CLAY 19.33%) has AEV as a modest competitor splitter representing 74% of CLAY's improvement score.

3.3.3 Findings

All eight CART-based regression trees modeling soil hydraulic properties (KFS and AEV) for shallow and benched measurements yielded some information to predict high and low air-entry tensions, summarized in Table 3.6.

Table 3.6 Soil conditions and associated average AEV

Soil conditions and associated average AEV for shallow measurements			
Shallow conditions	Shallow AEV (cm water)	Shallow conditions	Shallow AEV (cm water)
CLAY≤8%	-16.75	CLAY>8%	-10.56
SILT≤68% (CLAY≤8%)	-19.61	SILT>68% (CLAY≤8%)	-11.48
STRUCTURE=0,2,3,4 (SILT≤68% & CLAY≤8%)	-20.30	STRUCTURE=5 (SILT≤68% & CLAY≤8%)	-5.61
ORG>5% (SILT>68% & CLAY≤8%)	-19.24	ORG≤5% (SILT>68% & CLAY≤8%)	-9.33
TILL=0 (CLAY>8%)	-14.33	TILL=1 (CLAY>8%)	-9.29
BD>1.44 g cm^{-3} (ORG≤6% & TILL=0 & CLAY>8%)	-58.12	BD≤1.44 g cm^{-3} (ORG≤6% & TILL=0 & CLAY>8%)	-14.93
KFS≤0.02 cm s^{-1}	-20.68	KFS>0.02 cm s^{-1}	-11.97
0.02≤KFS≤0.23 cm s^{-1}	-14.02	KFS>0.23 cm s^{-1}	-8.14
Soil conditions and associated average AEV for benched measurements			
Benched conditions	Benched AEV (cm water)	Benched conditions	Benched AEV (cm water)
CLAY≤2%	-24.65	CLAY>2%	-14.48
2%<CLAY≤21%	-15.48	CLAY>21%	-9.40
TILL=0 (CLAY>21%)	-17.51	TILL=1 (CLAY>21%)	-7.41
KFS≤0.03 cm s^{-1}	-16.76	KFS>0.03 cm s^{-1}	-7.79
KFS≤0.01 cm s^{-1} (CLAY>2%)	-16.97	0.01 cm s^{-1}≤ KFS≤0.03 cm s^{-1}	-13.31

Table 3.7 lists conditions associated with rapid and slow field saturated hydraulic conductivity.

Table 3.7 Soil conditions and associated average K_{fs}

Conditions and associated average K_{fs} for shallow measurements			
Shallow Conditions	Shallow K_{fs} (cm s^{-1})	Shallow Conditions	Shallow K_{fs} (cm s^{-1})
CLAY>19%	0.05	CLAY ≤ 19%	0.01
STRUCTURE = 5 (CLAY≤20%)	0.05	STRUCTURE=0,1,2,3,4 (CLAY≤19%)	0.01
SAND>45% (CLAY>19%)	0.13	SAND≤45% (CLAY>19%)	0.04
TILL= 1 (SAND≤45% & CLAY>19%)	0.05	TILL=0 (SAND≤45% & CLAY>19%)	0.01
AEV>-10.71 cm water	0.04	AEV≤-10.71 cm water	0.01
AEV>-2.55 cm water (CLAY>19%)	0.18	AEV≤-2.55 cm water (CLAY>19%)	0.06
Conditions and associated average K_{fs} for benched measurements			
Benched conditions	Benched K_{fs} (cm s^{-1})	Benched conditions	Benched K_{fs} (cm s^{-1})
CLAY>22%	0.07	CLAY≤22%	0.01
AEV>-7.65 cm water (CLAY>22%)	0.11	AEV≤-7.65 cm water (CLAY>22%)	0.03

3.4 DISCUSSION & CONCLUSIONS

3.4.1 Texture

As can be noticed looking at variable relative importance score tables (Tables 3.4 and 3.5), soil texture is often identified as an indicator of soil hydraulic properties. Texture has been extensively used in the classical pedotransfer approach for estimating soil hydraulic properties (van Genuchten and Leji, 1982; Puckett et al., 1985; Lin et al., 1999; Timlin et al., 1999; Papchepsky et al., 2006). When hydraulic properties are excluded as exploratory variables, CLAY scores 100% for relative importance in terms of primary+surrogate improvement score (see Tables 3.4 and 3.5) in the context of shallow and benched CART-based modeling of KFS and in the context of benched AEV, whereas SAND scores 100% for relative importance in the context of shallow modeling of AEV, and CLAY 50.12%. SAND does not appear as a primary splitter in that regression tree but it is identified as a strong competitor splitter. Classification by texture results in homogeneous K_{sat} classes (King and Franzmeier, 1981). Table 3.7 lists threshold values for clays, silts and sands, and associated average K_{fs}. Fine sand and clay percentages are highly correlated with hydraulic conductivity (Rawls, 1982; Puckett et al., 1985). Average (shallow and benched) KFS can be estimated based on relatively higher or lower soils clay contents. The threshold value for clay contents is not the same for shallow and benched K_{fs}. For shallow measurements, the threshold value is 19.33% and it is 21.63% for benched measurements. Soils with CLAY>19.33% have an average shallow KFS=0.05 cm s^{-1} and soils with CLAY≤19.33% have an average shallow KFS=0.01 cm s^{-1}.

For benched measurements, soils with CLAY>21.63% have an average benched KFS=0.07 cm s^{-1} and soils with CLAY≤21.63% have an average of 0.01 cm s^{-1}. Shallow K_{fs} also stratifies according to SAND=44.55%. Soils with sand contents greater than 44.55% have an average shallow KFS of 0.13 cm s^{-1} and soils with sand contents lower or equal to 44.55% have an average of 0.04 cm s^{-1}. Sandy horizons are usually in the higher K_{sat} classes (McKeague et al., 1982).

3.4.2 Structure

Soil structural organization (presence or absence of preferential pathways; pores) overcomes the effect of texture when it comes to soil water infiltration. Textural-structural conditions are critically important factors determining the magnitude of K_{sat} (Bouma, 1983; Reynolds et al., 2000). The magnitude, range and pattern of variability of the K_{fs} measurements are primarily controlled by the well-developed and stable soil structure at the field scale, rather than by texture, organic carbon or surface topography, as classical pedotransfer functions suggest (Reynolds and Zebchuk, 1996). The classical pedotransfer approach is insufficient to assess macropore flow, which is held responsible for diffuse contamination of surface and groundwater (Lin et al., 1999; Papchepsky et al., 2006). Lin et al., (1999) established that soil structure is crucial in characterizing hydraulic behavior in the macropore flow region whereas texture has its major impact on hydraulic properties controlled by micropores (Lin et al., 1999).

Clay horizons riddled with biopores, especially those having strong fine blocky structure, have K_{sat} values as high or higher than coarse textured soils in which, K_{sat} is controlled, presumably by packing voids (McKeague et al., 1982). Fine textured soils tend to have greater tendency for 'bypass flow' (gravity-driven flow) to occur than coarse textured soils (Lin et al., 1999). Pore size distribution is more uniform in sandy soils, which tends to reduce the likelihood of bypass flow (Lin et al., 1997, 1999).

Horizons having low saturated hydraulic conductivity values are generally massive, compressed, and clayey with very few or no macropores (McKeague et al., 1982). These conditions are represented by structural classifications "0" and "1" within the context of this study. Soils with such STRUCTURE (0 and 1) have an average shallow KFS of 0.01 cm s^{-1} and an average shallow AEV of -16.98 cm water. Abundant biopores, and fine to medium blocky structure are identified as the major factor contributing to high saturated hydraulic conductivity values (McKeague et al., 1982). STRUCTURE is identified in this study as a robust indicator of regional shallow KFS and shallow AEV. A soil with unconsolidated structure (structure code = 5) is estimated to have an average shallow KFS of 0.05 cm s^{-1} and an average shallow AEV of -5.61 cm water. STRUCTURE does not appear as a significant independent predictor for soils hydraulic properties at depth.

3.4.3 Crop and tillage

Understanding macropore flow under a variety of agricultural and land management regimes has been stated has a priority for addressing environmental planning and policy in order to prevent surface and groundwater degradation and contamination (McKeague et al., 1982; Heppell et al., 2002). Tillage and crop types have been included to this analysis as ordinal values in order to explore their relationship to other soil physical and hydraulic properties. TILL relates to KFS as the third strongest correlation for KFS with

a coefficient of 0.31 for shallow measurements and 0.28 for benched ones (see Table 3.3 for details). The correlation coefficient between TILL and AEV is also the third in strength for AEV (0.28 for shallow measurements and 0.22 for benched ones). CROP does not relate strongly to any of the target variables and was not identified as a robust indicator for neither of them. Soil tillage generally increases soil porosity and changes the pore-size distribution, leading to changes in the soil water retention curve and hydraulic conductivities (Ahuja et al., 1998). Tillage impact was verified within the results from this shallow and benched soil hydraulic properties modeling effort. Both shallow and benched AEV modeling was affected by the tillage therapy (Tables 3.4 and 3.6). Long-term-no-till fields can be expected to have a shallow average AEV of -14.33 cm water and a benched one of -17.51 cm water. Tilled fields don't witness such high air-entry tensions; average shallow AEV for tilled fields =-9.29 cm water and benched average AEV = -7.41 cm water. Shallow values of field saturated hydraulic conductivity can be estimated according to the presence or absence of tillage. No-till fields are assumed to have a slower average shallow KFS of 0.01 cm s^{-1} as compared with conventionally tilled fields' shallow average KFS of 0.05 cm s^{-1}.

3.4.4 Hydraulic properties

On one hand, both shallow and benched AEV are best modeled by KFS (scores 100% for shallow and benched models, according to Table 3.4) when considered as an independent predictor of AEV. Slow infiltration rates (KFS≤0.01 cm s^{-1}) are indicative of high air-entry tensions (average AEV = -16.97 cm water). Higher saturated soil air-entry tensions are due to low soil porosity. The air-entry potential provides a measure of the largest continuous pore (Timlin et al., 1999). In other words, this means that slow K_{fs} readings may be indicative of the absence of biopores/macropores, or their low numbers. Biopore abundance is a major factor contributing to higher saturated hydraulic conductivity (McKeague et al., 1982), and biopore abundance could very well be reflected by AEV measurements used in this study. On the other hand, low shallow air-entry tensions (AEV>-2.55 cm water) can be indicative of rapid soil shallow water infiltrations (average K_{fs} = 0.18 cm s^{-1}). Shallow AEV is a robust indicator of shallow KFS but this relationship is not true for benched measurements. KFS and AEV have a higher Spearman's rank correlation coefficient (0.54) for shallow measurements than for benched ones (0.47) (see Table 3.3). Also, in CART-based shallow KFS modeling, AEV scores 100% as an independent predictor when included in the model, but AEV does not appear as an important variable (in terms of primary+surrogate improvement score) for modeling benched KFS (see Table 3.5). Benched KFS appears to be controlled by soil physical properties, namely CLAY, ORG and STRUCTURE.

Even if KFS and AEV are strongly related, and they appear as the most important variable for predicting each other, KFS and AEV may not be the most desirable indicator for surface and groundwater contamination risk. Strong surrogate and competitor variables have been identified as alternative robust indicators of these hydraulic properties, namely CLAY and SAND, for example. Soil physical/crop/management properties are better indicators of macropore flow risk since they are more stable in time and space and easier to measure. Good indicators need to be easily measured, spatially continuous, and tangible for policy makers and other non-experts.

3.4.5 Overview

Potential indicators of macropore flow incidence have been identified in this chapter. Even though the regression trees R^2 are not very high (the highest one being 0.48 for shallow AEV modeling based on soil physical/crop/management properties) some progress has been made towards a better understanding of macropore flow precursors. Saturated soil air-entry tensions (AEV) and field saturated hydraulic conductivities (K_{fs}) appear to be strong mutual predictors. However, AEV and K_{fs} are not representative of the desirable indicators needed to assess surface and shallow groundwater contamination risks at the regional scale. Clay and sand contents offer a more valid alternative to the aforementioned hydraulic properties. Sandy sites tend to exhibit low water retention capacities as water infiltration on such soils is rapid. Clay content is more significant in terms of macropore flow since fine textured soils have a greater tendency to be scattered with biopores than coarser soils. Higher clay contents are associated with faster K_{fs} and lower saturated soil air-entry tensions; this is probably due to the structural organization of clayey soils; i.e., the presence of macropores. Finally, tillage operations modify the soil's structure by loosening it. Looser soils are associated with higher preferential flow incidence even if permanent (nearly vertical) structures, such as earthworm channels, are destroyed by tilling applications. Further research on regional macropore flow incidence could use this study as a guide for hypothesis formulation. More work on the relationships between texture classes and the associated soil structures with regards to soil hydraulic behavior would be of great benefit for the preservation of good water quality.

3.5 REFERENCES

AAFC 2003. Canadian Adaptation and Rural Fund; Canada – Ontario Research Development Program, Performance Report, 2002-2003. http://www.agr.gc.ca/acaaf/card/pdf/councils.aac.cordp.03.pdf

AAFC 2005. Environmental sustainability of Canadian agriculture: agri-environmental indicator report series – report #2. http://www.agr.gc.ca/env/naharp-pnarsa/index_e.php?page=aei

Ahuja, L.R., Fiedler, F., Dunn, G.H., Menjamin, J.G., and Garrison, A. 1998. Changes in soil water retention curves due to tillage and natural reconsolidation. Soil Science Society of America Journal 62: 1228-1233

Arya, L.M., Leji, F.J., Shouse, P.J. and van Genuchten, M., Th. 1999. Relationship between the hydraulic conductivity function and the particle-size distribution. Soil Science Society of America Journal 63: 1063-1070

Bagarello, V. and Sgroi, A. 2004. Using the single-ring infiltrometer method to detect temporal changes in surface soil field-saturated hydraulic conductivity. Soil and Tillage Research 76: 13-24

Ball Coehlo, B.R., Roy, R.C., Topp, E., and Lapen, D.R., 2007. Tile water quality following liquid swine manure application into standing corn. Journal of Environmental Quality 36: 580-587

Beven, K. and Germann, P., 1982. Macropores and water flows in soils. Water Resources Research 18: 1311-1325

Bouma, J., Dekker, L.W., and Muilwijk, C.J. 1981. A field method for measuring short-circuiting in clay soils. Journal of Hydrology 52: 347-354

Bouma, J. 1983. Use of soil survey data to select measurement techniques for hydraulic conductivity. Agricultural Water Management 6: 177-190

Bouma, J. and van Lanen, J.A.J. 1987. Transfer functions and threshold values: from soil to land qualities. Pp. 106-110. In: Beek, K.J. *et al.* (Ed.) Quantified land evaluation. International Institute Aerospace Survey and Earth Science. ITC Publication 6: 106-110

Bouma, J., 1989. Using soil survey data for quantitative land evaluation. In: Stewart, B.A. (Ed.), Advances in soil science, vol. 9, Springer Verlag, New York, pp. 177-213

Bouma, J. 2006. Hydropedology as a powerful tool for environmental policy research. Geoderma 131: 275-286

Bouwer, H. 1978. Pages 123-124 in:Groundwater hydrology. McGrawHill, Toronto.

Breiman, L., Freidman, J., Olshen, R., and Stone, C. Classification and Regression trees. Pacific Grove: Wadsworth, 1984

Culley, J.L.B., 1993. Density and Compressibility pp: 529-530 in: Carter, M.R. (Ed.) Soil sampling and methods, Lewis Publishers, for the Canadian Society of Soil Science, 823pp.

Edwards, W.M., Shipitalo, J.M., Owens, L.B., 1993. Gas, water and solute transport in soils containing macropores: a review of methodology. Geoderma 57: 31-49

Fallow, D.J., and Elrick, D.E., 1996. Field measurement of air-entry and water-entry soil water pressure heads. Soil Science Society of America Journal 60: 1036-1039

Gregorich, E.G, Reynolds, W.D., McGovern, M.A., Culley, J.L.B., and Curnoe, W.E. 1992. Changes in some physical properties with depth in a structurally degraded soil under no-tillage. Soil and Tillage Research 26: 289-299

Hastie, T., Tibshirani, R., and Friedman, J. 2001. The elements of statistical learning: Data mining, inference and prediction, Springer Series in Statistics, New York, pp.266-278

Heiri, O., Lotter, A.F., and Lemcke, G., 2001. Loss on ignition as a method for organic and carbonate content in sediments: reproductibility and comparability of results. Journal of Paleolimnology 25: 101-110

Heppell, C.M., Worrall, F., Burt, T.P., and Williams, R.J. 2002. A classification of drainage and macropore flow in an agricultural catchement. Hydrological Processes 16: 27-46

King, J.J., and Franzmeier, D.P. 1981. Estimation of saturated hydraulic conductivity from soil morphology and genetic information. Soil Science Society of America Journal 45: 1153-1156

Kladivko, E.J., Van Scoyoc, G.E., Monke, E.J., Oates, K.M. and Pask, W., 1991. Pesticide and nutrient movement into subsurface tile drains on a silt loam soil in Indiana. Journal of Environmental Quality 20: 264-270

Lapen, D.R., and Martz, L.W., 1996. An investigation of the spatial association between snow depth and topography in a Prairie agricultural landscape using digital terrain analysis. Journal of Hydrology 184: 277-298

Lapen, D.R., Topp, G.C., Gregorich, E.G., Hayhoe, H.N., and Curnoe, W.E., 2001. Divisive field-scale associations between corn yields, management, and soil information. Soil and Tillage Research 58: 193-206

Lin, H.S., McInnes, K.J., Wilding, L.P., and Hallmark, C.T. 1997. Low tension water flow in structured soils. Canadian Journal of Soil Science 77: 649-654

Lin, H.S., McInnes, K.J., Wilding, L.P., and Hallmark, C.T., 1999. Effects of soil morphology on hydraulic properties II. Hydraulic pedotransfer functions. Soil Science Society of America Journal 63: 955-961

McKeague, J.A., Wang, C., and Topp, G.C. 1982. Estimating saturated hydraulic conductivity from soil morphology. - Soil Science Society of America Journal 46: 1239-1244

Mualem, Y. 1992. Modeling the hydraulic conductivity of unsaturated porous media. Pp. 15-36. In: van Genuchten *et al.* (Ed.) Proc. Int. workshop on indirect methods of estimating the hydraulic properties of unsaturated soils. 11-13 Oct. 1989. U.S. Salinity laboratory and Department of Soil and Environmental Science, University of California, Riverside.

Papchepsky, Y.A., Rawls, W.J., and Lin, H.S., 2006. Hydropedology and pedotransfer functions, Geoderma 131: 308-316

Puckett, W.E., Dane, J.H., and Hajek, B.F. 1985. Physical and minerologicla data to determine soil hydraulic properties. Soil Science Society of America Journal 49: 831-836

Rawls, W.J., Brakensiek, D.L., and Saxton, K.E. 1982. Estimation of soil water properties. Trans. ASAE. 25: 1316-1320

Reynolds, W.D., Elrick, D.E., 1990. Ponded infiltration from a single ring: I. Analysis of steady flow. Soil Science Society of America Journal 54: 1233-1241

Reynolds, W.D., 1993. Saturated hydraulic conductivity: field measurements; chapter 56 pp: 599-613 in: Carter, M.R. (Ed.) Soil Sampling and Methods of Analysis, Lewis Publishers, Boca Raton, Florida, 823 pp.

Reynolds, W.D., and Zebchuk, W.D. 1996. Hydraulic conductivity in a clay soil: two measurement techniques and spatial characterization. Soil Science Society of Americal Journal 60: 1679-1685

Reynolds, W.D., Bowman, B.T., Brunke, R.R., Drury, C.F., and Tan, C.S. 2000. Comparison of tension infiltrometer, pressure infiltrometer, and soil core estimates of saturated hydraulic conductivity. Soil Science Society of America Journal 64: 478-484

Sheldrick, B.H., and Wang, C., 1993. Particle size distribution pp: 499-512 in: Carter, M.R. (Ed.) Soil sampling and methods, Lewis Publishers, for the Canadian Society of Soil Science, 823pp.

Steinberg, D., and Colla, P. CART – Classification and Regression Trees. Sand Diego, CA. Salford Systems, 1997

Timlin, D.J., Ahuja, L.R., Papchepsky, Ya., Williams, R.D., Gimenez, D. and Rawls, W. 1999. Use of Brooks-Corey parameters to improve estimates of saturated conductivity from effective porosity. Soil Science Society of America Journal 63: 1086-1092

Topp, G.C., Zebchuk, W.D., and Dumanski, J., 1980. The variation of in-situ measured soil water properties within soil map units. Canadian Journal of Soil Science 60: 497-509

van Genuchten, M.Th., 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Science Society of America Journal 44: 892-898

van Genuchten, M.Th., and Leji, F. 1992. On estimating the hydraulic properties of unsaturated soils. Pp. 1-14 In M.Th. In: van Genuchten *et al.* (Ed.) Proc. Int. workshop on indirect methods of estimating the hydraulic properties of unsaturated soils. 11-13 Oct. 1989. U.S. Salinity laboratory and Department of Soil and Environmental Science, University of California, Riverside.

Wilson, G.V., and Luxmoore, R.J., 1988. Infiltration, macroporosity, and mesoporosity distributions on two forested watersheds. Soil Science Society of America Journal 52: 329-335

Chapter 4

Tree-Structured Relationships and Interactions Between *In Situ* Soil Physical/Hydraulic Properties, Land Management, and Earthworm Biomass; Eastern Ontario, Canada

Ouellet, G[1]., Lapen, D.R[2]., Sawada, M[1]., Topp, E[3]., and Edwards, M[2].

1. University of Ottawa, Department of Geography, Ottawa, ON.
2. Agriculture and Agri-Food Canada, Ottawa, ON
3. Agriculture and Agri-Food Canada, London, ON

ABSTRACT

Earthworm burrows are dominant preferential flow pathways and it is important to be able to assess earthworm abundance in order to delineate shallow and ground water contamination risk. In this study, we attempt to uncover relationships among worm biomass, soil physical/hydraulic parameters, and soil management variables in order to build robust regionally-based predictions of environmental conditions where worm burrowing activity may or may not influence vadose zone contaminant transport processes. Forty-three agricultural fields with various management treatments and a wide texture range were sampled and analyzed in eastern Ontario, Canada. Earthworms were sampled according to the "hot-mustard method", while infiltration parameters and soil physical properties were measured following peer-reviewed and accepted methodologies. It was found that the presence or absence of tillage induces clear distinctions in earthworm biomass modeling. No-till fields exhibit much higher worm biomass than conventional treatments. No-till fields' earthworm sampled biomass stratifies on the basis of soil physical properties (bulk density and texture) as opposed to conventional soil treatment's samples that organize on the basis of residue coverage. Relatively greater

residue coverage yields relatively higher earthworm biomass. More research is needed to successfully represent the spatial variability of macropores, or earthworm burrows, in order to assess tile and shallow groundwater contamination risk.

4.1 INTRODUCTION

Earthworms are a significant part of the macrofauna in many soils, affecting soil properties and processes through their feeding, casting, and burrowing activities (Kladivko, 2001). Earthworms can promote soil aeration and water infiltration to soil horizons well below cultivation depths via their burrowing activities (Shipitalo and Protz, 1987; Turpin et al., 2007). There is a large body of research that has characterized the role of burrowing activity on the preferential transport of water and contaminants in the vadose zone (Edwards et al., 1990; Schrader and Joschko 1991; Lachnicht et al., 1997). In many soils the largest, deepest, and most continuous macropores are associated with worm burrows produced by *Lumbricus terrestris* (Ehlers, 1975, Beven and Germann, 1982; Edwards et al., 1988; Nuutinen and Butt, 2003). These large pores are often inferred to contribute a large portion of pore-weighted flow and drainage in the vadose zone (Bouma et al., 1982; Shipitalo and Gibbs, 2000). The effect of such porosity on contaminant transport should be considered when selecting a method to apply amendments such as liquid manures and biosolids to land surfaces. It has been shown that liquid amendments can reach tile and groundwater depths within minutes following application as a result of rapid gravity flow through these large burrows (Kladivko et al., 1991, 1999; Lapen et al., 2007), and that disrupting these burrows, among tillage induced increases in surface soil porosity, can reduce, at application, the total mass flux of contaminants to tile drain depths by several orders of magnitude (Edwards et al., 1988; Malone et al., 2003; Shipitalo et al., 2004). Thus the capacity to predict the soil/agricultural management conditions that express this intrinsic burrow-based macroporosity could be critical to informing the type of soil therapy (eg. Pre-tillage of the soil, etc.) required to mitigate against surface and subsurface water contamination resulting from land application practices.

Several soil biodiversity-based studies have found modest relationships between 'state' soil physical properties and worm biomass (Fox, 2003; Fox et al., 2004; Joschko et al., 2006; Chan and Barchia, 2007); but the findings are very sensitive to methodological approaches (Joschko et al., 2006). Moreover, examination of direct relationships between worm activity and *in situ* soil infiltration properties, has primarily focused on direct infiltration through biopores (Edwards et al., 1988, 1990; Lachnicht et al., 1997; Willoughby et al., 1997) rather than recharge rates or matrix flow. Nevertheless, being able to predict worm activity/biomass using empirical models built from soil, land management, climate etc., (Joshko et al., 2006) is naturally extendable to informing beneficial land application practices and other contaminant transport process models (Lefebvre et al., 2005).

In this study, we attempt to uncover relationships among worm biomass, soil physical/hydraulic parameters, and soil management variables with the aim of building

robust regionally-based predictions of environmental conditions where worm burrowing activity influences vadose zone contaminant transport processes. The specific objective of this study is: use machine-driven regression tree techniques to build models to uncover relationships and interactions between *in situ* measured soil hydraulic properties derived from pressure infiltrometers (Reynolds, 1993), worm biomass as estimated using standard soil extraction techniques (http://sci.agr.ca/london/faq/worm-ver_e.htm), tillage/cropping system, and *in situ* measured soil physical variables such as bulk density, soil texture, and organic matter.

4.2 MATERIALS AND METHODS

This study took place on agricultural fields located throughout eastern Ontario, Canada (Minimum latitude: 43.1429°N, maximum latitude: 45.4206°N; minimum longitude: 74.8515°W and maximum longitude: 80.7973°W) representing diverse soil textural classes, soil management, (eg., no-tillage, conventional tillage), and cropping systems (eg., bare fields, corn and soybeans).

4.2.1 Field and Laboratory Measurements

Earthworms were collected in the field between October 3rd and October 24th 2005. The 'hot mustard' worm extraction method used here (http://sci.agr.ca/london/faq/worm-ver_e.htm) is considered to work best for worms that have burrows that open to the soil surface, such as *Lumbricus terrestris* (http://sci.agr.ca/london/faq/worm-ver_e.htm). For this study, metal frames (20 cm wide x 40 cm long x 20 cm high) were inserted into the soil to a depth of 15 cm. Subsequently, 5L of hot mustard/water solution, consisting of 17.5 g mustard to 1L water (http://sci.agr.ca/london/faq/worm-ver_e.htm) was poured over the contained area. After approximately 10 minutes post application, the worms that emerged to the surface were hand collected. The worms were stored in glass jars containing 30% ethanol until they were cleaned and weighed to determine their wet biomass. For each field site, six replicate measurements over a 15 m by 10 m area were made representing surface residue coverage ranging from no cover (0%) to full cover (100%).

Soil cores (5.5 cm diameter core by 10 cm length) used to measure soil bulk densities (Culley, 1993) were collected near individual worm sampling sites at 5 cm depth in each field. Twelve replicates were made per field. The geometric mean of bulk density for each field was used in subsequent statistical analyses. Grab surface soil samples (15 cm diameter by 15 cm depth) were collected at each field site near worm measurement sites for soil texture and organic carbon analyses. For soil texture, organic matter was removed from the samples according to Sheldrick and Wang (1993). Particle fractions greater than 63 μm were determined using wet sieving and the particle fractions smaller than 63 μm were determined using a Lecotrac LT-100 Particle Size Analyser (PSA) (Sedimentology Laboratory, Earth Sciences Sector, Natural Resources Canada, Ottawa). Sub-samples taken from the soil texture samples were subjected to loss-on-ignition (LOI) analysis (Heiri et al., 2001) to determine the organic matter content of the soil.

Pressure infiltrometers were used to measure *in situ* field-saturated hydraulic conductivity (K_{fs}) (Reynolds and Elrick, 1990; Reynolds, 1993) at the soil surface (5 cm) and at 15 cm depth from surface using 10.5 cm diameter by 6 cm deep rings. Multiple head analyses were performed using 10 cm and 40 cm heads. Multiple head analysis provided information on the degree of gravity vs. capillary flow as expressed via the capillary length parameter (α) (Reynolds, 1993). The saturated soil air entry tension (AEV) was measured via a tensimeter (Soil Measurement Systems, Tucson, Arizona, USA) fitted to the pressure infiltrometer according to Fallow and Elrick (1996), and Ball Coelho et al., (2007). Air-entry tensions can be used as an indicator of continuous pore networks and the pore size that allows air into a previously saturated medium (Fallow and Elrick, 1996). The greater the tension, the smaller the contributing pore networks. There were approximately 12 infiltrometer replicates per depth, summing up to 24 replicates per field. The measurements were spaced 1.5 m apart within a 15 m by 10 m area located near the worm biomass measurements. Infiltrometer measurements were made well into the crop growing season to allow for soil structural maturity, especially for the tilled sites (Lapen et al., 2001). All variables included in this analysis are described in Table 4.1.

Table 4.1 Variables description and units

Variable	Description and units
BIOMASS	Wet weight of earthworms (g)
BD	Soil bulk density (g cm^{-3})
ORG	Soil organic matter content (%)
CLAY	Soil clay content (%)
SILT	Soil silt content (%)
SAND	Soil sand content (%)
CROP	Bare field or no new crop (coded 2), corn (coded 1) and soy beans (coded 0).
TILL	Conventional tillage (coded 1), no-tillage (coded 0)
RESIDUE	Amount of residue covering the soil: [0, 5%[(coded 0), [5, 25%[(coded 1), [25, 50%[(coded 2), [50, 75%[(coded 3), [75, 100%[(coded 4) and 100% (coded 5).
AEV	Soil water air-entry values (cm water)
KFS	Field saturated hydraulic conductivity (cm s^{-1})

For each field site, means were employed to reflect 'average' site conditions for subsequent statistical analysis. Detailed summary statistics are shown in Table 4.2.

Table 4.2 Summary statistics of non-categorical variables

Summary statistics of non-categorical variables involved in prediction modelling Of earthworm biomass									
Variable	**N**	**MIN**	**1st Quartile**	**Median**	**Mean**	**3rd Quartile**	**MAX**	**STDEV**	**NA**
BIOMASS (g)	243	0.01	2.37	6.15	9.03	12.95	44.63	8.82	0
KFS (cm s^{-1})**	231	$6.62(10)^{-5}$	$5.63(10)^{-3}$	$9.29(10)^{-3}$	$1.41(10)^{-2}$	$1.61(10)^{-2}$	$5.69(10)^{-2}$	0.01	12
AEV (cmwater)**	243	-22.77	-17.71	-9.86	-11.86	-7.82	-3.74	5.19	0
BD (g cm^{-3})**	243	1.12	1.31	1.41	1.43	1.51	1.71	0.14	0
ORG* (%)	43	1.58	4.38	5.16	5.51	6.56	8.95	1.47	0
CLAY* (%)	43	0.49	4.87	18.01	14.66	20.70	22.73	8.58	0
SILT* (%)	43	4.35	39.79	48.46	53.47	66.74	86.70	16.03	0
SAND* (%)	43	9.05	15.25	25.03	31.81	50.03	95.16	18.53	0

* These variables were measured at the field scale rather than on a site by site basis.

** These variables were measured prior to earthworm biomass sampling and therefore, field averages were used for this analysis rather than individual measurement values.

4.2.2 Statistical Analyses

Summary statistics were generated and non-parametric Spearman Rank correlation analyses were conducted for all variables given in Table 4.1 with the exception of CROP which ranking does not make physical sense. The well documented recursive, test-learn, binary, and non-parametric CART-based regression tree methodology (Breiman et al., 1984; Steinberg and Colla, 1997) was used to uncover relationships (and interactions) between worm biomass (dependent) and a suite of independent variables. A CART-based regression tree model is generated by dividing dependent data groupings recursively into sub-groups on the basis of independent variable splitting criteria (primary data splitters) using simple variance minimizing algorithms (expressed as improvement scores in CART). The dependent variable data are partitioned into a series of descending nodes (right and left child nodes derived from parent nodes). The predicted value of the dependent variable for a specific node is presented here as the average value and the standard deviation of the dependent variable observations assigned to that node. Tree size is constrained by statistics that maximize predictive capacity (i.e. R^2) while minimizing model complexity and size. CART performs cost-complexity pruning to determine the optimal number of terminal nodes (Steinberg and Colla, 1997; Hastie et al., 2001) to avoid model overfitting. An ancillary component to CART is characterizing the variance minimizing capacity of variables that mimic the primary splitting variable for a nodal split on a case by case basis. The variable that is most strongly associated (Breiman et al., 1984) with the primary splitting variable is deemed the strongest surrogate splitting criteria; and surrogate information is useful for identifying potential substitutes for primary splitting variables as well as identifying potential multicollinearity. CART provides an indication of variable importance for a specific tree

model. A variable's relative importance is calculated here as the sum of the primary split and surrogate improvement scores throughout the selected tree. The values are scaled relative to the best performing variable (score of 100) (Steinberg and Colla, 1997). For this study, CART software default parameters were used exclusively, as per Lapen et al. (1996, 2001). Two regression tree models were developed for this study: 1) predict worm biomass from soil physical (eg. CLAY, SILT, SAND, ORG, BD, CROP and RESIDUE) variables given in Table 4.1, 2) predict worm biomass from all variables given above in Table 4.1, which include soil hydraulic variables derived from infiltrometry.

4.3 RESULTS

The 43 fields evaluated in this study reflected a diverse suite of soil textures and soil management conditions (12 no-till fields and 31 conventionally tilled fields). Surface crops consisted of bare fields (2), standing corn (39), and standing soybean (2). The soil conditions in terms of soil water content and soil temperature during the worm biomass measurements are given in Table 4.3.

Table 4.3 Summary statistics of soil conditions during earthworm sampling season

	Soil temperature (°C)	Soil water content (%)
N	43	43
Min	7.8	22.1
Max	16.5	36.1
Average	11.9	30.2
Stand. Dev.	2.6	2.7

Temperatures ranged between 7.7 and 16.5° C and soil water contents ranged between 13.9 and 33.1% volume. Spearman's Rank correlation analysis was conducted for all the data variables used in the analysis to identify correlations among the variables. Spearman's Rank correlations for no-till and tilled sites, specifically (Table 4.4) were also included.

Table 4.4 Spearman's Rank correlation matrix for all variables

	BIOMASS	RESIDUE	ORG	CLAY	SILT	SAND	TILL	CROP	BD	KFS	AEV
BIOMASS											
All data	1	0.35*	0.04	-0.05	-0.09	0.09	-0.32*	-0.11	0.25*	-0.05	-0.12
Till	1	0.32*	0.07	0.08	-0.07	0.03	---	-0.09	0.03	0.11	0.04
No-till	1	0.28*	-0.03	-0.28*	-0.26*	-0.20*	---	-0.41*	0.56*	-0.30*	-0.09
RESIDUE											
All data	N=243	1	0.17*	0.05	0.07	-0.04	-0.17*	-0.06	0.10	-0.02	-0.08
Till	N=176	1	0.15	0.07	0.07	-0.05	---	-0.04	0.04	0	0
No-till	N=67	1	0.32*	0.09	0	-0.03	---	-0.28*	0.09	0.04	0.09
ORG											
All data	N=243	N=243	1	0.65*	0.33*	-0.53*	-0.02	0.12	-0.28*	0.55*	0.38*
Till	N=176	N=176	1	0.63*	0.30*	-0.54*	---	0.29*	-0.29*	0.56*	0.50*
No-till	N=67	N=67	1	0.61*	0.45*	-0.49*	---	-0.33*	-0.34*	0.36*	0.50*
CLAY											
All data	N=243	N=243	N=243	1	0.03	-0.37*	0.20*	0.01	-0.30*	0.66*	0.56*
Till	N=176	N=176	N=176	1	-0.05	-0.36*	---	0.25*	-0.21*	0.69*	0.61*
No-till	N=67	N=67	N=67	1	0.38*	-0.49*	---	-0.27*	-0.53*	0.57*	0.51*
SILT											
All data	N=243	N=243	N=243	N=243	1	-0.89*	-0.12	0.23*	-0.26*	0.15*	0.22*
Till	N=176	N=176	N=176	N=176	1	-0.86*	---	0.04	-0.07	0.06	0.13
No-till	N=67	N=67	N=67	N=67	1	-0.94*	---	0.57*	-0.70*	0.42*	0.81*
SAND											
All data	N=243	N=243	N=243	N=243	N=243	1	0.02	-0.25*	0.45*	-0.42*	-0.47*
Till	N=176	N=176	N=176	N=176	N=176	1	---	-0.18*	0.33*	-0.40*	-0.41*
No-till	N=67	N=67	N=67	N=67	N=67	1	---	-0.41*	0.63*	-0.51*	-0.92*
TILL											
All data	N=243	N=243	N=243	N=243	N=243	N=243	1	-0.29*	-0.14*	0.21*	0.42*
Till	N=176	N=176	N=176	N=176	N=176	N=176	1	---	---	---	---
No-till	N=67	N=67	N=67	N=67	N=67	N=67	1	---	---	---	---
CROP											
All data	N=243	N=243	N=243	N=243	N=243	N=243	N=243	1	-0.41*	0.15*	0.13
Till	N=176	N=176	N=176	N=176	N=176	N=176	N=176	1	-0.36*	0.27*	0.21*
No-till	N=67	N=67	N=67	N=67	N=67	N=67	N=67	1	-0.56*	0.14	0.21
BD											
All data	N=243	N=243	N=243	N=243	N=243	N=243	N=243	N=243	1	-0.48*	-0.24*
Till	N=176	N=176	N=176	N=176	N=176	N=176	N=176	N=176	1	-0.45*	-0.12
No-till	N=67	N=67	N=67	N=67	N=67	N=67	N=67	N=67	1	-0.66*	-0.50*
KFS											
All data	N=231	N=231	N=231	N=231	N=231	N=231	N=231	N=231	N=231	1	0.56*
Till	N=164	N=164	N=164	N=164	N=164	N=164	N=164	N=164	N=164	1	0.55*
No-till	N=67	N=67	N=67	N=67	N=67	N=67	N=67	N=67	N=67	1	0.57*
AEV											
All data	N=243	N=243	N=243	N=243	N=243	N=243	N=243	N=243	N=243	N=243	1
Till	N=176	N=176	N=176	N=176	N=176	N=176	N=176	N=176	N=176	N=176	1
No-till	N=67	N=67	N=67	N=67	N=67	N=67	N=67	N=67	N=67	N=67	1

* Identifies statistically significant correlations (P-values < 0.05)

Most of the strongest correlations appear within the no-till field sites. The strongest correlation of all is between SILT and SAND for the no-till sites with an R= -0.94. The second strongest correlation occurs between SAND and AEV with an R= -0.92, closely followed by the correlation between SILT and AEV (R=0.81) both for no-till sites. For

all data, the strongest correlation is also found between SILT and SAND with an R= -0.89, followed by the correlation between CLAY and KFS (R=0.66). The same correlation pattern can be observed for the tilled sites; with R equal to -0.86 and 0.69 respectively. Interestingly, no-till field sites yield stronger associations with AEV whereas tilled sites tend to exhibit stronger correlations with KFS. For the BIOMASS variable, the strongest correlation is between BIOMASS and BD but this is only true for the no-till sites (R=0.56). For tilled sites, this correlation is negligible (R=0.03). The second strongest correlation including BIOMASS is also found on no-till sites between BIOMASS and CROP (R= -0.41). BIOMASS and RESIDUE interrelate moderately (R=0.35) when all field sites are considered. For more details see Table 4.4.

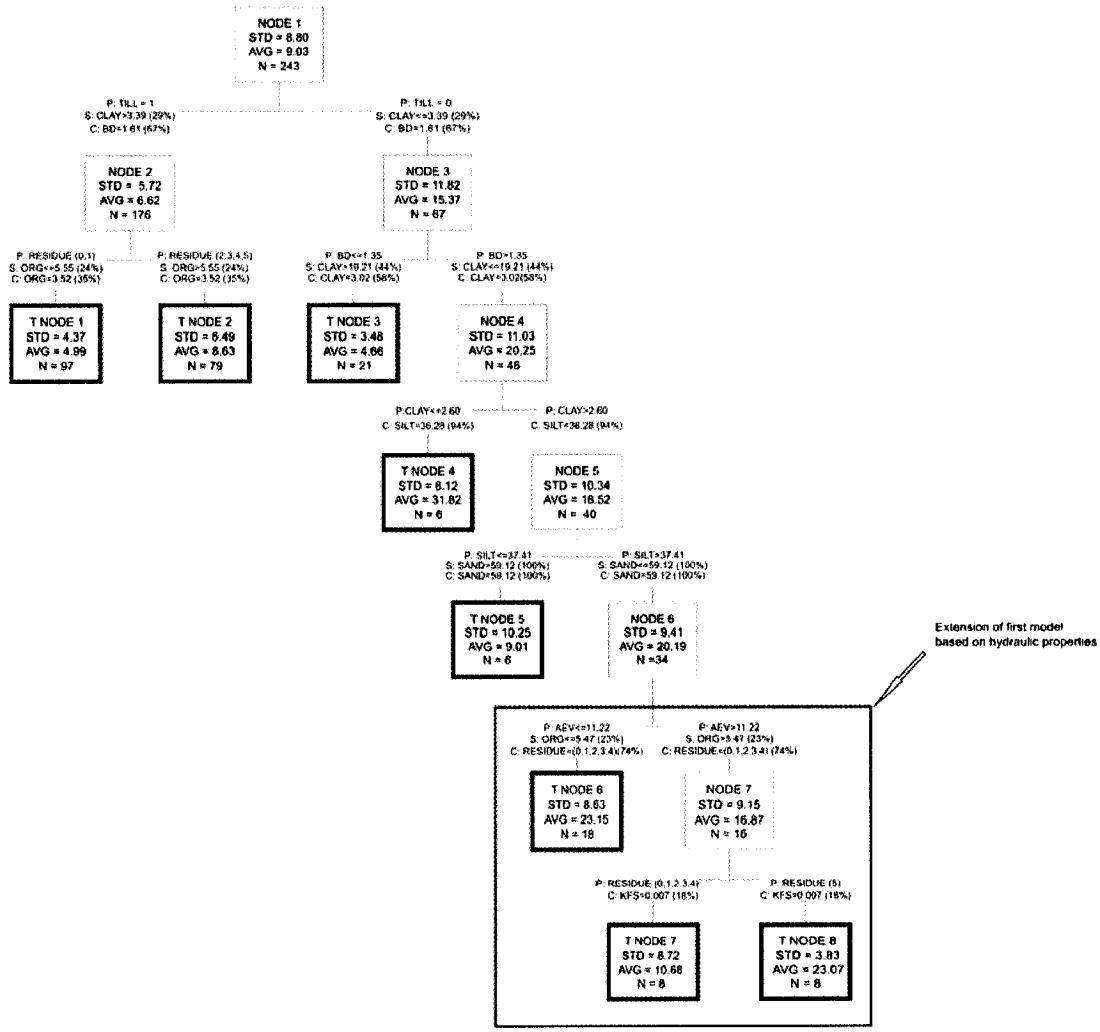


Figure 4.1 Regression tree predicting BIOMASS from SAND, SILT, CLAY, ORG, BD, TILL, RESIDUE, CROP, KFS and AEV

The cross-validated CART-based regression tree predicting BIOMASS from soil texture, organic matter content, bulk density, and soil management variables (see Figure 4.1) contains 6 terminal, or end nodes (designated as T NODE #). The rationale for using these predictor variables without the soil hydraulic data is that the latter can be highly transient in nature and this regression tree modeling effort provides information on relationships and interactions between worm biomass and soil variables that are often considered more temporally stable for spatial scaling endeavors (van Genuchten, 1980; Lin et al., 1999; Papchepsky et al., 2006). Thus the model could be used as a framework to map BIOMASS at regional scales of investigation. Soil hydraulic data, on its own, is less scalable in this manner given its high degree of temporal and spatial variability (Topp et al., 1980; Bararello and Sgroi, 2004; Turpin et al., 2007). The R^2 of the regression tree model is 0.50. The model produced the lowest cross-validated relative error of 0.62±0.07 indicating it was the optimal tree in terms of predictive power and

minimizing model complexity. Table 4.5 (on page 84) indicates that the most important variables in terms of primary+surrogate split improvement scores in the regression tree are BD (100%), CLAY (92%), and TILL (78%). Node 1 includes all of the 243 BIOMASS measurements; the average BIOMASS of all samples is 9.03 g (stdev = 8.80 g). Node 1 data split on the basis of tillage into tilled (NODE 2 with BIOMASS average = 6.62 g) and no-tilled (NODE 3 with BIOMASS average = 15.37 g) nodal groupings. The surrogate and competitor scores were poor to modest for the initial NODE 1 split, suggesting tillage was a robust predictor at this primary level of data support. The 176 tilled data are further (Node 2) subdivided on the basis of relatively 'lower' (T. Node 1) and relatively 'higher' (T. Node 2) RESIDUE classifications; the higher the surface residue (T. Node 2), the higher the average worm biomass. Again, the modest surrogate and competitor scores are indicative of a strong, robust NODE 2 primary splitting criteria. NODE 3 is comprised of 67 no-tilled sites. NODE 3 data were split on the basis of BD = 1.35g cm^{-3}; the lower the bulk density, the lower the average BIOMASS. The surrogate and competitor split criteria are based on CLAY but are modest in terms of variance reducing capacity. NODE 4 data split on the basis of CLAY = 2.60%. Field sites with CLAY $\leq$ 2.60% have an average BIOMASS of 31.82 g with a standard deviation of 8.12 g. The competitor variable (SILT = 36.28%) had a very high improvement score and thus could be interpreted as at least an equally competent split criteria at this stage in the tree.

NODE 5 data further splits on the basis of SILT = 37.41% into T. NODE 5 and NODE 6 (NODE 6 being an actual terminal node in the regression tree built using from all data except soil hydraulic data). The surrogate (SAND = 59.12%) and competitor (SAND = 59.12%) variable improvement scores contributed equally to the primary splitting variable improvement score (100%) indicating equal importance in terms of subdividing NODE 5 data. Essentially, the higher the relative silt content (37.41% threshold value), the higher the worm biomass whereas, for SAND, the lower the relative sand content (59.12% threshold value), the higher the BIOMASS.

Table 4.5 Comparison table of variables' relative importance scores according to the modeling scenario

BIOMASS predicted on the basis of SAND, SILT, CLAY, ORG, BD, TILL, RESIDUE, and CROP		BIOMASS predicted on the basis of SAND, SILT, CLAY, ORG, BD, TILL, RESIDUE, CROP, KFS and AEV	
Variable	Importance score	Variable	Importance score
BD	100	BD	100
CLAY	92.08	TILL	96.05
TILL	78	CLAY	93.42
SILT	53.21	SILT	53.4
SAND	43.66	SAND	44.36
CROP	35.48	KFS	44.04
RESIDUE	21.32	CROP	43.69
ORG	16.31	AEV	36.34
		RESIDUE	31.95
		ORG	2.53

When the soil hydraulic variables were included in regression tree modeling activities with the other soil/crop/management variables, an 8 terminal node regression tree model with a cross-validated relative error of 0.62±0.08 and a R^2 = 0.55 was produced. Essentially the difference between the first model (Figure 4.1) and the one including the soil hydraulic data, is that NODE 6 data splits on the basis of AEV into child nodes with 18 observations (AEV ≤ -11.22 cm water) (T NODE 6) and 16 observations (AEV > -11.22 cm water) (NODE 7). The average BIOMASS is 23.14 g (stdev = 8.63 g) for T NODE 6, while the average BIOMASS is 16.87 g (9.15 g) for NODE 7. The surrogate variable for the NODE 6 split, ORG, has a low improvement score representing 23% of the primary splitter. The first competitor variable, RESIDUE, represents 74% of the AEV improvement score. NODE 7 was subsequently split on the basis of RESIDUE (KFS provided the strongest competitor comprising only 18% of the RESIDUE split criteria improvement score). T NODE 8 was characterized by data with RESIDUE classes 0, 1, 2, 3 & 4; the average BIOMASS for this nodal grouping was 23.07 g (3.83 g). The T NODE 7 data was composed of RESIDUE class 5 data and had an average BIOMASS 10.68 g (8.72 g). There is no surrogate variable for that split (was not significant) and the first competitor for this split is KFS, representing 18% of the primary splitter. The importance scores were essentially similar to those produced for the first regression tree examined (see Table 4.5).

4.4 DISCUSSION & CONCLUSIONS

Joschko et al. (2006) indicated that earthworm populations are better estimated with soil properties such as nitrogen content and pH than soil management practices such as tillage and crop rotation. However, in this study, tillage was an important variable for predicting

worm biomass from all field data. Most of the correlation coefficients are stronger for the no-till field sites (Table 4.4). Tomlin and Miller, (1987) also noted that agricultural practices that cause the greatest changes in populations of soil fauna are deforestation followed by conventional tillage practices. Ploughing activity has adverse effects on worm population density and diversity due to mechanical disturbance of the soil (Kladivko, 2001). For tilled soils, the burrow must be reformed after any tillage operations (Kladivko, 2001) in order for the worm to access food (i.e. surface residues) at the surface. This is less true for no-till situations, a practice that minimizes soil disruption and therefore requirements for burrow reformation. Surface residue is more ubiquitous on the soil surface of no-till systems than it is for tilled systems (Willoughby et al., 1997) where residue materials are incorporated in the plough zone. Hence, as found in this study, worm activity is relatively greater for tilled soils where surface residue is relatively higher. For no-till soils, surface residue is more ubiquitous at the surface (spatially homogenous); therefore, the residue predictor variable was not as important for differentiating worm biomass as it was for tilled soils. Overall, surface residues are less spatially ubiquitous for tilled soils, thus worms tend to cluster in areas where residues are most pronounced, which causes greater statistical differentiation in worm biomass between areas where residues are richer and poorer. The heterogeneity of worms' response on tilled soils is also reflected in Table 4.2: BIOMASS observations for tilled field sites exhibit greater relative standard deviation than for no-till sites (6.23 g±92% vs. 15.37 g±78%). The modal and median residue values are the same for tilled sites; that is to say 1.0. For no-till sites, the modal residue value is also 1.0 but the median residue value = 2.0.

Kladivko et al. (1997) identified statistically significant correlations between shallow dwelling earthworm numbers and soil variables (such as clay content (r=0.62), sand content (r=-0.56) and organic matter (r=0.46)) for no-till sites, but no significant relationships existed between these variables for conventionally tilled observation sites. Similar findings were made in this study, in that regression tree development for the no-till dataset significantly splits on the basis of soil physical properties such as BD, CLAY, SILT, and AEV (when soil hydraulic variables were included in analysis). The conventionally tilled data set, however, only stratified on the basis of residue coverage. Due to the minimal soil disturbance imposed by no-tillage systems, soil properties at the surface will be less temporally transient, relative to conventional tillage system where structural components at the surface are frequently disrupted (Lapen et al., 2004). Hence, it is not surprising that the no-tillage worm biomass data was stratified significantly on the basis of a suite of soil physical properties. For instance, for both regression tree models, soil bulk density was the most important global predictor as a regression tree primary and surrogate splitting variable. The correlation coefficient between bulk density and earthworm biomass was 0.25 for all data; for the no-till sites this correlation was 0.56 and it was negligible (0.03) for tilled sites. In fact, the highest worm biomasses, on average, were associated with no-till management in soils with a bulk density >1.35 g cm^{-3}.

Thompson and Davies (1974) found that high bulk density could compensate for lower soil moisture levels in supporting *Eiseina rosea*'s activity; however, such cause and effect inferences can not be made here. This study found, generally, that worm biomass

was generally poorly correlated with soil texture classes. Thompson and Davies, (1974) indicated that earthworms were more numerous within higher moisture and lower sand content environments; perhaps the latter effect is a result of the lower water holding capacity of sands relative to finer texture soils.

Kladivko et al. (1997) established that on no-till fields, earthworm numbers were positively correlated with clay content and negatively with sand content. Regression tree analysis indicated here that for no-tilled sites where bulk densities were greater than 1.35 g cm^{-3}, greatest groupings of "highest" worm biomass were found where clay content was $\leq$2.6%, and where clay and silt content were respectively >2.6% and >37.4%. While soil textural classes were deemed important variables in regression tree modeling, their impact would appear to be context-dependent and interactive. Thus, texture data alone may not be an adequate predictor of biomass for these agricultural soil systems. As Chan and Barchia (2007) pointed out, many standard soil tests, while useful soil quality indicators (eg. for crops) might not be sensitive indicators of environmental conditions for earthworms. Clay percentage, for example, does not provide a good indication on the soil structure because soils with similar clay percentages can have very different forms of soil structure.

Nevertheless, the regression tree model predicting biomass from soil physical/management variables is a heuristic effort that may help generate additional hypotheses on how to best predict worm biomass from environmental information. Given the many studies that have drawn relationships between worm burrow activity and preferential flow (Beven and Germann, 1982; Edwards et al., 1990, 1993; Lachnicht et al., 1997), prediction of earthworm biomass could be a more stable preferential flow indicator than many *in situ* measures of infiltrability. In fact, Lapen et al. (2007) found that tension and pressure infiltrometer estimates of near and field saturated hydraulic conductivity underestimated the rate of transport of surface applied liquid municipal biosolids to tile drain systems. This was likely a result of inadequate spatial representation of dominant contributing worm burrows. Kladivko et al. (1997) indicated that the presence or absence of a "significant" earthworm population may be an important determiner of whether infiltration rates increase or decrease in no-till vs. conventional systems. Edwards et al. (1990) found that infiltration and therefore, potential groundwater recharge, could increase by more than 100 mm yr^{-1} in watersheds farmed with no-till practices as compared to similar fields that were conventionally tilled. The difference in hydrologic response appeared to be due to differences in soil physical properties, possibly those attributable to earthworm activity.

Along these lines we attempted to evaluate direct relationships among worm biomass and *in situ* infiltration properties. Globally, the correlations between field saturated hydraulic conductivity and biomass were small and negative, although, for no-till sites, BIOMASS interrelates with KFS with a coefficient of R=-0.30. The correlation between AEV and KFS was modestly positive (R = 0.56 for all sites and R = 0.57 for no-till sites) suggesting that soil water infiltration processes may be driven, or influenced by the effective porosity. AEV is an indicator of the maximum continuous network pore radii and therefore, could be used as an indication of burrowing activity. Nuutinen et al.

(2001) found that infiltration rates were positively correlated to *L. terrestris* numbers (R=0.31) and biomass (R=0.33). According to our findings, these variables are negatively correlated (Table 4.4) (r=-0.16 and r=-0.22) with KFS and AEV respectively. Infiltration via the soil matrix is slow when compared to water flow in macropores (Chan, 2004). Therefore, Pressure Infiltrometer measurements tend to underestimate *in situ* K_{fs}. The effects of earthworm abundance/activity on burrow continuity/stability and hence infiltration rate are not clear (Chan, 2004) and therefore, it is difficult to grasp the spatial variability of macropores leading to preferential flow. Shipitalo and Butt (1999) also concluded that no significant relations between infiltration rates and burrow geometrical properties or earthworm weight could be detected (Shipitalo and Butt, 1999).

In situ data relating earthworm biomass and soil water infiltration parameters are very scarce and therefore, it becomes difficult to compare our findings to the literature. This project confirms that preferential flow is a complex and dynamic process that cannot be estimated by soil "state" properties only. Soil physical and managerial characteristics are useful indicators of earthworm biomass but are not sufficient to derive an estimation of the infiltration rates. First, we learnt that the absence of tillage favors the development and preservation of burrowing earthworm's environment, maintaining macropores and channels continuously opened, available for preferential flow mechanisms to occur. Also, on conventionally tilled fields, the residue coverage is an important indicator of earthworm's abundance probably because the mulch cover offers them food and protection. Finally, coarse-textured soils appear to be uninviting for earthworm communities, probably because these soils have lower water holding capacities than fine-textured soils.

In this attempt to uncover robust relationships between earthworm biomass and soil water transport rates, we have to conclude that we could not represent the spatial variability of macropores. Still, we achieved, as a starting point for future research, to express these soils parameters to an adequate degree.

4.6 REFERENCES

Bagarello, V. and Sgroi, A. 2004. Using the single-ring infiltrometer method to detect temporal changes in surface soil field-saturated hydraulic conductivity. Soil and Tillage Research 76: 13-24

Ball Coehlo, B.R., Roy, R.C., Topp, E., and Lapen, D.R., 2007. Tile water quality following liquid swine manure application into standing corn. Journal of Environmental Quality 36: 580-587

Beven, K. and Germann, P., 1982. Macropores and water flows in soils. Water Resources Research 18: 1311-1325

Bouma, J., Belman, C.F.M., and Dekker, L.W., 1982. Water infiltration and redistribution in a silt loam subsoil with vertical worm channels. Soil Science Society of America Journal 46: 917-921

Breiman, L., Freidman, J., Olshen, R., and Stone, C. Classification and Regression trees. Pacific Grove: Wadsworth, 1984

Chan, K.Y., 2004. Impact of tillage practices and burrow of a native Australian anecic earthworm on soil hydrology. Applied Soil Ecology 27: 89-96

Chan, K.Y. and Barchia, I., 2007. Soil compaction controls the abundance, biomass and distribution of earthworms in a single dairy-farm in south-eastern Australia. Soil and Tillage Research 94: 75-82

Culley, J.L.B., 1993. Density and Compressibility pp: 529-540 in: Carter, M.R. (Ed.) Soil sampling and methods, Lewis Publishers, for the Canadian Society of Soil Science, 823pp.

Edwards, W.M., Shipitalo, M.J., and Norton, L.D., 1988. Contributions of macroporosity to infiltration into a continuous corn no-tilled watershed: implications for contaminant movement. Journal of Contaminant Hydrology 3: 193-205

Edwards, W.M., Shipitalo, M.J., and Norton, L.D. 1988. Contributions of macroporosity to infiltration into a continuous corn no-tilled watershed: implications for contaminant movement. Journal of contaminant hydrology 3: 193-205

Edwards W.M., Shipitalo M.J., Owens L.B. and Norton L.D., 1990. Effect of *Lumbricus terrestris* L. burrows on hydrology of continuous no-till corn fields. Geoderma 46:73-84

Edwards, W.M., Shipitalo, J.M., Owens, L.B., 1993. Gas, water and solute transport in soils containing macropores: a review of methodology. Geoderma 57: 31-49

Ehlers, W., 1975. Observations on earthworm channels and infiltration on tilled and untilled loess soil. Soil Science 119: 242-249

Fallow, D.J., and Elrick, D.E., 1996. Field measurement of air-entry and water-entry soil water pressure heads. Soil Science Society of America Journal 60: 1036-1039

Fox, C.A., 2003. Characterizing soil biota in Canadian agroecosystems: state of knowledge in relation to soil organic matter. Canadian Journal of Soil Sciences 83: 245-257

Fox, C.A., Jarvis, I., Behan-Pelletier, V., Dalpe, Y., Clapperton, J., Prevost, D., Joschko, M., and Lentzsch, D., 2004. Progress towards developing a soil biodiversity indicator for Canada. In: Francaviglia, R. (Editor) Agricultural impacts on soil erosion and soil biodiversity: developing indicators for policy analysis, proceedings from an OECD expert meeting. Rome, Italy, March 2003. Available online at: http//www.oecd.org/agr/env/indicators.htm pp. 439-451

Hastie, T., Tibshirani, R., and Friedman, J., 2001. The elements of statistical learning: Data mining, inference and prediction, Springer Series in Statistics, New York, pp. 266-278

Heiri, O., Lotter, A.F., and Lemcke, G., 2001. Loss on ignition as a method for organic and carbonate content in sediments: reproductibility and comparability of results. Journal of Paleolimnology 25: 101-110

Joschko, M., Fox, C.A., Lentzsch, P., Kiesel, J., Hierold, W., Kruck, S., and Timmer, J., 2006. Spatial analysis of earthworm biodiversity at the regional scale. Agriculture, Ecosystems and Environment 112: 367-380

Kladivko, E.J., Van Scoyoc, G.E., Monke, E.J., Oates, K.M. and Pask, W., 1991. Pesticide and nutrient movement into subsurface tile drains on a silt loam soil in Indiana. Journal of Environmental Quality 20: 264-270

Kladivko E.J., Akhouri N.M. and Weesies G., 1997. Earthworm populations and species distributions under no-till and conventional tillage in Indiana and Illinois. Soil Biology and Biochemistry 29: 613-615

Kladivko, E.J., Grochulska, J., Turco, R.F., Van Scoyoc, G.C. and Eigel, J.D., 1999. Pesticide and nitrate transport into subsurface tile drains of different spacings. Journal of Environmental Quality 28: 997-1004

Kladivko, E.J., 2001. Tillage systems and soil ecology. Soil & Tillage Research 61: 61-76

Lachnicht, S.L., Parmelee, R.W., McCartney, D., and Allen, M., 1997. Characteristics of macroporosity in a reduced tillage agroecosystems with manipulated earthworm populations: Implications for infiltration and nutrient transport. Soil Biology and Biochemistry 29: 493-498

Lapen, D.R., and Martz, L.W., 1996. An investigation of the spatial association between snow depth and topography in a Prairie agricultural landscape using digital terrain analysis. Journal of Hydrology 184: 277-298

Lapen, D.R., Topp, G.C., Gregorich, E.G., Hayhoe, H.N., and Curnoe, W.E., 2001. Divisive field-scale associations between corn yields, management, and soil information. Soil and Tillage Research 58: 193-206

Lapen, D.R., Topp, G.C., Edwards, M.E., Gregorich, E.G., and Curnoe, W.E. 2004. Least limiting water range indicators of soil quality and corn production, eastern Ontario, Canada. Soil and Tillage Research 78: 151-170

Lapen, D.R., Topp, E., Edwards, M., Sabourin, L., Curnoe, W., Gottschall, N., Bolton, P., Rahman, S., Ball-Coehlo, B., Payne, M., Kleywegt, S., and McLaughlin, N., 2007. A comparison of two methods for land applying liquid municipal biosolids with respect to potential contamination of tile and groundwater with nutrients and bacteria. Journal of Environmental Quality. *In Press.*

Lefebvre, A., Eilers, W., and Chunn, B. (Editors), 2005. Environmental sustainability of Canadian agriculture –Agri-environmental indicators report series, report 2. Agriculture and Agri-food Canada, Ottawa, Ontario, 232 pages. Available online at: http://www.agr.gc.ca/env/naharp-pnarsa/pdf/2005_AEI_report_e.pdf

Lin, H.S., McInnes, K.J., Wilding, L.P., and Hallmark, C.T., 1999. Effects of soil morphology on hydraulic properties II. Hydraulic pedotransfer functions. Soil Science Society of America Journal 63: 955-961

Malone, R.W., Logsdon, S., Shipitalo, M.J., Weatherington-Rice, J., Ahuja, L., and Ma, L., 2003. Tillage effect on macroporosity and herbicide transport in percolate. Geoderma 116: 191-215

National Science Programs, Agriculture and Agri-Food Canada
http://sci.agr.ca/london/faq/worm-ver_e.htm

Nuutinen, V., Poyhönen, S., Kejota, E. and Pitkänen, J., 2001. Abundance of the earthworm *Lumbricus terrestris* in relation to subsurface drainage pattern on a sandy clay field. European Journal of Soil Biology 37: 301-304

Nuutinen, V. and Butt, K.R., 2003. Interactions of Lumbricus terrestris L. burrows with field subdrains. Pedobiologia 47: 578-581

Papchepsky, Y.A., Rawls, W.J., and Lin, H.S., 2006. Hydropedology and pedotransfer functions, Geoderma 131:308-316

Reynolds, W.D., Elrick, D.E., 1990. Ponded infiltration from a single ring: I. Analysis of steady flow. Soil Science Society of America Journal 54: 1233-1241

Reynolds, W.D., 1993. Saturated hydraulic conductivity: field measurements; chapter 56 pp: 599-612 in: Carter, M.R. (Ed.) Soil Sampling and Methods of Analysis, Lewis Publishers, Boca Raton, Florida, 823 pp.

Schrader, S., and Joschko, M., 1991. A method for studying the morphology of earthworm burrows and their function in respect to water movement. Pedobiologia 35: 185-190

Sheldrick, B.H., and Wang, C., 1993. Particle size distribution pp: 499-512 in: Carter, M.R. (Ed.) Soil sampling and methods, Lewis Publishers, for the Canadian Society of Soil Science, 823pp.

Shipitalo, J.M., and Protz, R., 1987. Comparison of morphology and porosity of a soil under conventional and zero tillage. Canadian Journal of Soil Science 67: 445-456

Shipitalo, M.J., and Butt, K.R., 1999. Occupancy and geometrical properties of Lumbricus terrestris L. burrows affecting infiltration. Pedobiologia 43: 782-794

Shipitalo, M.J., and Gibbs, F., 2000. Potential of Earthworm burrows to transmit injected animal wastes to tile drains. Soil Science Society of America Journal 64: 2103-2109

Shipitalo, J.M., Nuutinen, V., Butt, K.R., 2004. Interaction of earthworm burrows and cracks in a clayey, subsurface-drained, soil. Applied Soil Ecology 26: 209-217

Steinberg, D., and Colla, P. CART – Classification and Regression Trees. Sand Diego, CA. Salford Systems, 1997.

Thompson A.J. and Davies D.M., 1974. Mapping methods for studying soil factors and earthworm ditributions. Oikos 25: 199-203

Tomlin, A.D., and Miller, J.J., 1987. Composition of the soil fauna in forested and grassy plots at Delhi, Ontario. Canadian Journal of Zoology 65: 3048-3055

Topp, G.C., Zebchuk, W.D., and Dumanski, J., 1980. The variation of in-situ measured soil water properties within soil map units. Canadian Journal of Soil Science 60: 497-509

Turpin, K.M., Lapen, D.R., Robin, M.J.L., Topp, E., Edwards, M., Curnoe, W.E., Topp, G.C., Ball Coehlo, B., and Payne, M. Slurry-application implement tine modification of soil hydraulic properties under different soil water content conditions for silt-clay loam soils. Soil and Tillage Research (2007), doi:101016/j.still.2006.12.001

van Genuchten, M.Th., 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Science Society of America Journal 44: 892-898

Willoughby, G.L., Kladivko, E.J. and Savabi M.R., 1997. Seasonal variations in infiltration rate under no-till and conventional (disk) tillage systems as affected by *Lumbricus terrestris* activity. Soil Biology and Biochemistry 29: 481-484

Chapter 5

Conclusion

Preferential (or macropore) flow poses a serious threat for surface and shallow groundwater contamination and degradation. Regional/landscape easily-interpretable indicators of surface and shallow groundwater contamination potential would be a useful tool to better inform policy-making regarding water and environmental preservation. Agriculture and Agri-Food Canada (AAFC) undertook the development of agro-environmental indicators to assess how well agricultural systems manage and conserve natural resources and determine how compatible they are with the natural systems and processes in the broader environment. The response to the tragedy of Walkerton (2000) is a concrete example of the seriousness attached to water quality. Since then, great political efforts have been directed to the better understanding of diffuse pollution to surface and shallow groundwater by agricultural activities. This thesis explored the possibility of using CART-based regression tree models to identify potential indicators of regional preferential flow mechanisms.

The leading question for this thesis was: what significant relationships exist among soil physical properties, crop systems, tillage therapy and *in situ* soil hydraulic behaviour? This was an exploratory, hypotheses-forming endeavour. Seventy-eight field sites were analysed in terms of texture, organic matter, bulk density, field saturated hydraulic conductivity (K_{fs}) and soil saturated air-entry tensions (AEV). In addition, the earthworm biomass of 43 of these field sites was assessed in order to derive the relationships relating it to field-measured soil physical properties. Spearman's Rank correlations were computed for all variables included in the analyses and CART (a data-mining technique) was used to derive regression trees, modeling soil hydraulic properties (K_{fs} and AEV) and worm biomass based on *in situ* soil physical/crop/management properties.

A total of ten CART-based regression trees were described and analyzed. Their maximized predictive capacities (R^2) range from 0.28 to 0.55, suggesting that further work is needed to improve the validity of identified robust indicators of macropore flow

incidence. K_{fs} and AEV modeling efforts yielded some leads for macropore flow precursors. High K_{fs} were identified on high sand content soils and more so, on tilled fields. Tillage loosens the soil surface and increases soil porosity, allowing soil water transport processes to occur more rapidly. Clay was also associated with high K_{fs} and low AEV. The presence of (abundant) macropores in fine-textured soils may be responsible for gravity flow to occur. The soil structural organization must be accounted for when trying to estimate regional soils' infiltrability; texture alone misrepresents soil's macroporosity ant the associated preferential flow risk. Texture is a better indicator of soil water processes though micropores. More than tillage, the presence of (nearly-vertical) worm burrows is responsible for preferential flow events. Low AEV precursors were also identified in the study and they are very similar to the precursors of rapid infiltrations. This is due to their strong multicollinearity.

In an attempt to represent the regional variability of biopores (open earthworm burrows) field measured earthworm biomass was modeled accordingly to *in situ* soil physical, managerial and hydraulic properties. Tillage activities disrupt worm channels and integrate surface organic residues into the soil matrix creating a less favourable environment for earthworms than no-tillage systems. Greater worm biomasses were observed on no-till field sites. Due to minimal soil disturbances imposed by the absence of tilling activities, soil surface properties are less temporally transient and consequently, the modeled earthworm biomass, for no-till treatments, stratified according to numerous physical properties, as opposed to the conventionally-tilled dataset that can be estimated based on the residue coverage. On no-till fields, surface residues are more ubiquitous than on tilled sites. Greater residue cover offers food supply and protection to the earthworms, favouring greater sampled biomasses. Texture appeared to be an inadequate predictor of earthworm biomass since it is interactive and context dependent, giving strength to the necessity of considering soil structure in order to assess macropore flow incidence. Worm biomass would be a stable indicator of preferential flow mechanisms but since it is an endothermic being, it is extremely sensitive to soil temperature and humidity conditions. To avoid the necessity of earthworm sampling, AEV could be used as an indicator of burrowing activity, since it represents continuous pore networks. Additionally, AEV relates strongly to K_{fs} (Spearman's Rank correlation coefficient = 0.56), once more proving that soil's voids control infiltration rates. Nevertheless, valid indicators of preferential flow would ideally be easily measured and understandable for 'non-experts', discrediting the value of AEV as a desirable robust regional macropore flow indicator. The regression trees presented in this thesis identify potential robust indicators of macropore flow (in the form of high Kfs, low AEV and high earthworm biomass) and these independent variables could serve as guidelines to mitigate groundwater contamination risks.

The spatial representativity of this study has been impaired by the difficulties to obtain access to the land and by theoretical assumptions but still, the indicators that were identified here provide a good sense of better and worse agricultural practices in terms of water conservation. Further work on regional/landscape soil's hydraulic properties with an emphasis of the structure could yield beneficial discoveries relating to best-

management practices and policy-making regarding water and environmental preservation.

Launching the 'Decade of water for life', the United Nations took a firm stand on earth's water essentiality. Being able to assess regional preferential flow mechanisms represents the first step towards better-informed environmental policy-making and therefore, towards the preservation of our most valuable resource: water.