NEW APPROACHES TO DETERMINE RELATIONSHIPS BETWEEN SPATIO-TEMPORALLY VARIABLE CROP YIELDS AND ROOT ZONE SOIL INFORMATION

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Rintaro Kinoshita

August 2016

NEW APPROACHES TO DETERMINE RELATIONSHIPS BETWEEN SPATIO-TEMPORALLY VARIABLE CROP YIELDS AND ROOT ZONE SOIL INFORMATION

Rintaro Kinoshita, Ph.D.

Cornell University 2016

Spatio-temporally variable crop yields exist, which makes it difficult for growers to adjust their inputs, while reductions in yields are both economic and food security concerns. In this dissertation, we focused on transition and subsoil layers that are below the usually sampled top 15 cm of soil profiles, which have been identified as important sources of soil moisture and nutrients, especially under moisture-limited conditions. This dissertation contained six chapters and each contained one or more of the following specific goals: i) developing approaches to utilize yield monitor information to effectively assess within-field spatio-temporal yield variations, ii) developing laboratory and in-situ soil property estimation and mapping approaches for soil profile-scale biological, chemical, and physical properties, and iii) addressing the effects of crop and soil management on soil profile properties.

A multivariate statistical approach of standardized principal component analysis (stdPCA) was applied to grain harvester yield monitor information from farm fields in the Mid-Atlantic US. This approach successfully determined characteristic within-field yield patterns and their relationship to in-season precipitation conditions.

In the same region, in situ estimation of soil-profile properties using a proximal sensor of apparent electrical conductivity showed promising results for mapping subsoil water retention

parameters especially within the Coastal Plain physiographic province. However, in situ estimation of soil organic matter (SOM) related properties using visible and near-infrared reflectance spectroscopy were confounded by clay mineralogical differences. This was further confirmed in the study using 1977 US samples that showed the effects of Fe-oxide related spectral peaks on SOM and soil texture prediction.

The proximal sensor estimated soil profile information and topographical variables were related to stdPCA processed yield data using Random Forest, which helped to determine potential yield constraints and soil productivity.

Finally, soil profile scale assessment of long-term maize cropping systems suggested no-till cropping systems to allow better access of subsoil nutrients and water, but unfertilized nutrients can be mined in the long-term especially when residue is harvested.

In conclusion, this dissertation emphasized the importance of within-field variable profile-scale soil information on site-specific crop and soil management decision making, and presented practical approaches for assessment, data application, and soil conservation strategies.

BIOGRAPHICAL SKETCH

Rintaro Kinoshita was born in Tokyo, Japan to parents Tetsuo and Makiko Kinoshita. His father spent his career as a translator of English and French on art related publications while his mother worked as a kimono (traditional Japanese clothing) tailor. He has a younger sister, Natsume. After high school in both Tokyo, Japan and Cardiff, Wales, Rintaro attended the University of Wales in Bangor, North Wales where he received BS in agriculture in 2008.

After graduation, Rintaro spent 6 months in Costa Rica as a research assistant for a project led by CIRAD (French organization for agricultural research for development). After a brief time working as a farm trainee in Otofuke, Japan, Rintaro began his MS/Ph.D. studies under the supervison of Prof. Harold van Es.

Rintaro's passion for soil science came from his experience in growing up in the countryside of Japan during his youth and from an inspirational book written by Lester Brown called "Who Will Feed China?".

He now works at Obihiro University of Agriculture and Veterinary Medicine in Obihiro, Japan as an assistant professor in soil scince.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to Prof. Harold van Es, my major advisor and great mentor during the 6 years I spent at Cornell. This body of work would not exist without his tremendous academic support throughout my time in Ithaca. I also would like to acknowledge Mr. Bob Schindelbeck, my mentor from my very first days at Cornell. He not only gave me a lot of great advice and help in the field and in the laboratory, but he also always gave me extraordinary ideas for tackling challenging research questions. My minor adviser Dr. David Rossiter opened my eyes to geostatistics and gave his time to explain the deep details of the science. I also would like to thank Dr. Jeff Melkonian for his willingness to support whenever I needed his help. Prof. Timothy Fahey, too, provided me with great insights for improving the contents of this dissertation.

Throughout the course of my study, I had tremendous support from my friends. I would like to thank Sonam Sherpa for all the great discussions and distractions, which ended up giving me fresh ideas for research in soil science. I want to extend thanks to Eduardo Carrillo, Kirsten Kurtz, Yuko Jingu, Eliel Ruiz May, Francis Chen, Souvik Chakrabarty, Yuuka Kihara, Satoshi Ito, and Shiori Kano. Also, this work would not have been possible without the support from my family.

Further acknowledgement is required for Dr. Olivier Roupsard, who encouraged me to pursue my graduate studies at Cornell and provided me with continuous support.

I am grateful to all my funding bodies: Willard Agri-Service Inc., Heiwa Nakajima Foundation, Saltonstall Family, Cornell Graduate School, NRCS, World Bank, EU, and the government of Japan.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 GENERAL BACKGROUND

Understanding within-field spatial variation of yield constraints for precision agriculture (PA) has the potential to improve crop yields and reduce nutrient losses to the surrounding environment by adjusting the necessary inputs and alleviating identified yield constraints (Oliver, 2010; Schafer et al., 1984). The concept has been successfully applied on small-scale farms for generations by growers who recognized and managed crop production relative to local soil conditions and other environmental factors in subplots (Heuvelink and Webster, 2002; Webster, 1997). For current large scale farming, this concept has been largely motivated by technological development of farm machinery including remote sensing equipment (Norman, 2013; Schueller, 1997) that allow variable rate soil nutrient management (Ma et al., 2013), pest control, and hybrid selection (Katsvairo et al., 2003). Nevertheless, the adjustments of agronomic inputs are justified only when important within-field yield constraints can be identified, and the relationship between these constraints and crop responses are verified, otherwise the variable rate management can have adverse effects (Baveye and Laba, 2015).

In order to justify the need for variable rate crop management, the magnitude of spatio-temporal grain yield variation also needs to be determined (Baveye, 2002), which can be affected by factors such as topography (Basso et al., 2009; Jiang and Thelen, 2004), genotype (Yang et al., 2009), management (Katsvairo et al., 2003), soil properties (Jiang and Thelen, 2004; Miller et al., 1988; Shahandeh et al., 2005), and their interactions. The concept of yield gap assessment (van Ittersum et al., 2013) can be effectively applied to PA, which has not been explicitly assessed to date even though within-field variations of soil properties (Cambardella et al., 1994), crop yield

levels, and temporal yield stability exist (Abuzar et al., 2004; Lawes et al., 2009). Yield gap assessment quantifies the gap between actual yields and yield potential (Yp) or water-limited yield potential (Yw), which is the Yp in rainfed systems (van Ittersum and Rabbinge, 1997). There are several different methodologies that are used for yield gap assessment including the use of Soil-Plant-Atmosphere (SPA) system models (Grassini et al., 2015; Hochman et al., 2012), upper percentiles of growers' yield distributions, maximum measured yields in experimental stations or yield contests (van Ittersum et al., 2013), and the use of baseline functions (Calviño et al., 2003; Calviño and Sadras, 1999). Models have been used to assess Yp or Yw at national (Hochman et al., 2012) and regional scales (Grassini et al., 2015), and some have a functionality of landscape-scale runoff and erosion (Bonilla et al., 2008, 2007; Molling et al., 2005) but the input data required for these models are extensive. The existing soil information such as the Natural Resources Conservation Service (NRCS) Soil Survey database can help to provide required information such as soil texture, soil organic matter content, bulk density, and the depth to water table (B. N. Moebius-Clune et al., 2014; Molling, 2011; Yang et al., 2013) but may not have appropriate spatio-temporal resolution when considering soil properties that can be altered in a short time period by management or areas smaller than the minimum delineation size (Zhu et al., 2001). Therefore, easily accessible and practical methods are needed to acquire the necessary information about soil properties, topography, and climate (Bianchini and Mallarino, 2002; Kweon et al., 2013; Lauzon et al., 2005).

The crop growth factors can be divided into yield defining factors (i.e. hybrid, growing season temperature, and growing season incident solar radiation), yield limiting factors (i.e. water and nutrients), and yield reducing factors (i.e. pests and diseases; van Ittersum et al., 2013; van Ittersum and Rabbinge, 1997). The availability and the cost of obtaining the data for

identifying the constraining growth factors, as well as the management practices to remove them, depend upon their spatio-temporal variability. Therefore, there is a need to i) distinguish manageable yield variations in spatially-referenced grain yield data (Pringle et al., 2003), ii) identify regional scale yield constraining factors, and iii) identify within-field scale yield constraining factors.

In crop production world-wide, the susceptibility to drought conditions is one of the major yield-constraining factors where reduction in yield gaps is possible (Gaiser et al., 2012; Kirkegaard et al., 2007; Passioura and Angus, 2010). At present, soil property assessment in individual fields is largely based on soil nutrient concentrations and/or soil organic matter (SOM) contents at the surface layer (Karlen et al., 2001; Moebius-Clune et al., 2008; Sojka et al., 2003; Spectrum Analytic Inc, 2010). This offers a very small sample size per field, which does not provide the necessary information for the estimation of in-field variable Yw. Shallow soil sampling has been justified due to both the difficulties of sampling to deeper depths and because of the relative importance of topsoil when adequate growing conditions are met (Kautz et al., 2013; Kirkegaard et al., 2007). Consequently, subsoil has been largely ignored for soil management decision-making. However, under drought conditions, plant roots need to access water and nutrients deeper in the profile, if available, in order for the crop to approach Yw (Garz et al., 2000; Kautz et al., 2013; Mengel and Scherer, 1981). Peigné et al. (2013) described this layer at the interface between topsoil and the part of subsoil most affected by soil management, the "transition layer", and identified the physical properties of this layer as being important for assessing a root development potential. Therefore, there is a need to evaluate the importance of acquiring soil property information at depth for assessing Yw. Targeted soil analyses may make it possible to obtain important properties for Yw, such as available water capacity (AWC),

rooting depth, and topography (van Ittersum et al., 2013). In this circumstance, proximal sensing equipment, including visible-near-infrared reflectance spectroscopy (VNIRS; Kinoshita et al., 2016, 2012; Kuang et al., 2012) and apparent soil electrical conductivity (Corwin and Lesch, 2005), has a potential in predicting the targeted soil property information and increasing the spatial as well as temporal coverage (Viscarra Rossel et al., 2011).

1.2 RESEARCH OBJECTIVES

The main objectives of this dissertation are to investigate the spatio-temporal grain yield variations in maize/soybean production fields in the Mid-Atlantic US, and to identify potential yield constraints in those fields. In order to achieve the spatial coverage of soil information, the use of proximal sensing equipment and geostatistics in predicting soil physical, chemical, and biological property values were investigated both in-situ and in laboratory. Lastly, the impacts of long-term no-till and residue return on soil property values at the topsoil, transition layer, and soil horizons below the transition layer at a plot scale were assessed.

The specific objectives were:

- To assess area specific behavior of spatio-temporal yield variation for maize and soybeans in two physiographic provinces in the Mid-Atlantic US
- To determine potential yield constraining edaphic and topographical factors in the region.
- To assess the usefulness of in-situ proximal soil sensing equipment in the region.

- To investigate the predictability of laboratory-based VNIRS in assessing soil physical, chemical, and biological property values as a substitute for conventional analytical assessments.
- To investigate the effects of 40-years of no-till management and residue return on soil physical, chemical, and biological properties in the topsoil, transition layer, and soil horizons below the transition layer.
- To compare strategies to predict the spatial variation of topsoil soil organic carbon combining VNIRS, geostatistics, and data mining approaches.

1.3 DISSERTATION OUTLINE

This dissertation consists of four interrelated projects with the overall objective of improving within-field yield gap analysis by incorporating spatially referenced grain yield and soil profile information.

The first project (Chapters 2 – 4) was in collaboration with Willard Agri-Service (Frederick, MD) with the goal of developing a framework for assessing in-field yield gap in two physiographic provinces of the Mid-Atlantic region of the US. Chapter 2 is an assessment of the magnitude of existing spatio-temporal yield variations using baseline functions at a whole field scale and standardized principal component analysis at a within field scale. Chapter 3 is a general economic assessment of the justification for PA. Chapter 4 presents prediction models for Yp or Yw for each physiographic province in the US Mid-Atlantic region. These models consisted of statistical and machine learning approaches that used spatially referenced soil property measurements, proximal soil sensing data, and digital elevation models as input. Chapter 5

focuses on assessing the feasibility of VNIRS in predicting soil physical, chemical, and biological properties for soil samples collected in the Northeast and Midwest US. Chapter 6 is an analysis of the changes in measureable soil physical, chemical, and biological properties across soil profiles by long-term tillage and stover management treatments in Chazy, NY. Chapter 7 (developed into a manuscript; Kinoshita et al., 2016), contains results of the assessments of approaches for assessing spatial variation of topsoil soil organic carbon in Costa Rica.

CHAPTER 2: STANDARDIZED PRINCIPAL COMPONENT ANALYSIS AND GEOSTATISTICS FOR SITE-SPECIFIC CROP AND SOIL MANAGEMENT DECISION MAKING

2.1 ABSTRACT

Understanding spatio-temporal variation of grain yields is an important step towards strategizing site- and time-specific soil and crop management. In this study, we utilized multi-year grain yield monitor information of maize and soybean under rainfed and irrigated cropping systems in the Mid-Atlantic US for regional, field, and within-field scale assessments. Regional scale relationship between in-season precipitation and temporal yield variation was assessed using baseline function with critical period rainfall as a predictor. We identified soybean yields to show less sensitivity to moisture-limited conditions and showed lower correlation to in-season precipitation variation in the region. Maize had a higher association with in-season precipitation in the Coastal Plain province compared to the Piedmont. The function fit was better in the Coastal Plain and indicated a relative similarity of each field on the relationship between field-average yield variation and in-season precipitation in the province. Subsequently, standardized principal component analysis (stdPCA) was utilized, which revealed within-field yield pattern to be present only under moisture-limited conditions for maize in the Coastal Plain. In the Piedmont, yield pattern was more consistent under a range of in-season precipitation conditions compared to the Coastal Plain. These indicated the importance of both within-field site- and time-specific management for maize in the Coastal Plain, whereas site-specific management is more important in the Piedmont for both crop types depending on the nature of management.

2.2 KEY WORDS

Baseline function; Maize; Principal component analysis; Soybean; Yield monitor

2.3 INTRODUCTION

Understanding yield potential (Yp) and associated yield constraints are important steps toward improving the efficiency of resource investment in terms of profitability and grain yields while avoiding adverse environmental effects (van Ittersum and Rabbinge, 1997). In the Mid-Atlantic region of the United States (US), significant variations of geological, topographical, and climatic characteristics cause grain yields to vary spatially from year-to-year at various geographical scales (Kaul et al., 2005). From a field management perspective, growers can modify their management practices if the patterns of within-field variation are known and their causes are verified. Therefore, this is a two-part study where we attempt to extract practical information from grain yield monitor data in Part I (Chapter 2), and to identify within-field variable yield constraints and to predict yield patterns in Part II (Chapter 4).

Recent availability of grain yield monitor information has enabled to quantify within-field yield variations (Abuzar et al., 2004; Blackmore, 2000; Blackmore et al., 2003; Florin et al., 2009), and has been extensively utilized to establish management zones (Brock et al., 2005; Cox and Gerard, 2007; Pringle et al., 2003). Although Oliver et al. (2010) emphasized the importance of identifying the nature of yield constraints and their position in the soil profile rather than the

spatial variation of Yp, many of farm fields in the Mid-Atlantic US are leased (≈46%; USDA, 2009) and the farmer knowledge of yield patterns might be limited.

Yield potential is controlled by yield defining factors (van Ittersum et al., 2013) including CO_2 level, incident solar radiation, temperature, and pre-determined cultivar features. Under field conditions especially under rainfed agricultural systems, water is the primary yield constraint limiting crop growth, and therefore the yield potential under this system is defined as water-limited yield potential (Yw). Actual yield (Ya) produced by growers are further influenced by factors including weeds, pests, diseases, as well as pollutants

Therefore, strategies are needed to predict within-field Yp. Within-field Yp may be predicted using approaches including baseline function (Calviño et al., 2003; Calviño and Sadras, 1999), machine learning (Kaul et al., 2005), and Soil-Plant-Atmosphere (SPA) system models (Grassini et al., 2015; Hochman et al., 2012; Oliver et al., 2010) although applications of these for within-field assessment have been limited by the data available to parameterize the models. Within-field assessment requires the input data to have high spatial resolution, and the data acquisition may become extensive and costly. Therefore, we need to first assess the potential benefits and the feasibility of within-field management using yield data collected by a grain harvester yield monitor. Once this is verified, common yield constraints should be identified for a particular region and verify their relationship to the spatio-temporal yield variation for developing empirical models (Kaul et al., 2005), or utilize various sensing equipment to collect and estimate required information for the SPA system models (Adamchuk et al., 2004; Graham, 2012). In this part of the study, we proposed to first assess field-scale spatio-temporal yield variation relative to growing season precipitation, one of the most important yield constraints in the region (Kaul et al., 2005) using baseline function (Calviño et al., 2003; Calviño and Sadras,

1999). Subsequently, within-field spatio-temporal yield variation was assessed using spatially referenced standardized principal component analysis (Eastman and Filk, 1993; stdPCA), which is a method capable of identifying characteristic as well as hidden yield patterns and assess their relationship to growing season specific environmental variables. This has been applied for the spatial assessment of time-series satellite images for vegetation (Eastman and Filk, 1993) or county-average grain yield data (Potgieter et al., 2002) but has not been applied to within-field scale grain yield information. The linear transformation of the original data of this method allows for noise reduction i.e. spatio-temporally sporadic yield pattern or data collection errors, while it attempts to retain the maximum amount of variation in the original data. The assessment of the first principal component (PC) reveals the magnitude of spatial vs. temporal yield variation in the dataset while the remaining PCs also represent particular yield patterns and can be assessed for their correlations to factors such as in-season climate conditions (Eastman and Filk, 1993). In-season climate conditions have been found important and known to control the effects of soil and topographical variables on grain yields (Jiang and Thelen, 2004). Verification of a stdPCA calculated yield pattern to a particular in-season precipitation condition may allow for the identification of potential yield constraints among measurable soil and topographical properties spatially, which will be discussed in Chapter 4.

Therefore, this is the first of a two-part study to develop strategies to assess within-field Yp in the Mid-Atlantic US and to assess existing potential yield constraints. In this chapter, we investigated the spatio-temporal yield variation of maize and soybean at regional, field, and within-field scales using well-calibrated grain yield monitor information across three physiographic provinces in the Mid-Atlantic US, and assessed their interactions with in-season precipitation information. The objectives of this study were to i) apply baseline function to reveal

the regional-scale relationship between temporal yield variation and in-season precipitation, ii) apply stdPCA to reveal within-field scale spatio-temporal grain yield variation and their associations with in-season precipitation, and iii) utilize the gained knowledge for strategizing possible site and time-specific crop and soil management.

2.4 MATERIALS AND METHODS

2.4.1 Site description

The research fields are located in the states of Delaware, Maryland, Virginia, West Virginia, and Pennsylvania, between 75° 33' 51" and 77° 54' 49" W, and between 38° 56' 10" and 39° 50' 23" N, the total area of around 19,630 km^2 (Figure 2.1). There are two climate regions, which are warm temperate climates in the southern part and hot summer continental climates in the northern part (Peel et al., 2007).

Three physiographic provinces (Fenneman, 1938) and five distinctively different areas of soil characteristics according to the National Cooperative Soil Survey exist in the research area (Soil Survey Staff et al., n.d.). The first physiographic province (Area 1) is the Coastal Plain province, and the latter two are the Piedmont and Blue Ridge provinces (Area 2). Area 1 is mainly associated with Typic Hapludults and predominantly formed on coastal plain deposits (sandy loam; Figure 2.1) below varying depths (40-to-100 cm depth) of aeolian silt deposits that are very acidic (Simonson, 1982). Closer to the Chesapeake Bay on the Eastern Shore of Maryland, the surface silt deposits could reach more than 150 cm deep (Foss et al., 1978). Area 2

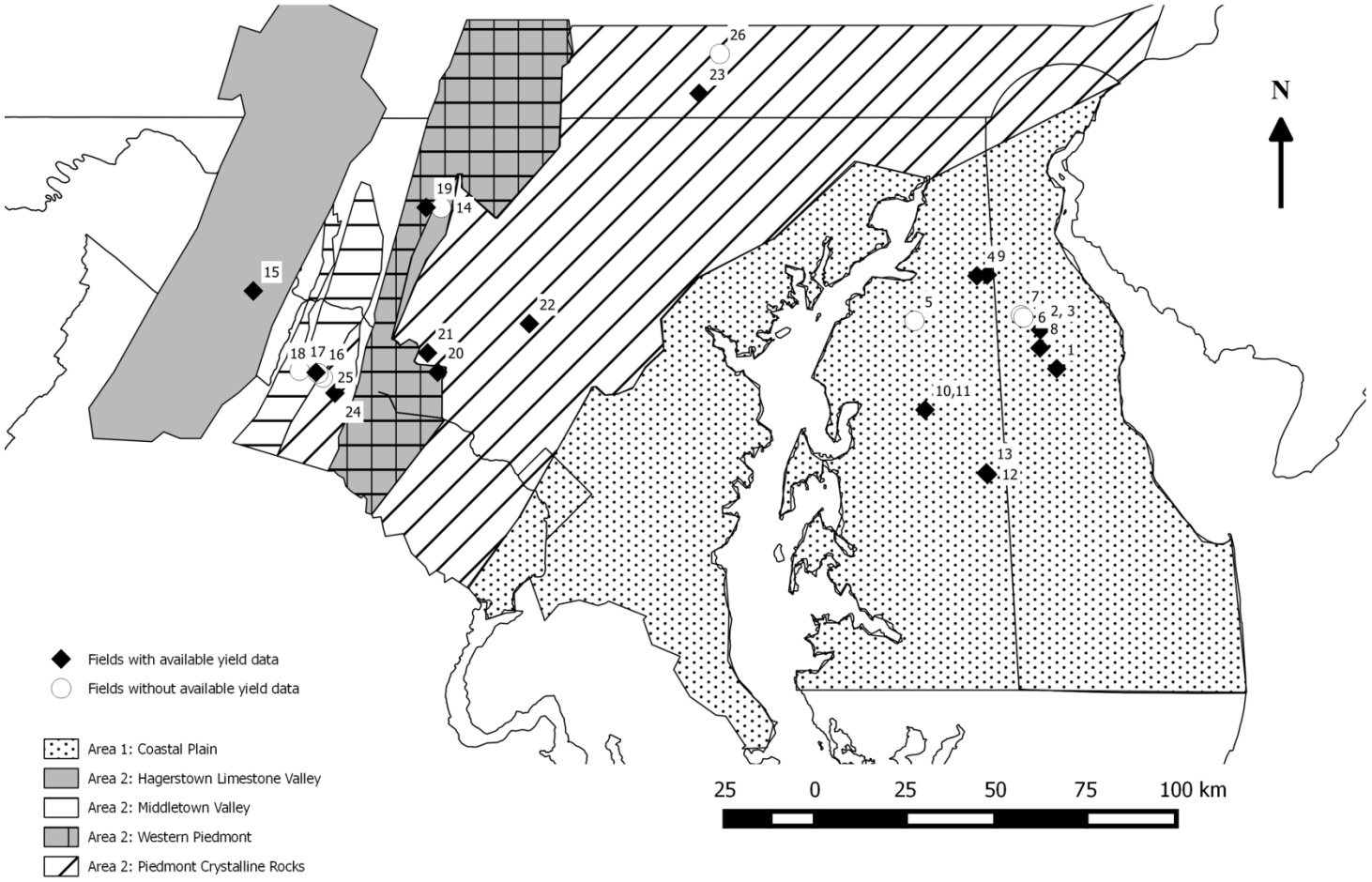

Figure 2.1 Map of the study site with indications of each characteristic soil region and field locations.

had undulating topography with four distinct soil characteristic areas, which are Hagerstown Limestone Valley, Middletown Valley, Western Piedmont, and Piedmont Crystalline Rocks (Figure 2.1). Hagerstown Limestone Valley is mainly associated with Typic Hapludalfs and the soil pH is neutral to slightly acid and has low rock content; Middletown Valley is associated with Typic Hapludults on moderate to strongly acid soil and has high rock contents; Western Piedmont is associated with Ultic Hapludalfs formed on reddish parent materials of Triassic period and is moderately acidic and extremely rocky; and Piedmont Crystalline Rocks is associated with Typic Hapludults and has thin to intermediate depth (< 75 cm) aeolian silt deposits on the surface (Simonson, 1982; Weaver, 1967) and is moderately acidic and rocky.

We selected 18 fields from Area 1 and 2 based on the availability of yield data of more than three growing seasons for maize (*Zea mays*, L.) or soybean (*Glycine max*, L.; Table 2.1). Twelve of the 18 fields were rainfed (rainfed set) and six fields were irrigated (irrigated set), and fertilizer, pest, weed, and water management on each field followed individual farm's management scheme. All of the selected fields had a similar sequence of crop rotation: maize, soybean followed by wheat or barley. In some cases, double-crop soybean followed the harvest of wheat or barley. The earliest planting date recorded for maize in the dataset was the April 12th and the latest was the May 30th. For soybean, the earliest was May 13th, and July 18th was the latest, which was under the double-cropping system. In most fields, multiple varieties were planted.

Table 2.1 Availability of grain yield and soil information for each study field.

Area	Field ID	Dryland Yield Study	Irrigated Yield Study	Deep Soil Samples	Shallow Soil Samples
Area 1	1	*		*	*
	2	*		*	*
	3	*			*
	4	*			*
	5			*	
	6			*	
	7				*
	8		*		*
	9		*		*
	10		*		*
	11		*		*
	12		*		*
	13		*		*
Area 2	14			*	*
	15	*			*
	16			*	*
	17				*
	18	*			*
	19	*		*	*
	20	*			*
	21	*		*	*
	22	*		*	*
	23	*			*
	24	*			*
	25				*
	26				*

2.4.2 Yield data collection and processing

Yield data were collected using well-calibrated yield monitors. Post-processing of the data was done using the Yield Editor 2.0.7 software (Sudduth and Drummond, 2007) for flow delays and slow combine velocity at the beginning and end of each pass. The data were then rasterized (6 m by 6 m) using the SAGA function (Conrad et al., 2015) within the QGIS environment (QGIS Development Team, 2015).

The yield data were collected between 2001 and 2014 and we adjusted for the potential effects of technological improvements over the period (Grassini et al., 2013). We normalized the data for the 2001 level by accounting for annual yield increases of 114 and 27 kg ha^{-1} $year^{-1}$ for maize (Grassini et al., 2013) and soybean (Good and Irwin, 2012), respectively. Also, 18 m of field borders were removed from the analyses where we expected and visually observed unusually low grain yields by factors including reduced incident solar radiation by trees, soil compaction due to higher traffic, yield monitor recording errors associated with slow tractor velocity, and pest damage.

It was also important to estimate the number of growing season yield records required until we have stable estimates of the mean yield levels. We calculated the percent change of the last sequential averaged yield values for each crop (van Ittersum et al., 2013). At present, there are no established thresholds for the sequential average yields, therefore we simply presented the data.

2.4.3 Climate information

High-resolution precipitation and temperature information was acquired from the Northeast Regional Climate Center and the Cornell Center for Advanced Computing for the resolution of 4 by 4 km per grid. The temperature information was generated utilizing the National Oceanic & Atmospheric Administration's (NOAA) Rapid Update Cycle (RUC) weather forecast model as well as data obtained from the Applied Climate Information System (ACIS; Belcher and DeGaetano, 2005). The precipitation information was generated utilizing the NOAA's Doppler radars and data obtained from the ACIS (Ware, 2005; Wilks, 2008). We estimated the growing degree days (GDDs) using the 10 °C base.

$$GDD = \frac{Tmax+Tmin}{2} - 10 \quad (1)$$

Precipitation information was converted to abundant and well-distributed rainfall (AWDR), which incorporates precipitation amount as well as distribution by multiplying the cumulative precipitation of critical periods for each crop type by the Shannon diversity index (SDI; Tremblay et al., 2012).

$$Cumulative\ precipitation\ (PPT) = \sum(Rain) \quad (2)$$

where Rain is the daily rainfall in mm.

$$SDI = \frac{[-\sum pi \ln(pi)]}{\ln(n)} \quad (3)$$

where pi = Rain/PPT and n is the number of days in the period.

$$AWDR = PPT \times SDI \quad (4)$$

We used different critical time periods for calculating the AWDR for each crop type. For maize, Calviño et al. (2003) suggested that precipitation from 30 days before to 20 days after maize flowering to correlated the most to grain yield variation. We assumed the number of GDD

requirement to reach anthesis as 1135 (Neild and Newman, 1990). For soybean, the critical time period for yield determination has been suggested to range from the R3 to R6 stage. For maturity group III, the GDD of 520 is necessary to reach the R1 stage (Casteel, 2011), and an additional 15 days to the R3 stage (Pedersen, 2009). Therefore, the soybean critical period was 35 days while the maize critical period was 50 days. In order to make the AWDR values for maize and soybean to be comparable, we multiplied the soybean AWDR by a factor of 1.42.

2.4.4 Baseline function

We used baseline function (Calviño and Sadras, 1999) to explore the relationship between field-averaged grain yield information and in-season precipitation records for the region using the dryland set. This simple function relies solely on crop-specific critical period precipitation amount, and primarily applied to assess the effects of precipitation (French and Schultz, 1984a), soil characteristics and crop management on grain yields (Calviño et al., 2003) or to identify potential yield constraints (Calviño and Sadras, 1999; French and Schultz, 1984b). Precipitation distribution has not been incorporated into the function even though the importance has been emphasized (Calviño and Sadras, 1999). In this study, we used the AWDR (Tremblay et al., 2012), which incorporates precipitation distribution to improve the function.

All yield data were standardized for each field, therefore we effectively removed inter-field variations in the mean yields.

The relationship between yield and the AWDR was described by the model (Calviño et al., 2003):

$$Y = a[1 - e^{-\frac{(AWDR - AWDR_0)}{b}}] \quad (5)$$

where a = grain yield with no water limitation, $AWDR_0$ = threshold AWDR for grain set, and b = degree of curvature. The first derivative of the model is then:

$$Y' = a \times b^{-1} \times e^{-\frac{(AWDR - AWDR_0)}{b}} \quad (6)$$

A lower $AWDR_0$ value shows the capacity of the soil environment to produce grain at a lower moisture input through precipitation and varied according to the non-linear least-squares estimates. A lower value of curvature can be interpreted as grain yields only responding to a narrow range of precipitation input and the yield level plateaus quickly. Due to the small number of yield records per field, we only built baseline functions per each crop type followed by fitting the function to each area separately for maize.

In-season, more timely management of fertilizer input has been suggested to increase profitability and grain yield while avoiding over-applications beyond crop uptake (Scharf et al., 2011). Hong et al. (2007) measured residual soil nitrate to be lower after maize harvest when fertilizer rate is decided through a yield-estimate based function. Therefore, we altered the critical periods used for the calculation of the AWDR to test whether the estimation of grain yield is possible. We used May and May through June cumulative precipitation, which are before and during the usual time-windows for side-dress N application for maize.

2.4.5 Standardized principal component analysis

We utilized stdPCA (Eastman and Filk, 1993; Machado-Machado et al., 2011), a feature-space multivariate statistical method to decompose the spatio-temporal variation of within-field multi-year grain yield monitor information to explore the opportunities for site-specific soil and crop management. The assessment was carried out for the dryland set using the

'princomp" module (R Core Team, 2014) in the R statistical computing environment. The dataset was formatted for each growing season to be represented as variables and each 6 m by 6 m cell location to be represented as samples for the stdPCA analysis. It is an orthogonal linear transformation of the original dataset to explain as much variance of the original data in the first principal component (PC) and subsequent PCs with less original variance explained (Varmuza and Filzmoser, 2009). The variables of the original dataset were standardized to avoid a variable with higher absolute variance to dominate but tried to capture existing yield patterns of each growing season. The correlation between the original variables and each PC was denoted as a loading and it is analogous to Pearson's correlation coefficient. Eastman and Filk (1993) undertook visual assessments of the variation of loadings and identified the trend to match particular environmental events that coincided temporally for each PC. This helped to explain the causes of the spatial pattern of the PC values (i.e. PC scores).

2.4.6 Geostatistical assessment of the PC score maps

Standardized PCA is a feature-space method and it is independent of any spatial dependencies of the original samples, and therefore the calculated PC scores do not necessarily present spatial structures. From a field management perspective, opportunities for site-specific management increases when PC score maps have both a large magnitude of variation and a high spatial structure (Pringle et al., 2003). In order to assess the strength of spatial structures present, we adopted the index *S* proposed by Pringle et al. (2003) where they calculated the magnitude of spatial structure by first fitting a quartic trend surface:

$$Y(X,Y) = (Int. + X + Y + X^2 + Y^2 + XY + X^3 + Y^3 + X^2Y + XY^2 + X^4 + Y^4 + X^3Y + X^2Y^2 + XY^3) + \varepsilon \quad (7)$$

where X is Easting coordinates, Y is Northing coordinates, Y(X,Y) is yield as a function of the coordinates, Int. is the intercept of the regression, and ε is the residual term. We then calculated the 'integral scale' (Ja; Russo and Bresler, 1981) of the trend-surface residuals, which shows the range of influence in 2-dimensional space (Warrick et al., 1986) based on a semi-variogram. We undertook a least-squares fit of spherical model to the semi-variogram using the *gstat* package (Pebesma, 2004) in the R statistical computing environment and the *Ja* was calculated as follows:

$$J_a \approx \frac{\left\{2\int_0^{\infty}(1-\frac{\gamma(h)}{(Sill)}\right\}hdh}{10000} \quad (8)$$

where γ(h) is the modeled semi-variogram, Sill is the total sill of the residuals, and the denominator converts square meters to hectares.

Finally, the index of spatial structure *S* was calculated as:

$$S = P_t A + (1 - P_t)J_a \quad (9)$$

where *Pt* is the percent total variance explained by a fitted trend-surface of the yield information, *A* is the area of the field in ha, and *Ja* is the calculated integral scale. This index represents the largest average area of autocorrelation (Pringle et al., 2003).

2.5 RESULTS AND DISCUSSION

2.5.1 Grain yields and climate

The number of years of yield data acquired from each field was different and ranged from three to seven per crop type (Table 2.2). Sequential average yield revealed the percent change from 1.95 to 22.5 % for rainfed maize, 0.881 to 4.41 % for irrigated maize, and from 2.72 to 12.0 % for soybean (Table 2.2). We identified contrasting mean yields and the CVs in each area and also for water management. For maize, the mean yield levels were generally higher for the irrigated set compared to the rainfed set. Among the rainfed set, the temporal CV was substantially higher in Area 1 (mean CV = 45.8 %) compared to Area 2 (mean CV = 18.0 %) and the precipitation variation was also higher in Area 1 (mean CV = 54.9 %) than Area 2 (mean CV = 31.1 %). Irrigation reduced the temporal CV substantially (mean CV = 12.0 %; Table 2.2), which is in line with the past findings of irrigation decoupling temporal yield variation from climate variation (Troy et al., 2015). For soybean, Field 19 had the lowest AWDR (Table 2.3) due to a late planting date in mid-July.

2.5.2 Baseline function

We assessed the regional-scale relationship between standardized yields and AWDR and determined whether regional or physiographic province-specific functions can be built. Baseline function only relies on precipitation related variables where we used the AWDR parameter to incorporate the distribution of precipitation (Tremblay et al., 2012). Using the entire data set (i.e. global models), we found a better model fit for maize (R^2 = 0.46; Figure 2.2a) compared to soybean (R^2 = 0.31; Figure 2.2d). Higher variation in the planting date for soybean and late planting under double cropping systems could cause lower total incident solar radiation,

lower temperature as well as lower available soil moisture, and therefore lower yields as a result (Egli and Bruening, 2000). Although the model fit was relatively low for both crop types, the $AWDR_0$ value was substantially lower for soybean (0.078 vs. 30; Table 2.4) indicating a lower critical period precipitation requirement compared to maize. Soybean is known to have higher tolerance to water deficiency compared to maize (Sadras and Calviño, 2001), and maize is known to be extremely susceptible to environmental stress around anthesis (Tollenaar and Dwyer, 1999). Soybean can tolerate moisture stress partly because of its ability to forage resources more precisely (Grime, 1998), a beneficial characteristic when it is grown in shallow soil (Sadras and Calviño, 2001).

The maize dataset was separated to each area, and the baseline function fit was tested. We found a substantially better fit for Area 1 (R^2 = 0.78; Figure 2.2b) compared to Area 2 (R^2 = 0.14; Figure 2.2c) and the Area 1 also showed a larger b coefficient (39 vs. 13; Table 2.4) indicating a larger response of the standardized yield to added precipitation. Although both the maximum and the minimum AWDR values were found in Area 1, there was a larger increase in the standardized yield level between the AWDR values of 50 and 150 compared to Area 2 (Figure 2.2b and 2.2c). This indicates the dominance of precipitation-related yield constraints in Area 1. Also, relatively similar relationships between precipitation and yield levels among individual fields were found in this area. In Area 2, the effects of precipitation-related yield constraints were shown less important at a whole-field scale and the yield levels appeared to be controlled by field-specific constraints observed by a larger scatter of recorded yield levels around the baseline function.

Table 2.2 Summary statistics of the recorded yield data with both temporal and spatial variations.

Maize															
							Temporal Variation					Spatial Variation			
Field ID	Area	Soil Order	Irrigation	Area	N †	Percent change	Min	Max	Mean	SD	CV	Min CV	Max CV	Mean CV	
				ha		%	Mg ha^{-1}				%	%			
1	1	Ultisols	No	17.2	5	22.5	3.80	13.3	7.70	4.30	55.4	9.70	45.8	27.8	
2	1	Ultisols	No	12.0	5	9.15	4.20	11.3	8.10	3.00	37.6	9.80	46.8	29.5	
3	1	Ultisols	No	12.9	5	13.1	3.30	11.6	7.80	3.60	46.2	5.90	51.7	28.3	
4	1	Ultisols	No	38.6	4	6.24	5.70	13.7	8.30	3.60	44.0	14.8	56.8	36.9	
8	1	Ultisols	Yes	23.7	4	4.41	9.38	13.7	11.5	2.10	18.4	6.08	49.1	19.3	
9	1	Ultisols	Yes	14.2	4	4.02	11.6	14.7	13.3	1.30	9.49	10.4	29.5	16.6	
10	1	Ultisols	Yes	14.1	4	4.33	10.5	13.9	12.1	1.50	12.4	15.8	33.2	22.7	
11	1	Ultisols	Yes	47.2	3	1.93	10.9	12.2	11.3	0.800	6.75	25.9	29.4	27.4	
12	1	Ultisols	Yes	23.9	6	0.881	8.78	13.4	10.8	1.50	14.1	13.4	63.1	29.9	
13	1	Ultisols	Yes	16.4	4	1.26	9.18	11.7	11.0	1.20	10.9	10.0	37.2	22.6	
15	2	Alfisols	No	25.3	1	na	na	na	na	na	na	na	na	na	
18	2	Ultisols/ Alfisols	No	13.1	1	na	na	na	na	na	na	na	na	na	
19	2	Alfisols/ Inceptisols	No	11.8	1	na	na	na	na	na	na	na	na	na	
20	2	Ultisols/ Alfisols	No	7.63	5	2.62	5.80	10.3	8.30	2.10	25.3	17.4	41.4	27.6	
21	2	Ultisols	No	17.3	5	5.63	10.4	13.8	11.4	1.40	12.7	11.6	25.0	16.3	
22	2	Ultisols	No	14.6	4	3.99	6.90	10.3	8.70	1.70	19.9	16.6	36.8	25.7	
23	2	Ultisols/ Inceptisols	No	7.73	4	1.95	8.70	10.9	9.80	1.00	10.0	12.1	17.4	15.4	
24	2	Ultisols/ Alfisols	No	6.07	7	3.11	5.10	10.0	7.80	1.70	21.9	19.5	65.8	38.9	

† N = number of years of available yield data

Table 2.2 (Continued)

							Soybean							
							Temporal Variation					Spatial Variation		
Field ID	Area	Soil Order	Irrigation	Area	N	Percent change	Min	Max	Mean	SD	CV	Min CV	Max CV	Mean CV
				ha		%	Mg ha^{-1}				%	%		
1	1	Ultisols	No	17.2	3	7.61	2.90	3.90	3.40	0.500	14.9	9.40	12.1	10.7
2	1	Ultisols	No	12.0	1	na	na	na	na	na	na	na	na	na
3	1	Ultisols	No	12.9	1	na	na	na	na	na	na	na	na	na
4	1	Ultisols	No	38.6	2	na	na	na	na	na	na	na	na	na
8	1	Ultisols	Yes	23.7	2	na	na	na	na	na	na	na	na	na
9	1	Ultisols	Yes	14.2	1	na	na	na	na	na	na	na	na	na
10	1	Ultisols	Yes	14.1	1	na	na	na	na	na	na	na	na	na[1]
11	1	Ultisols	Yes	47.2	0	na	na	na	na	na	na	na	na	na
12	1	Ultisols	Yes	23.9	0	na	na	na	na	na	na	na	na	na
13	1	Ultisols	Yes	16.4	1	na	na	na	na	na	na	na	na	na
15	2	Alfisols	No	25.3	7	2.85	2.30	4.30	3.00	0.600	21.3	14.6	40.8	20.5
18	2	Ultisols/ Alfisols	No	13.1	3	12.0	2.80	5.40	3.90	1.30	33.5	15.1	31.8	20.8
19	2	Alfisols/ Inceptisols	No	11.8	4	2.72	1.40	4.50	2.50	1.40	57.1	14.6	52.5	36.0
20	2	Ultisols/ Alfisols	No	7.63	5	10.1	1.10	3.10	2.20	0.800	37.4	18.7	38.5	27.8
21	2	Ultisols	No	17.3	3	4.24	3.30	4.80	3.90	0.800	21.1	11.9	18.2	14.6
22	2	Ultisols	No	14.6	2	na	na	na	na	na	na	na	na	na
23	2	Ultisols/ Inceptisols	No	7.73	2	na	na	na	na	na	na	na	na	na
24	2	Ultisols/ Alfisols	No	6.07	5	3.22	2.30	3.50	3.10	0.600	18.9	19.0	28.5	24.2

† N = number of years of available yield data

Table 2.3 Summary statistics of the abundant and well-distributed rainfall (AWDR) for each rainfed field.

	Maize						Soybean					
Field ID	N†	Min	Max	Mean	SD	CV	N†	Min	Max	Mean	SD	CV
1	5	41.8	213.6	104.8	70.9	67.7	3	69.8	197.9	143.7	66.3	46.1
2	5	45.4	162.9	82.7	46.2	55.9	1	na	na	na	na	na
3	5	45.4	162.9	82.7	46.2	55.9	1	na	na	na	na	na
4	4	57.6	126.1	79.1	31.7	40.1	2	na	na	na	na	na
15	1	na	na	na	na	na	7	37.9	137.8	73.7	38.8	52.6
18	1	na	na	na	na	na	3	48.5	108.1	69.8	33.2	47.6
19	1	na	na	na	na	na	4	19.9	125.2	82.2	45.1	54.9
20	5	70.4	173.9	108.0	40.9	37.9	5	25.7	130.4	69.3	39.2	56.6
21	5	55.9	104.3	75.6	18.4	24.3	3	41.9	143.2	87.6	51.4	58.7
22	4	66.7	125.9	90.9	25.0	27.5	2	na	na	na	na	na
23	4	68.2	137.2	103.9	28.8	27.7	2	na	na	na	na	na
24	7	56.7	170.7	103.6	39.6	38.2	5	31.6	154.9	77.0	47.4	61.6

† N = number of years of available yield data

We also assessed whether within-field spatial CVs are related to AWDR. For maize in Area 1, the spatial CV had a negative correlation (ρ = -0.71; Figure 2.3a) to AWDR where AWDR of higher than 100 showed the spatial CV of less than 20 %. The relationships in Area 2 were not clear (ρ = -0.03; Figure 2.3b) and showed higher spatial CVs compared to Area 1 for high precipitation conditions. These results indicate that within-field yield patterns exist even in non-moisture limited growing seasons in Area 2, while yield patterns in Area 1 are largely absent in non-moisture limited growing seasons. Area 2 is known to have complex topography due to the mixture of parent materials with a varying degree of resistance to erosion (Hack, 1980), which could affect crop growth directly by affecting soil moisture distribution and indirectly on soil nutrient availability and soil physical properties including soil organic matter, soil temperature, and soil texture (Bennett et al., 1972; Franzmeier et al., 1969; Jiang and Thelen, 2004; Stone et al., 1985). In addition, infiltration rates are generally lower in Area 2 (6-15 cm h^{-1}; Markewich et al., 1990) compared to Area 1 (13-28 cm h^{-1}) that could cause relatively high spatial CVs to be present under non-moisture limited or moisture-excess growing seasons.

For soybean, there was a moderate negative relationship (ρ = -0.45; Figure 2.3c) between the AWDR and spatial CVs, with lower spatial CVs compared to maize in Area 2 (Figure 2.3b and 2.3c) under high precipitation conditions. Due to the shorter critical period for soybean, we normalized the AWDR values to the maize level, which artificially inflated the AWDR values for soybean, and therefore similar levels of AWDR could have caused more severe moisture-excess for maize.

From a field management perspective, it is important to assess whether yields can be predicted using early season precipitation information since crop inputs such as N application is often based on a yield goal (Stanford, 1973). Therefore, we assessed the predictability of field-

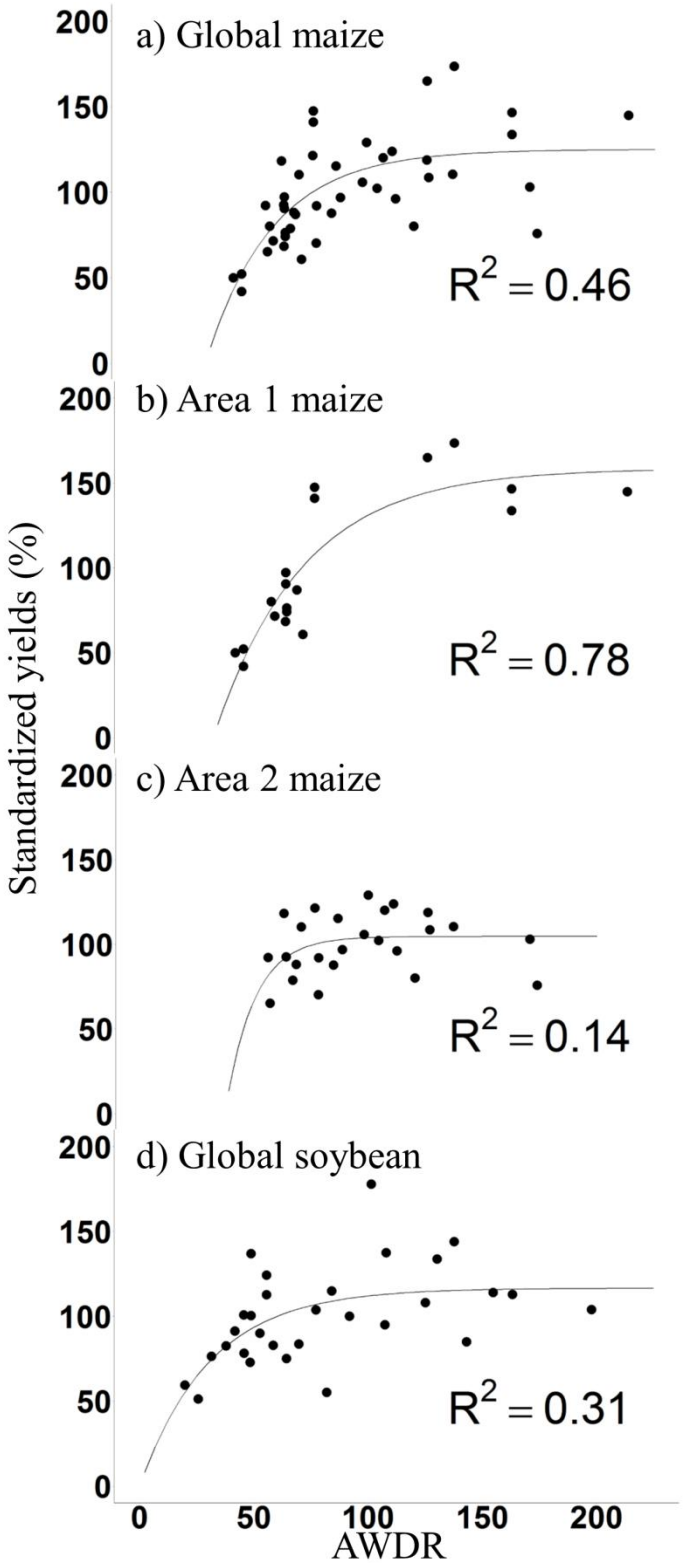

Figure 2.2 Baseline functions of standardized grain yields against abundant and well-distributed rainfall (AWDR) for a) maize in all fields, b) maize in Area 1, c) maize in Area 2, and d) soybean in all fields.

Table 2.4 Coefficients for the baseline functions for standardized yield of maize and soybeans in each Area.

	Maize			Soybeans		
	a†	$AWDR_0$	b	a	$AWDR_0$	b
Global	125	30.0	29.0	117	0.078	31.0
Area 1	159	32.0	39.0	na	na	na
Area 2	105	37.0	13.0	na	na	na

† a = grain yield with no water limitation; $AWDR_0$ = threshold AWDR for grain set; b = degree of curvature

averaged yields using early-season precipitation for maize in Area 1 where the baseline function fit was the highest (Figure 2.2b). Results show that baseline function using the AWDR of May has no predictability (R^2 = -0.009; Figure 2.4a) and the predictability with the May-June AWDR is modest (R^2 = 0.38; Figure 2.4b) compared to the full critical period AWDR (R^2 = 0.78; Figure 2.2b). These results indicate traditional mass-balance approach insufficient to account for the temporal variation of crop inputs requirement even incorporating in-season precipitation in this area. In Area 2, standardized yields varied less compared to Area 1 (Figure 2.2b and c), and therefore mass-balance approach may be possible to estimate the crop N requirement without varying the rate of application temporally. For individual growing season, the relationship between yield and economically optimum N rate has been found weak (Katsvairo et al., 2003; Lory and Scharf, 2003; Sawyer et al., 2006; Vanotti and Bundy, 1994) and the soil N supply and crop N uptake efficiency appear more important (Scharf et al., 2006). Nevertheless, crop N demand becomes important for the long-term and especially applicable where yield patterns are more consistent than the patterns of N supply, loss, and uptake efficiency (Scharf et al., 2006; Vanotti and Bundy, 1994). This appears challenging in Area 1 where yield levels are not consistent year-to-year. Melkonian et al. (2008) suggested the use of models that simulate processes including N mineralization, N loss and crop N uptake using input data of crop and soil information as well as in-season climate information. About 70 to 85 % of N is taken up before anthesis for maize (Chen et al., 2011), therefore the estimation of the availability of soil N is important before and during the early part of the critical period.

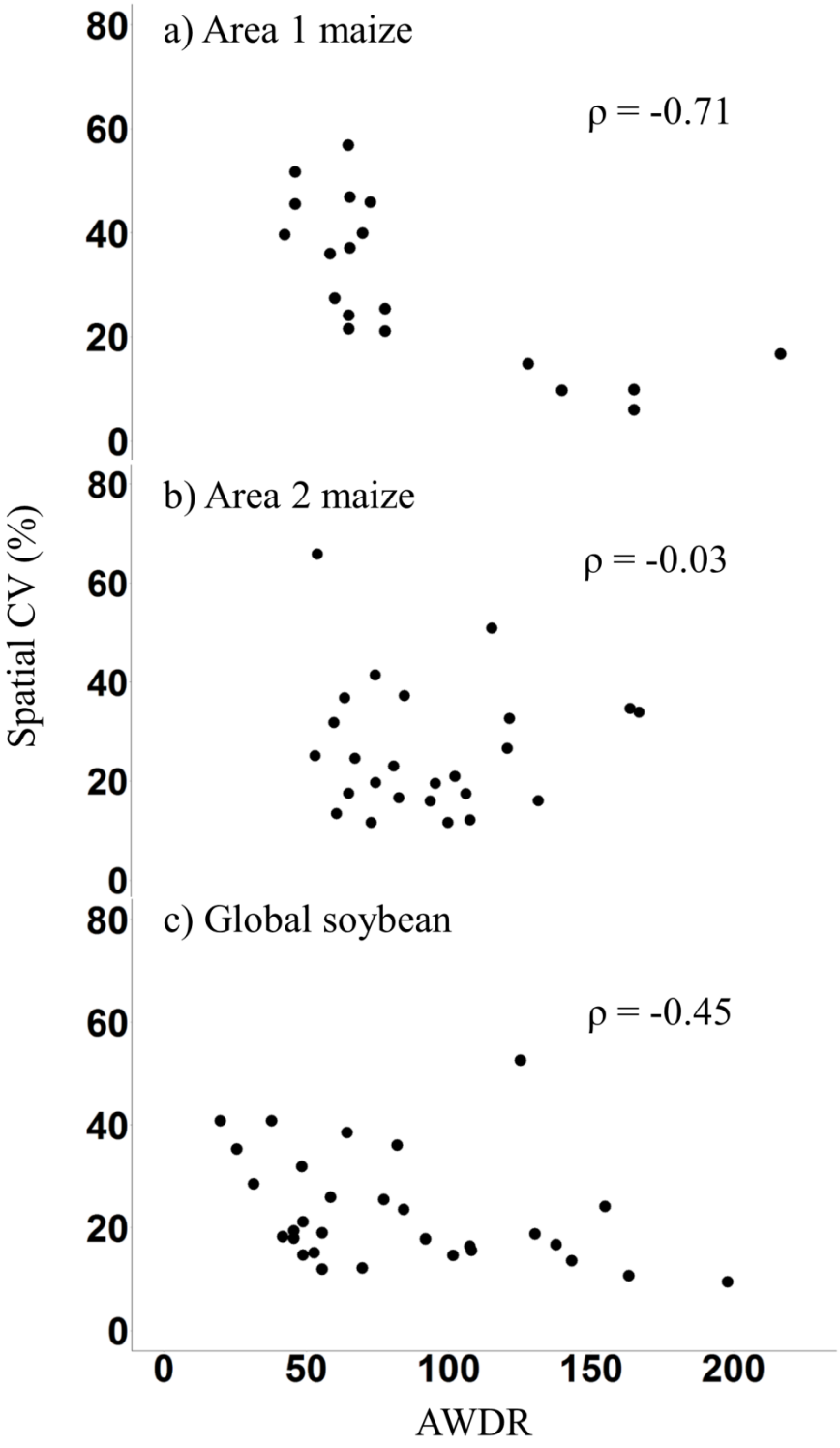

Figure 2.3 Spearman rank correlation between the abundant and well-distributed rainfall (AWDR) and the spatial coefficient of variation (CV) of field averaged yields for a) maize in Area 1, b) maize in Area 2, and c) soybean in all fields.

2.5.3 Standardized principal component analysis

Standardized principal component analysis was shown as a powerful tool to analyze spatio-temporal within-field yield pattern data acquired from a grain yield monitor in this study. There were two main outputs from this analysis; 1) loadings are the principal axes scaled by the square roots of the respective eigenvalues and show the correlation to the spatial structure of a particular variable and 2) scores are newly assigned transformed values in the PC space for each sample and thus they are spatially-referenced. In summary, the combination of correlation plots between loadings and AWDR, score maps, and the analysis of the spatial structure of the score maps allowed to 1) assess the magnitude of spatial yield variation relative to temporal yield variation, 2) visualize characteristic and latent yield pattern, and 3) assess whether the yield pattern contains autocorrleated manageable yields or nugget effects.

Maize: For the rainfed set, first PC (PC1) explained 43.0 to 66.0 % of the total variation, whereas 30.0 to 50.5 % of the variation was explained for the irrigated set (Table 2.5). The importance of assessing subsequent PCs depend on the variance explained by PC1 (Table 2.5) and the number of PCs to retain for further analysis can be assessed by methods including a visual assessment of a scree plot or utilizing one of the stopping rules (e.g. Kaiser rule; Kaiser, 1960). In some cases, such as Field 20, 23, and 24, PC1 explained a high percentage of variance (Table 2.5) and the necessity for further assessment was minimal. For some of the irrigated set, the variance explained by each PC was relatively similar showing that no consistent yield pattern was present (i.e. Field 8, 9, and 10). First, we assessed the correlation plots between the loadings from first four PCs and the AWDR to assess whether precipitation has an impact on the spatial yield pattern. The variation of the loadings along AWDR was lower in Area 2 compared to Area 1 (Figure 2.5) indicating that the PC1 score maps (i.e. spatial yield pattern) are more temporally

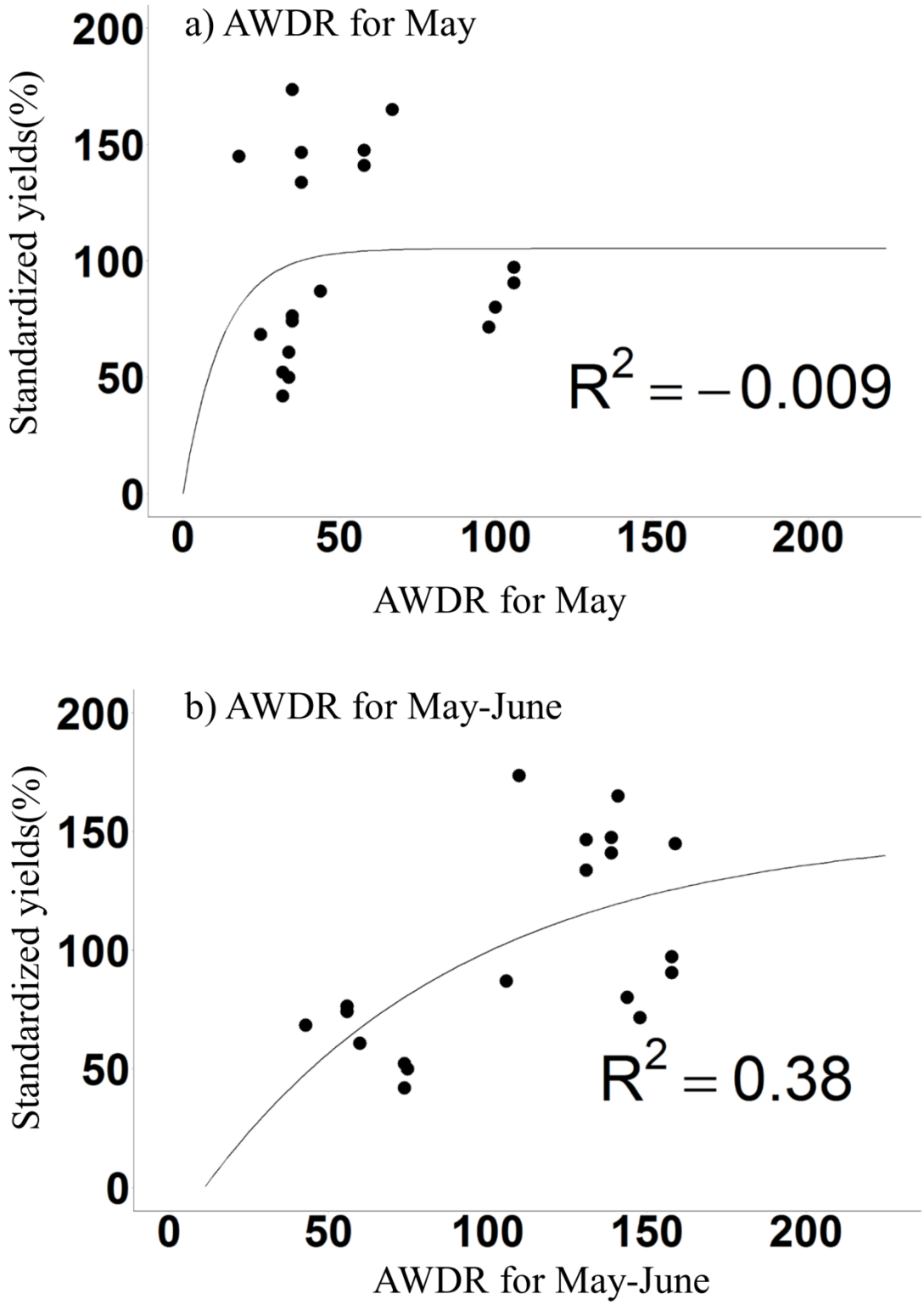

Figure 2.4 In-season prediction of maize yields in Area 1 using a) the AWDR information for May and b) the AWDR information of May and June.

stable in Area 2. Furthermore, this also shows that spatial yield variation dominates over temporal yield variation in this area. In addition, large loadings were only present in low AWDR (i.e. high moisture stress) values in Area 1 (Figure 2.5) indicating the PC1 score maps to represent moisture-limited conditions. In Area 2, positive correlations between the PC1 loadings and AWDR were identified for Fields 20, 21, and 24 (Figure 2.5). In Field 1, a slug outbreak could be the cause of a higher negative loading for the highest AWDR, which was identified through a discussion with a farmer. PC2 in Area 1 showed higher loadings for non-moisture limited conditions (AWDR > 75) and therefore the PC2 score maps represent non-moisture limited yield pattern. Nevertheless, the spatial CV was smaller than 25 % in these non-moisture limited conditions (Figure 2.3a), and therefore the benefit of spatial management based on these maps are low. In Area 2, the PC2 loadings showed a higher variation compared to PC1 and had higher loadings under non-moisture limited conditions (Figure 2.5). Here, the spatial CV was not necessarily lower for PC2 (Figure 2.3b), and it may need to be retained. However, some of the fields showed very small variance explained by PC2 and limits the utility (Table 2.5). The third and the fourth PCs showed small amount of variance explained apart from the fields without consistent spatial pattern (Field 8, 9, and 10; Table 2.5). Therefore, we decided to omit these PCs for further analysis. There were some exceptions to this in Field 3, 8, 9, 20, and 21 although PC1 always explained the most variance (Table 2.5; Varmuza and Filzmoser, 2009). We have shown that the PC1 score maps represent moisture-limited yield pattern and PC2 represent non-moisture limited yield pattern in Area 1 for maize, and visualized for Field 3 (Figure 2.6). Here, PC2 showed low spatial CV explained even though the largest area of autocorrelation was larger than PC1 (Table 2.6; Figure 2.6). The irrigated set of Field 8 and 9 did not have a consistent spatial pattern for management. PC1 captured most of the variance for Field 20 and the loadings were

Table 2.5 Percent variance explained by the first four components in each crop type.

		Maize					Soybean				
Field ID	Area	PC 1†	PC 2	PC 3	PC 4	Cumulative percentage in first 2 components	PC 1	PC 2	PC 3	PC 4	Cumulative percentage in first 2 components
1	1	43.0	21.9	15.9	12.6	64.9	41.7	34.5	23.9	na	76.2
2	1	49.4	21.1	18.2	6.00	70.5	na	na	na	na	na
3	1	50.0	21.0	17.4	7.10	71	na	na	na	na	na
4	1	66.0	21.9	7.2	4.90	87.9	na	na	na	na	na
8	1	34.0	28.0	22.0	16.0	62.0	na	na	na	na	na
9	1	30.0	24.7	23.5	21.7	54.7	na	na	na	na	na
10	1	35.5	25.7	20.8	18.1	61.2	na	na	na	na	na
11	1	49.4	30.8	19.8	na	80.2	na	na	na	na	na
12	1	44.4	19.8	15.8	10.3	64.2	na	na	na	na	na
13	1	50.5	22.0	19.6	7.87	72.5	na	na	na	na	na
15	2	na	na	na	na	na	27.4	15.3	12.8	12.5	42.7
18	3	na	na	na	na	na	54.4	25.5	20.1	na	79.9
19	4	na	na	na	na	na	34.7	27.7	20.1	17.5	62.4
20	4	58.1	14.3	11.7	10.1	72.4	55.6	17.6	9.72	9.42	73.2
21	5	48.5	17.1	15.6	11.2	65.6	49.6	29.7	20.7	na	79.3
22	5	53.5	21.4	17.4	7.60	74.9	na	na	na	na	na
23	5	62.0	18.7	11.2	8.10	80.7	na	na	na	na	na
24	5	65.7	10.1	7.9	6.10	75.8	54.7	14.5	12.4	9.54	69.2

† PC = principal component

consistent throughout different AWDR levels. Field 21 along with Field 23 had the smallest mean spatial CVs (Table 2.2) and also the smallest *S* (Table 2.6) and present small opportunity for spatial management.

Soybean: There were more fields with inconsistent spatial patterns as observed from low variance explained in PC1 (Field 1, 15, 19, and 21; Table 2.5) and higher variation in the loadings (Figure 2.7) but without an obvious trend with the AWDR. There were low spatial CVs present for Field 1 and 21 and the benefits of spatial management may be minimal. Conversely, Field 18, 20 and 24 showed a higher variance explained by PC1 and had relatively high mean spatial CV (Table 2.2). These fields also had consistent yield patterns across years (Figure 2.7). We assessed subsequent PCs for Field 15 and 19 where there were substantial spatial CVs (Table 2.2) but showed inconsistent spatial structures. In Field 15, PC2 only explained a small amount of total variance (Table 2.5) and the spatio-temporal yield structure could contain effects that are spatially random. In Field 19, PC2 contained a comparable amount of total variance (27.7 %; Table 2.5) compare to PC1, and the loadings changed the signs along a range of AWDR values (Figure 2.7) indicating "flip-flop" phenomenon (Florin et al., 2009).

2.5.4 Management perspectives

The primary objective of this study was to assess within-field Yp and determine the opportunity for possible within-field management. There are two possible precision management options exist; spatial and temporal, and the decision needs to be made whether either or both of these options are beneficial for individual field.

Maize and soybean showed contrasting behavior to precipitation variation under rainfed cropping systems using baseline function, and we confirmed relatively higher susceptibility of

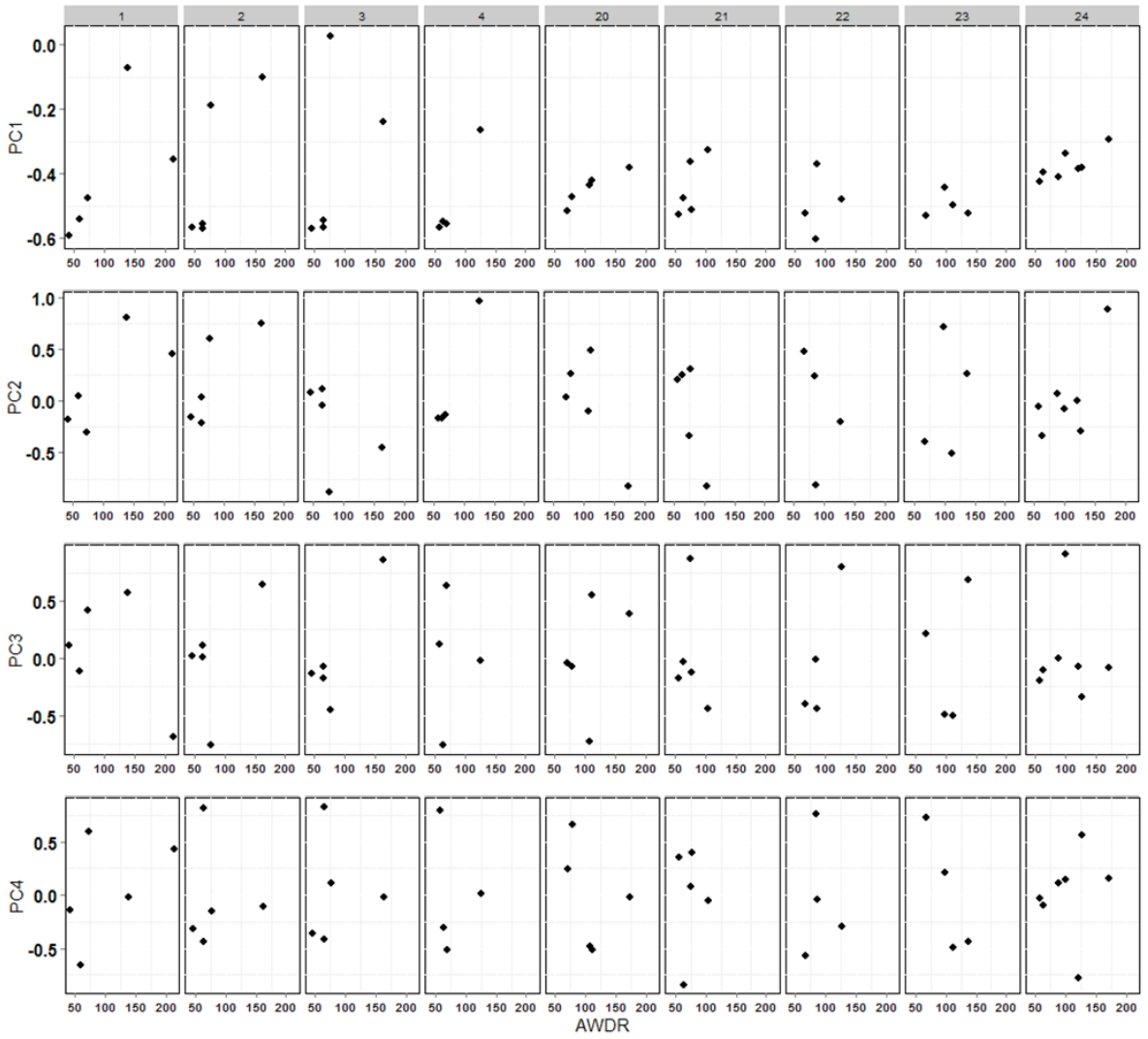

Figure 2.5 Correlation plots between the first four principal component (PC) loadings and the abundant and well-distributed rainfall (AWDR) for maize.

maize to moisture-limited conditions (Figure 2.2a and 2d; Sadras and Calviño, 2001). This was even more significant in Area 1 (Coastal Plain province; Figure 2.2b) than Area 2 (Piedmont province; Figure 2.2c) suggesting a higher need for temporal crop management in Area 1. In the area, high spatial CV was only present under moisture-limited conditions, therefore spatial management should be based on the PC1 score maps. Under non-moisture limited conditions, spatial CV was minimal and thus uniform management is adequate. Yield prediction was not feasible at the time of side-dress N application (Figure 2.4a and 4b) for maize, one of the most crucial in-season management. Therefore, spatial management is useful for pre-plant decision making and may become important for in-season management if underlying soil and topographical properties are significantly different to each other (Whelan and McBratney, 2003) to affect the outputs of decision making simulation models (Melkonian et al., 2008). The PC2 score maps can be useful in Area 1 to identify consistently low-yielding areas but these tend to be along the edge of the field and the areas were not included in this analysis. In Area 2, temporal yield variation was less important (Table 2.2) and consistent yield patterns were observed for each field through time (Figure 2.5). For management, this allows for spatial management using the PC1 score maps with higher confidence for the entire field. Areas with consistently low yielding areas can be identified, and profitability assessment is necessary if these areas are justified to be removed from production (Chapter 3). There were some fields with relatively small *S* values (Field 21 and 23; Table 2.6) and these fields had overall small spatial CV (Table 2.2). The decision needs to be made if spatial management in areas as small as 0.3 ha (≈ 0.8 acre; Table 2.6) is justifiable. Overall, we recommend pre-plant spatial management in both Area 1 and Area 2 based on the PC1 score maps. Temporal management of in-season fertilizer

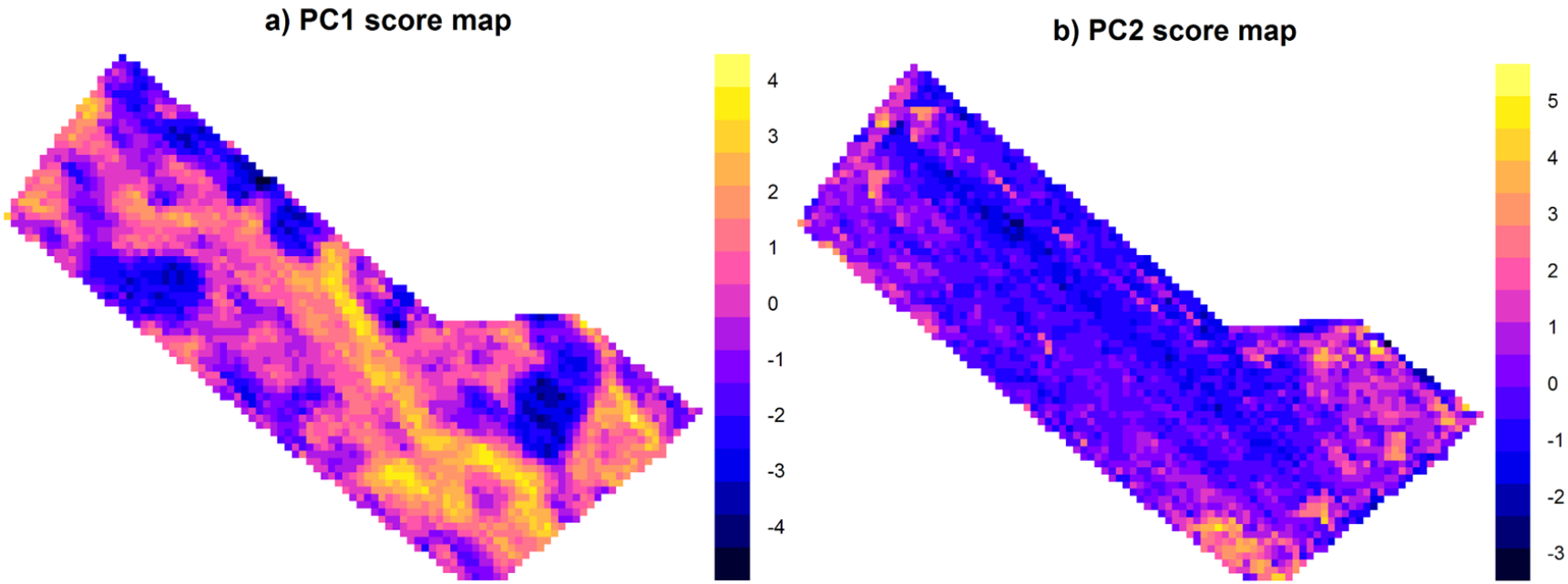

Figure 2.6 Spatially referenced information of a) the PC1 scores and b) the PC2 scores.

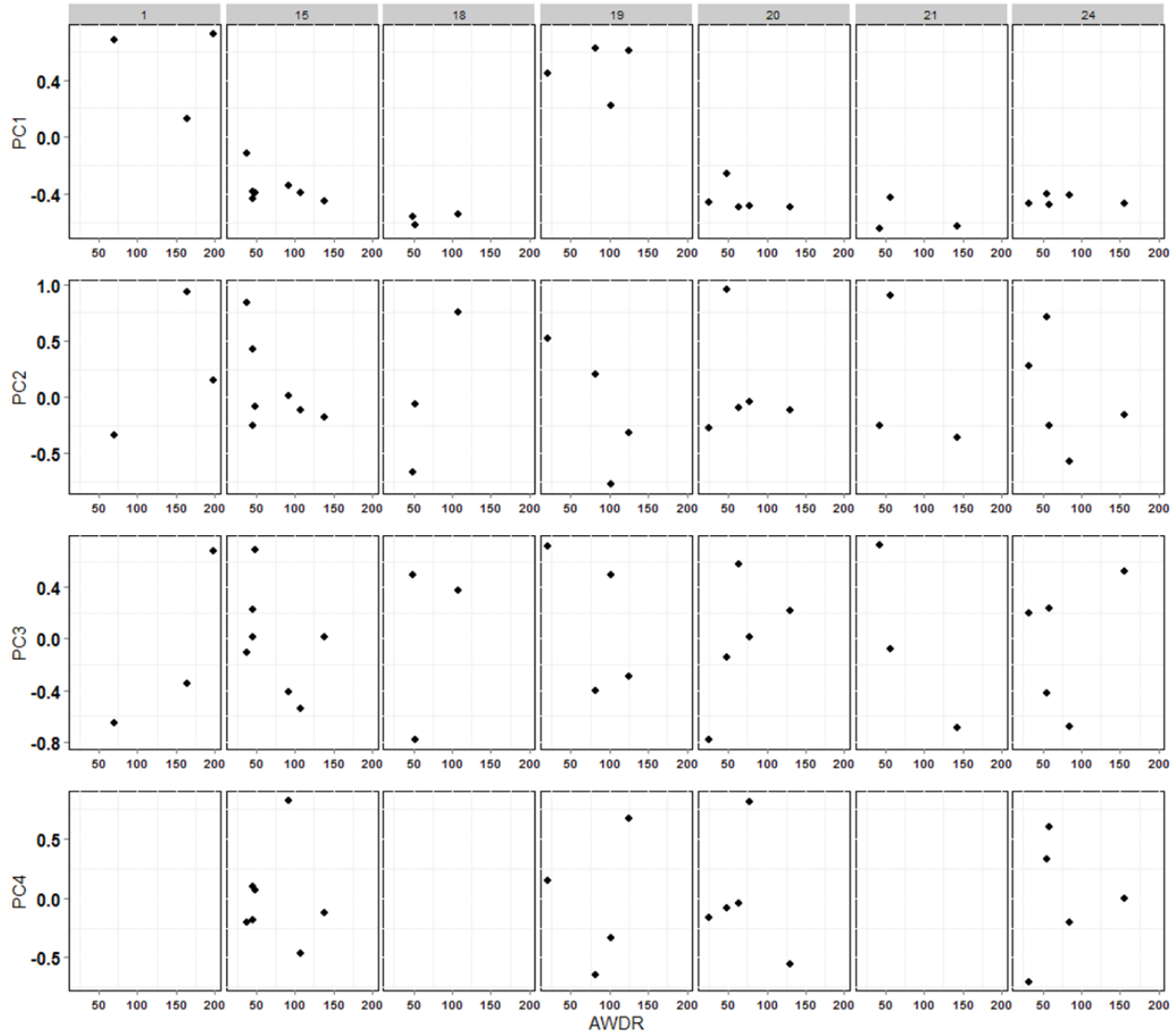

Figure 2.7 Correlation plots between the first four principal component (PC) loadings and the abundant and well-distributed rainfall (AWDR) for soybean.

application is more important in Area 1 and the possibility of spatial management for this depends on the feasibility of input data acquisition at the spatial resolution.

Under irrigated cropping systems, two scenarios were possible with some fields expressing no consistent yield patters, whereas other fields showed PC1 to account for a large amount of the yield variation. Further assessment on soil and topography are necessary to discern which field to follow one of the two scenarios.

For soybean, both spatial and temporal yield variation was high but it did not correlate with critical period precipitation. The presence of double-cropping for soybean may be an important parameter and the planting date may have a high explanatory power for the yields. Kane et al. (1997) noted for maturity group III cultivars planted in June can have significantly diminished yields. In this study, the planting dates were as late as July 18th and this could have been combined with moisture-limited conditions during the critical period with further yield penalties (Kaul et al., 2005). Nevertheless, spatial management can be feasible for several fields in Area 2 for soybean including Field 18, 20, and 24. In other fields, inconsistent yield patterns were observed (Field 1 and 21; Table 2.5) where the magnitude of spatial CV was small (Table 2.2). Therefore, a strategy to predict the magnitude of spatial CV is necessary such as the Random Forest model to predict the spatial variation of PC1 (Chapter 4). For Field 15, inconsistent and unexplainable yield variation caused low total variance explained in PC1, whereas both PC1 and PC2 appeared important for Field 19. Further discussions are necessary with farmers to identify the cause of the spatial variation before utilizing these results for management in these fields (Calviño et al., 2003; Oliver et al., 2010).

Table 2.6 Integral scale of the trend surface residual (J_a) and spatial structure (S) of each PCA score map.

		Maize							
Field ID	Area†	PC 1		PC 2		PC 3		PC 4	
		Ja‡	S	Ja	S	Ja	S	Ja	S
	ha	ha							
1	17.8	0.0323	6.41	0.0037	2.25	0.0195	2.95	0.0200	3.25
2	12.8	0.0389	5.28	0.0263	2.72	0.0203	3.94	0.0113	2.03
3	12.9	0.0392	3.64	0.00484	4.89	0.00198	1.29	0.0153	1.31
4	106	0.0999	26.7	0.00501	18.1	0.00623	2.38	0.00299	6.08
8	15.3	0.0234	2.19	0.0187	4.46	0.0187	1.07	0.00872	0.634
9	11.6	0.00503	1.67	0.00142	1.75	0.00203	0.661	0.00304	0.659
10	11.3	0.0208	3.29	0.0275	1.69	0.0073	2.06	0.00847	0.861
11	39.5	0.0310	10.7	0.0985	8.51	0.0148	2.24	na	na
12	20.5	0.176	9.00	0.0757	8.24	0.0487	3.31	0.0134	1.28
13	13.5	0.0336	8.44	0.00902	0.534	0.0105	2.11	0.0290	1.18
15	16.2	na	na	na	na	na	na	na	na
18	10.6	na	na	na	na	na	na	na	na
19	9.32	na	na	na	na	na	na	na	na
20	5.51	0.0194	1.57	0.0028	1.91	0.00182	0.9	0.00288	0.995
21	17.4	0.00709	0.488	0.00171	0.37	0.00973	1.29	<0.001	0.0704
22	9.26	0.0216	2.35	<0.001	0.215	<0.001	0.148	0.0114	0.707
23	3.93	<0.001	0.333	<0.001	0.0818	<0.001	0.169	<0.001	0.0659
24	3.56	0.0305	1.88	0.00387	0.744	0.0013	0.107	0.0032	1.105

† Area = field area after 18 m of field edges removed; PC = principal component.

‡ Ja = integral scale; S = spatial structure.

Table 2.6 (Continued)

		Soybean							
Field ID	Area†	PC 1		PC 2		PC 3		PC 4	
		Ja	S	Ja	S	Ja	S	Ja	S
	ha	ha							
1	17.8	0.0228	5.48	0.0184	3.34	0.0107	3.49	na	na
2	12.8	na	na	na	na	na	na	na	na
3	12.9	na	na	na	na	na	na	na	na
4	106	na	na	na	na	na	na	na	na
8	15.3	na	na	na	na	na	na	na	na
9	11.6	na	na	na	na	na	na	na	na
10	11.3	na	na	na	na	na	na	na	na
11	39.5	na	na	na	na	na	na	na	na
12	20.5	na	na	na	na	na	na	na	na
13	13.5	na	na	na	na	na	na	na	na
15	16.2	0.0160	3.92	0.00244	2.36	<0.001	0.961	0.0011	0.333
18	10.6	0.0686	4.41	0.0183	2.85	0.00486	1.68	na	na
19	9.32	0.00800	3.89	0.0198	1.88	0.00534	1.21	0.0177	2.27
20	5.51	0.00485	1.68	<0.001	0.150	<0.001	0.0486	<0.001	0.912
21	17.4	0.0907	5.72	0.0118	1.76	0.00625	1.40	na	na
22	9.26	na	na	na	na	na	na	na	na
23	3.93	na	na	na	na	na	na	na	na
24	3.56	0.0138	0.654	<0.001	0.686	<0.001	0.322	<0.001	0.109

† Area = field area after 18 m of field edges removed; PC = principal component.

‡ Ja = integral scale; S = spatial structure.

2.6 CONCLUSIONS

In this study, we utilized baseline function, standardized principal component analysis, and geostatistics to understand the spatio-temporal grain yield variation for precision agriculture at within-field scale. This study revealed 1) spatial yield pattern to exist only under moisture-limited conditions for maize in the Coastal Plain, 2) consistent spatial yield pattern under both moisture-limited and non-moisture limited conditions for maize in the Piedmont, and 3) higher tolerance of soybean for moisture stress compared to maize and less obvious spatial yield patterns to exist. From a management perspective, both spatial and temporal management are justifiable in the Coastal Plain, whereas temporal management maybe less important in the Piedmont. For maize cropping systems, the decisions around in-season nitrogen (N) fertilizer application and rate are important. In the Coastal Plain, the use of traditional mass-balance approach based on expected yields are unsuitable since the predictions of yields at pre-plant or pre-side dress are difficult. Therefore, a use of process-based simulation models to estimate soil N supply and N uptake efficiency are necessary. In the Piedmont, mass-balance approach may be feasible but crop N demand has been shown inadequate in deciding optimum N rates, and thus further investigations are necessary to estimate temporal variations in both soil N supply and N uptake efficiency to determine the benefits of temporally variable fertilizer N rates in this province.

2.7 ACKNOWLEDGEMENTS

I am grateful to all the participated growers in the Mid-Atlantic region of the US. Also, Mike Twining, Dave Yannacci, Chris Atkinson, Jamie Kimbles , Nelson Oberholzer, and David Hertel from Willard Agri-Service for their support in providing data and field work. I also would like to thank John Dantinne for the invaluable suggestions on data analysis. Dr. Kevin Packard of the Cornell Statistical Consulting Unit also gave me with helpful suggestions.

CHAPTER 3: WITHIN-FIELD VARIATION OF CROP PRODUCTION PROFITABILITY IN THE MID-ATLANTIC USA

3.1 KEY WORDS

Corn; Cost; Profitability; Soybean; Yield

3.2 INTRODUCTION

Arable fields in the Mid-Atlantic region of the United States (US) are known to have high field variability of crop performances and nutrient losses within and among them (Kaul et al., 2005) due to the presence of a wide range of soil characteristics owing to the nature of soil formation (Fenneman, 1938) as well as topography (Chapter 2 and 4). Precision agricultural management such as variable rate soil nutrient management (Ma et al., 2013), pest control, and hybrid selection (Katsvairo et al., 2003) are available with recent technological development, but they are generally adopted only when farmers have confidence that the management changes increase field-average profitability or reduce environmental impacts (Arnholt, 2001; Erickson and Widmar, 2015; Plant, 2001). In Chapter 2, we discussed different behaviors of within-field yield variation relative to climate and physiographic province in multiple fields in the Mid-Atlantic US. We also utilized geostatistical approaches (Pringle et al., 2003) to assess the spatial structure of within-field yield patterns. We found the Coastal Plain province to be more influenced by in-season precipitation to the field-averaged yields, and which is also more pronounced within-field spatial variation under moisture limited conditions. In the Piedmont

province, yield patterns are persistent across different in-season precipitation conditions and the temporal yield variation is lower compared to the Coastal Plain. An additional dimension is the overall expected profitability of different areas within fields. If fields contain areas of negative expected profitability, overall field profitability may be enhanced by taking such field zones out of crop production, potentially adding environmental benefits as landscape buffer areas.

This study was conducted to i) assess field-average expected profitability and the amount of field area with either positive or negative expected profit for 18 fields in the Mid-Atlantic US, and ii) determine opportunities for management adjustments and increase in field-average profitability.

3.3 METHODS

The study fields are located in the states of Delaware, Maryland, Virginia, West Virginia, and Pennsylvania, between 75° 33’ 51” and 77° 54’ 49” W, and between 38° 56’ 10” and 39° 50’ 23” N, a total area of around 19,630 km^2. In this study, we only retained the fields where there were more than three years of yield data available for a single crop type. There are two climate regions, which are warm temperate climates in the southern part, and hot summer continental climates in the northern part (Peel et al., 2007). Three physiographic provinces; Coastal Plain, Piedmont, and Blue Ridge (Fenneman, 1938); and five distinctively different areas of soil characteristics were present according to the National Cooperative Soil Survey in the research area (Soil Survey Staff et al., n.d.; Figure 3.1).

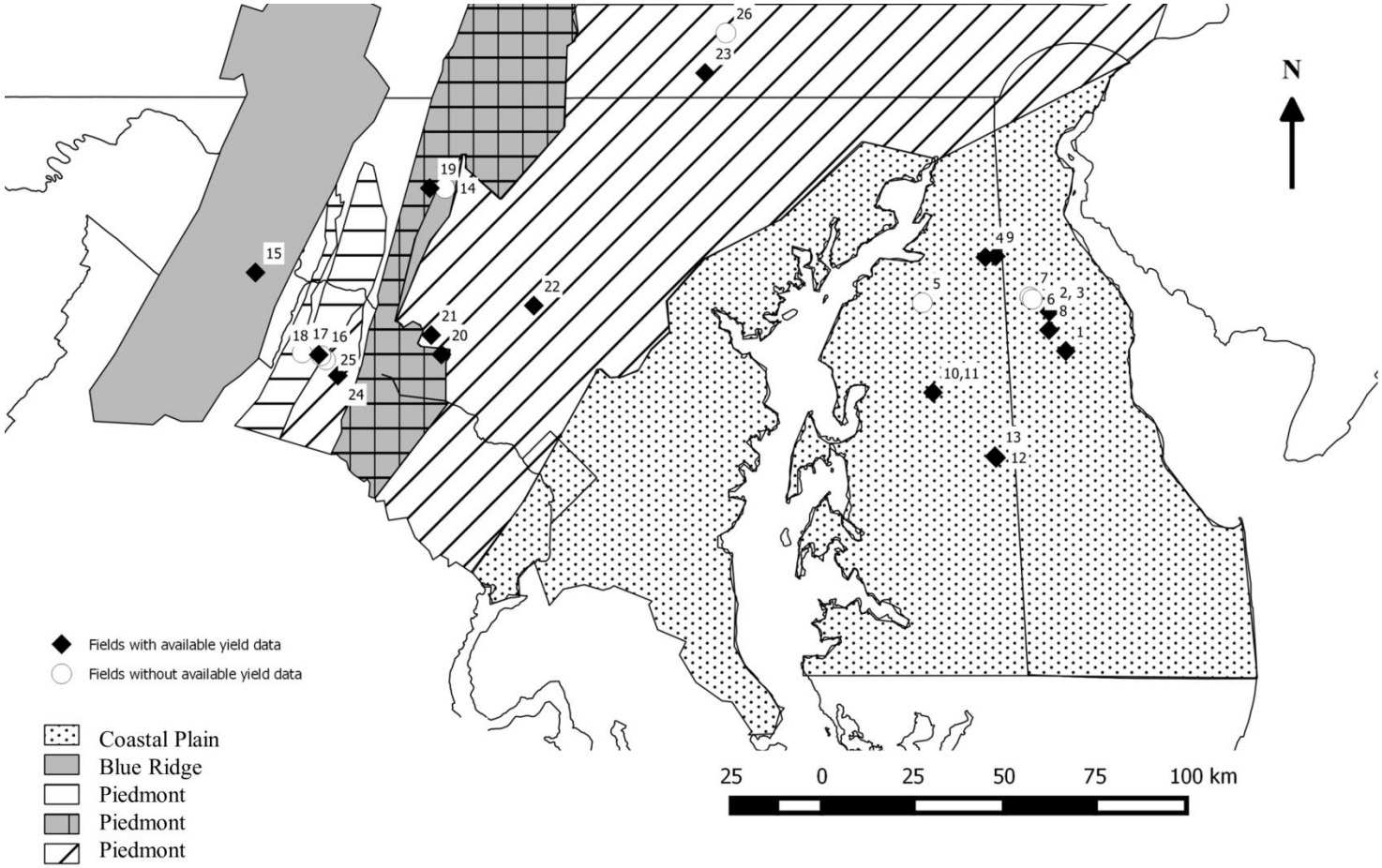

Figure 3.1 Map of the study site with indications of each physiographic province and field locations.

All of the selected fields had a similar crop rotation: corn (*Zea mays*, L.), soybean (*Glycine max*, L.) and wheat (*Triticum aestivum*. L.) or barley (*Hordeum vulgare*, L.). In some cases, double crop soybean were cultivated following the harvest of small grains. The earliest planting date for corn was the April 12th and the latest was the May 30th. For soybean, the earliest was May 13th, and July 18th was the latest for the double cropping systems. Soil nutrient, pest, weed, and irrigation water management on each field was decided according to the individual farm's management scheme.

Yield data were collected for corn and soybean with well-calibrated yield monitors on each of the combine harvesters used. The data were corrected for flow delays and slow combine velocity at the beginning and end of a pass using the Yield Editor 2.0.7 software (Sudduth and Drummond, 2007). All the point data were converted to raster files with a cell size of 6 m, which was the typical swath width of the combine harvesters.

We calculated profitability using the following relationship:

$$Profitability = E[yield] \times Price - Cost \qquad (1)$$

where E[yield] is the expected yield derived using historical average yield, Price is the average price of a particular crop, and Cost is the average cost of production. In this study, the unit for the profitability is United States \$ acre^{-1}.

We utilized the 10 year average (2004-2013) price of corn and soybean for the profitability calculation, which were \$4.29 and \$10.37 bushel^{-1}, respectively (University of Illinois, 2015). The cost of production was determined using the Farm Resource Regions of the United States Depart of Agriculture Economic Research Service (USDA-ERS, 2000) for 2014, and the study fields were distributed over three regions that were the Northern Crescent, Eastern Uplands, and Southern Seaboard. The Northern Crescent region was equivalent to the Piedmont

physiographic province; the Eastern Uplands region was equivalent to the Blue Ridge province, and the Southern Seaboard region was equivalent to the Coastal Plain province, therefore we will present the results according to the physiographic provinces.

The cost of production was substantially lower for soybean as compared to corn (USDA-ERS, 2015; Table 3.1) especially for fertilizer related costs. In addition, the land rental rate in the Coastal Plain was much lower for soybean compared to corn. We adopted two different scenarios for profitability calculation depending on field ownership. For the owned-field scenario, we subtracted the land rental rate from the total cost derived from the USDA-ERS data, whereas the rented-field scenario utilized the original data from the USDA-ERS (Table 3.2). We also estimated the cost of irrigation to be \$138 $acre^{-1}$ $year^{-1}$ (Tyson and Curtis, 2008) and it was added to the total cost (Table 3.2) when appropriate.

3.4 RESULTS AND DISCUSSION

3.4.1 Field-scale profitability

It is important to note that the cost of production used in this study was the best approximation for each field using publicly available data sources (USDA-ERS, 2015) and the actual cost of production could be widely different in individual cases.

Corn: Overall, the field-averaged profitability was higher in the Piedmont compared to the Coastal Plain, apart from Field 24. The cost of production was about \$50 $acre^{-1}$ lower for the

Table 3.1 Itemized cost of corn and soybean production in each physiographic province ($ acre^{-1}).

Cost Item	Coastal Plain		Blue Ridge		Piedmont	
	Corn	Soybean	Corn	Soybean	Corn	Soybean
Operating costs:						
Seed	82.97	55.71	69.22	55.69	93.01	62.99
Fertilizer	182.92	68.66	174.46	67.62	162.11	49.47
Chemicals	39.22	38.00	27.28	23.18	28.56	22.53
Custom operations	19.72	8.52	6.96	8.08	22.51	12.56
Fuel, lube, and electricity	42.30	16.87	24.71	17.07	29.68	17.33
Repairs	28.45	21.82	24.78	17.20	25.77	19.63
Purchased irrigation water	0.00	0.00	0.00	0.00	0.00	0.00
Interest on operating capital	0.13	0.06	0.10	0.06	0.12	0.06
Total, operating costs	395.71	209.64	327.51	188.90	361.76	184.57
Allocated overhead:						
Hired labor	4.53	2.85	2.57	3.93	3.94	1.71
Opportunity cost of unpaid labor	35.61	23.11	37.08	21.39	32.89	18.04
Capital recovery of machinery and equipment	97.19	74.29	85.38	66.42	87.82	73.45
Opportunity cost of land(rental rate)	92.45	55.71	100.40	101.81	114.28	125.97
Taxes and insurance	13.16	8.79	12.10	10.00	10.03	10.41
General farm overhead	28.07	20.86	25.55	18.30	26.07	22.82
Total, allocated overhead	271.01	185.61	263.08	221.85	275.03	252.40
Total costs listed	666.72	395.25	590.59	410.75	636.79	436.97

Table 3.2 Summarized cost of production for corn and soybean in each physiographic province ($ acre^{-1}).

		Corn		Soybean	
Physiographic Province	Water management	Owned	Rented	Owned	Rented
Coastal Plain	Rainfed	575	667	339	395
	Irrigated	713	805	477	533
Blue Ridge	Rainfed	491	591	309	411
Piedmont	Rainfed	523	637	311	437

Piedmont, and therefore the higher profitability due to higher yields was restricted to Field 21, 22, and 23. The lower cost of production was mainly identified for fertilizer and chemical related costs (Table 3.1). For the rented-field scenario, the Coastal Plain fields had overall losses (-$95.6 acre^{-1} < profitability < -$50.7 acre^{-1}; Table 3.3). In this region, irrigation appears to be an effective approach for improving profitability, and the profitability for irrigated fields ranged from -$2.23 acre^{-1} to $172 acre^{-1} (Table 3.3) even under the rented scenario. This indicates that soil moisture shortage is the major yield limiting factor in this region and an opportunity exists to achieve positive profits after accounting for the added cost of irrigation. In the Piedmont, losses were only observed under the rented scenario for Field 20 and 24.

Soybean: The profitability was overall higher compared to corn partly due to the lower cost of production (Table 3.2) by almost $200 acre^{-1}. Significant differences in the fertilizer cost, relative to corn, were identified, mostly related to a lower or to no application of N fertilizer for soybeans (Cornell University Cooperative Extension, 2012; Schmidt, 2015). In Field 1, the profitability of soybean under the owned-field scenario was $214 acre^{-1} compared to -$3.57 acre^{-1} for corn, and the difference in the cost of production was $236 acre^{-1} (Table 3.2). This was in line with the recent analysis where the return from soybean production was $75 acre^{-1} or more compared to corn in the Mid-Atlantic US (Schnitkey, 2015). In the Piedmont, the lowest profitability was the same for corn in Field 20. For Field 24, the soybean profitability was much higher at $188 acre^{-1} compared to $29.9 acre^{-1} but this can be explained by the higher cost for corn by $212 acre^{-1}. It is important to note that there were some late planting dates for soybean due to the double cropping systems, and farmers may be seeing lower profitably from this system.

Table 3.3 Summary of field-averaged profit and the amount of acreage in profit or loss for corn and soybean in each physiographic province.

ID	Physiographic Province	Water management	Area	N†	Average yield	Corn					
						Owned			Rented		
						Profitable	Non-profitable	Ave profit	Profitable	Non-profitable	Ave profit
			acre		bushel acre^{-1}	acre		$/acre	acre		$/acre
1	Coastal Plain	Rainfed	40.2	5	122.4	18.9	21.3	-3.57	3.9	36.3	-95.6
2	Coastal Plain	Rainfed	31.2	5	128.8	19.6	11.6	32.4	7.9	23.3	-59.6
3	Coastal Plain	Rainfed	31.2	5	124.0	18.3	12.9	15.5	5.7	25.5	-76.5
4	Coastal Plain	Rainfed	89.6	4	132.0	54.8	34.8	41.3	37.6	52	-50.7
8	Coastal Plain	Irrigated	58.4	4	209.9	44.6	13.8	118	41.3	17.1	25.9
9	Coastal Plain	Irrigated	35.2	4	211.5	34.4	0.845	264	32.9	2.29	172
10	Coastal Plain	Irrigated	34.7	4	192.4	32.4	2.27	192	29.3	5.36	99.6
11	Coastal Plain	Irrigated	116	3	203.5	89.6	26.4	115	73.7	42.3	22.8
12	Coastal Plain	Irrigated	59.1	6	171.7	43.9	15.2	103	37.4	21.7	11.1
13	Coastal Plain	Irrigated	40.3	4	174.9	32.3	7.99	89.8	25.5	14.8	-2.23
15	Blue Ridge	Rainfed	47.8	1	na	na	na	na	na	na	na
18	Piedmont	Rainfed	31.2	1	na	na	na	na	na	na	na
19	Piedmont	Rainfed	28.3	1	na	na	na	na	na	na	na
20	Piedmont	Rainfed	19.1	5	132.0	14.7	4.4	74.4	7.8	11.3	-39.6
21	Piedmont	Rainfed	49.9	5	181.3	49.2	0.712	283	46.7	3.22	169
22	Piedmont	Rainfed	28.9	4	138.3	25.4	3.49	137	18.1	10.8	22.9
23	Piedmont	Rainfed	17.2	4	155.8	16.7	0.52	224	15.2	2.01	110
24	Piedmont	Rainfed	14.4	7	124.0	8.25	6.15	29.9	5.01	9.39	-84.1

† N = number of years of available yield data

Table 3.3 (Continued)

				Soybean							
						Owned			Rented		
ID	Physiographic Province	Water management	Area	N†	Average yield	Profitable	Non-profitable	Ave profit	Profitable	Non-profitable	Ave profit
			acre		bushel acre^{-1}	acre		$/acre	acre		$/acre
1	Coastal Plain	Rainfed	40.2	3	50.7	40.1	0.0979	214	39.9	0.285	158
2	Coastal Plain	Rainfed	31.2	1	na	na	na	na	na	na	na
3	Coastal Plain	Rainfed	31.2	1	na	na	na	na	na	na	na
4	Coastal Plain	Rainfed	89.6	2	na	na	na	na	na	na	na
8	Coastal Plain	Irrigated	58.4	2	na	na	na	na	na	na	na
9	Coastal Plain	Irrigated	35.2	1	na	na	na	na	na	na	na
10	Coastal Plain	Irrigated	34.7	1	na	na	na	na	na	na	na
11	Coastal Plain	Irrigated	116	0	na	na	na	na	na	na	na
12	Coastal Plain	Irrigated	59.1	0	na	na	na	na	na	na	na
13	Coastal Plain	Irrigated	40.3	1	na	na	na	na	na	na	na
15	Blue Ridge	Rainfed	47.8	7	44.7	47.6	0.16	170	43.9	3.86	67.8
18	Piedmont	Rainfed	31.2	3	58.1	31	0.24	317	29.6	1.6	191
19	Piedmont	Rainfed	28.3	4	37.3	26	2.3	116	13.5	14.8	-10.5
20	Piedmont	Rainfed	19.1	5	32.8	15.2	3.86	53.5	2.9	16.2	-72.5
21	Piedmont	Rainfed	49.9	3	58.1	49.9	0	305	49.7	0.178	179
22	Piedmont	Rainfed	28.9	2	na	na	na	na	na	na	na
23	Piedmont	Rainfed	17.2	2	na	na	na	na	na	na	na
24	Piedmont	Rainfed	14.4	5	46.2	13.9	0.47	188	11.3	3.08	62.4

† N = number of years of available yield data

3.4.2 Spatial patterns of profitability and opportunities for alternative land uses

We identified three general categories in within-field variation of profitability: "economically sensitive" (Figure 3.2a and 3.2b), "clear profitability zones" (Figure 3.2c and 3.2d), and "all profitable" (Figure 3.2e and 3.2f). We used the owned-field scenario of corn for the visualization of Figure 3.2. "Economically sensitive" fields have the majority of the area with profitability ranging from -\$100 $acre^{-1}$ to \$100 $acre^{-1}$, where the change in the cost of production or the price of grain can readily move profits from positive to negative, or vice versa. "Clear profitability zone" fields have distinct zones of high (> \$100) and low (< -\$100) profitability, and the zones are little impacted by production prices. "All profitable" fields have most of the areas in high profitability (> \$100), and all areas generally remain profitable under price fluctuations.

The economically sensitive fields were in the Coastal Plain for which we identified high temporal yield variation in Chapter 2. This indicates that the profitability at a particular location can be high or low depending on the growing season environmental conditions. The small margin in profitability emphasizes the need for the reduction in cost as well as improvement in yields.

In the fields with clear profitability zones, spatially explicit management is recommended for the profitable areas. For the consistently unprofitable regions of the fields, consideration should be given to taking the field areas out of production, or identifying and ameliorating potential yield constraining factors like soil compaction and low pH

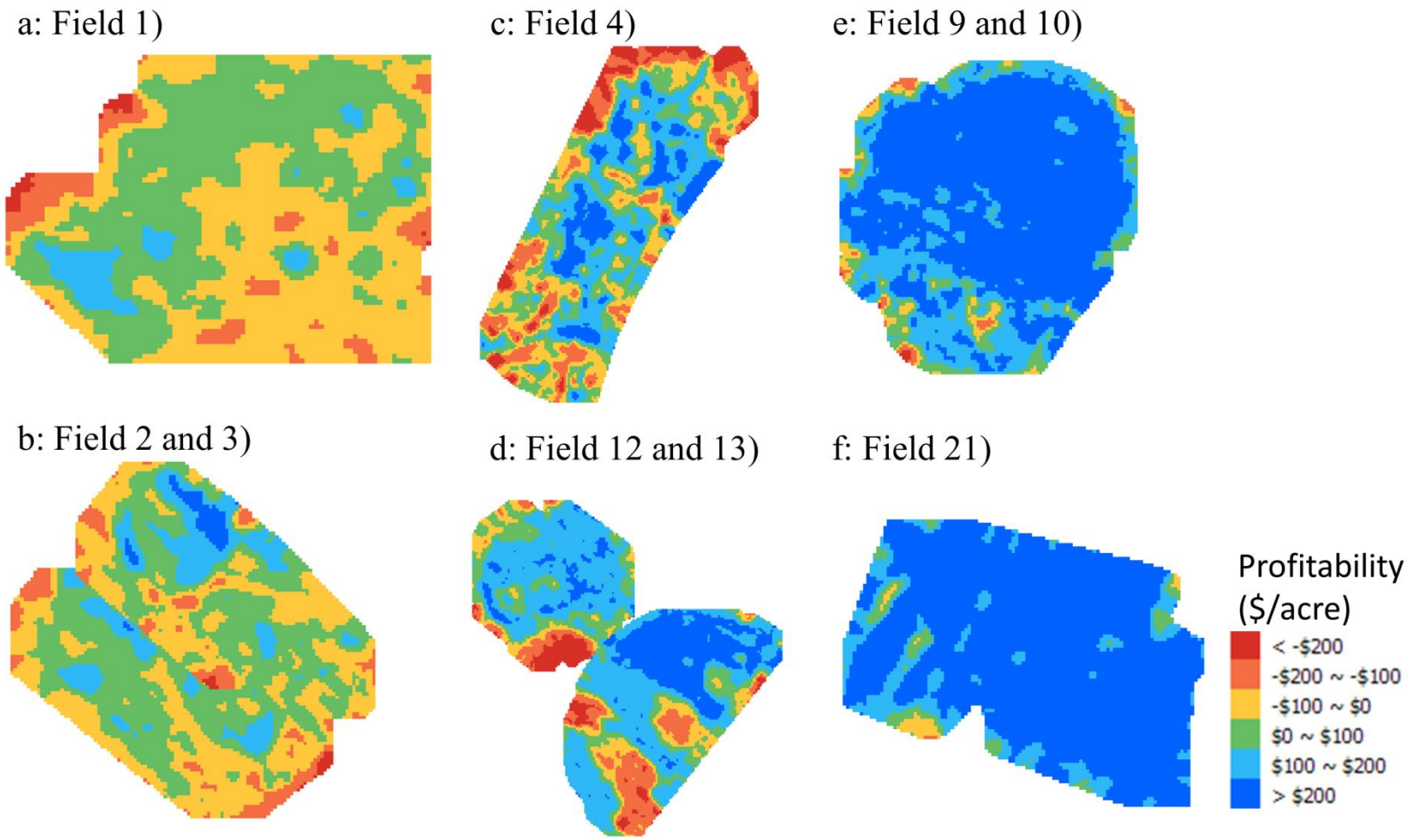

Figure 3.2 Maps of within-field profitability for the owned-field scenario for corn. There were three profitability categories; a and b) economically sensitive; c and d) clear profitability zones; and e and f) all profitable.

(Baveye and Laba, 2015; Oliver, 2010). Field 4 (Figure 3.2c) had a distinct area of low profitability along the edge of the fields on the northern side due to shading by surrounding woodland. Fields 12 and 13 (Figure 3.2d) had areas of low profitability; the combination of lower elevation and poorly drained soil (Soil Survey Staff et al., n.d.) were the constraining factor. In this field, low profitability was contingent on excessive early-season precipitation, and thus in-season cost-saving decisions may be considered (e.g., foregoing siderdress N application for those areas). In Field 4, the low profitability areas along the treeline can be removed from production and converted for other land uses including herbaceous buffer strips. Past research has shown that an introduction of herbaceous strips on the field borders can promote wildlife habitats and reduce nutrient losses (Borin and Bigon, 2002) while showing minimum effects on the yields of the main crop (Stamps et al., 2008). We assessed the opportunity of removing the low profitability areas (< -\$200 acre^{-1}) in Field 4 and recalculated the profitability. The removal of the unproductive areas increased the average field profitability from \$41.3 to \$67.2 acre^{-1}.

There were several fields where the majority of the areas were under high profitability (> \$200 acre^{-1}; Figure 3.2e and 3.2f). These fields may be already undertaking higher-than-average inputs to achieve higher yields and thus the cost of production could be higher than the scenarios we used.

3.5 CONCLUSIONS

We identified substantial variation in within- and among-field profitability across the Mid-Atlantic US, which revealed opportunities for precision agricultural management either temporally or spatially, or both. Under our assumptions of the cost of production and the price of

grain, soybean appeared to be more profitable overall compared to corn in this region under these scenarios. For corn, the field-averaged profitability was higher in the Piedmont compared to the Coastal Plain fields. In the latter region, high temporal variation related to in-season precipitation appeared to be the yield constraining factor (Chapter 2). Here, irrigation systems increased profitability substantially, even after accounting for the additional cost.

In general, there were three categories of within-field variation of profitability (economically sensitive, clear profitability zones, and all profitable). We identified the economically sensitive fields to match those with high temporal yield variation (Chapter 2), and thus there is a need for temporally- as well as spatially-explicit management. In fields with clear profitability zones, we showed the importance of identifying the potential yield constraints, or land use conversion for conservation purposes if the yield constraint cannot be easily removed, which can both enhance field-average profits.

3.6 ACKNOWLEDGEMENTS

I am grateful to all the participated growers in the Mid-Atlantic region of the US. Also, Mike Twining, Dave Yannacci, Chris Atkinson, Jamie Kimbles , Nelson Oberholzer, and David Hertel from Willard Agri-Service for their support in providing data and field work. I also would like to thank John Dantinne for the invaluable suggestions on data analysis.

CHAPTER 4: IDENTIFICATION OF WITHIN-FIELD SPATIAL PATTERNS OF YIELD POTENTIAL AND CROP GROWTH CONSTRAINTS IN THE MID-ATLANTIC USA

4.1 ABSTRACT

Predicting the within-field spatial patterns of yield potential (Yp) and the identification of site-specific yield constraints are necessary for determining soil and crop management strategies. Information regarding grain yields, proximal sensing data, and topographical and soil properties were collected from 26 arable fields in the Mid-Atlantic USA. The grain yield information was processed using standardized principal component analysis (stdPCA) in Chapter 2. We found the ratio of shallow (0-to-45 cm depth) and deep (0-to-90 cm depth) apparent electrical conductivity (ECR) collected by an on-the-go sensor as an effective method for estimating water holding capacity and subsoil soil textural variation in the Coastal Plain province (r = 0.77). This was the best predictor when estimating within-field spatial patterns of Yp in the area using a Random Forest (RF) method combined with surface pH and topographic wetness index (TWI). In the Piedmont province, the correlations of ECR to soil properties were lower (dry bulk density; r = 0.33), and topographical properties (aspect and slope) were more important when assessing Yp. We successfully applied pattern similarity distance (d) to compare the spatial patterns of measured and model predicted Yp, and the model predictability was higher in the Coastal Plain ($\bar{d}$ = 0.234, n = 3) compared to the Piedmont ($\bar{d}$ = 0.245; n = 8). The d values were effective in identifying areas where the RF model successfully explained the spatial patterns,

which indirectly suggests that the soil properties associated with the best predictor in RF are correlated to underlying yield constraints. Furthermore, large d-values represent the areas where those yield constraints failed to predict the spatial patterns of Yp, and can be used as the targeted soil sampling locations.

The combination of proximal sensing, measured soil property information, RF, and d values showed potential in estimating the spatial patterns of Yp and targeting potential yield constraints, and further assessment with larger number of calibration fields could improve model predictability.

4.2 KEY WORDS

Physiographic province; Proximal sensing; Random Forest; Subsoil; Topography

4.3 INTRODUCTION

Arable fields in the Mid-Atlantic region of the United States are known to have high within and among field variability in crop performance, nutrient losses, and profitability (Kaul et al., 2005) due to the presence of a wide range of soil characteristics owing to the nature of soil formation (Fenneman, 1938). In Chapter 2, we revealed the importance of physiographic province in the Mid-Atlantic US in understanding the behaviors of within-field spatio-temporal yield variation for maize (*Zea Mays*, L.) using standardized principal component analysis (stdPCA). We showed spatial yield patterns becoming more apparent under moisture-limited conditions in all the selected fields in the Coastal Plain province, while consistent yield patterns

existed for both moisture-limited and non-moisture limited conditions in the Piedmont province. Standardized PCA also helped capture characteristic yield patterns while removing the effects of sporadic yield patterns (Eastman and Filk, 1993). From a management perspective, this information indicated the opportunities for both spatially- and temporally-explicit management in the Coastal Plain while the latter maybe less important in the Piedmont.

Recent technological developments in precision agriculture have provided farmers the capability to undertake variable rate soil nutrient management (Ma et al., 2013), pest control, and hybrid selection (Katsvairo et al., 2003) but they are generally adopted only when farmers have confidence that the management changes increase field-average profitability or reduce environmental impacts (Arnholt, 2001; Erickson and Widmar, 2015; Plant, 2001). The sole use of grain yield monitor information may not be sufficient to formulate management plans since it does not explicitly identify yield constraints that can be caused by a wide range of factors including topography (Basso et al., 2009; Jiang and Thelen, 2004), genotype (Yang et al., 2009), fertilizer management (Katsvairo et al., 2003), soil properties (Jiang and Thelen, 2004; Miller et al., 1988; Shahandeh et al., 2005), and their interactions. The variable rate management may then not be beneficial if the yield constraints are not effectively addressed (Baveye and Laba, 2015).

In the Mid-Atlantic US, many of farm fields are leased (≈46%; USDA, 2009) and the farmer knowledge of the constraints of those fields may be limited. Also, there are challenges in collecting multiple years of yield data of a single crop due to the existence of multiple crop rotations, data handling issues, and possible limitations from inadequate yield monitor calibration. Therefore, there is a need to i) estimate site-specific yield potential (Yp; van Ittersum and Rabbinge, 1997) in the absence of grain yield data and ii) identify spatially explicit plant growth constraints in the region.

Within-field Yp may be predicted using approaches including multiple linear regression (Kitchen et al., 2003) machine learning (Kaul et al., 2005; Kitchen et al., 2003), and Soil-Plant-Atmosphere (SPA) system models (Grassini et al., 2015; Hochman et al., 2012; Oliver et al., 2010) although applications of these for within-field assessment have been limited by the data available to parameterize the models. Existing soil information such as the NRCS soil survey provide static soil information but the spatial resolution is often inadequate (USDA-NRCS, 2015), and temporal changes due to agricultural management are not well represented (Tugel et al., 2005). In grain production systems in the US, the applications of remote and proximal soil sensing techniques (Kuang et al., 2012) has provided information that could link spatio-temporal variation of crop performance to contributing environmental factors (Basso et al., 2009; Jiang and Thelen, 2004). Proximal sensing equipment such as the Veris Mobile Sensor Platform (MSP-3; Veris Technologies, Salina, KS) combines apparent EC (ECa; Corwin and Lesch, 2003), optical sensing (Kweon and Maxton, 2013), on-the-go pH metering (Adamchuk et al., 1999) and high-precision GPS to undertake continuous measurement of within-field soil and topographical variations. The relative correlation of the proximal sensor information to particular soil properties and the feasibility of building soil property prediction models depend on the interactions among the influential properties including soil salinity, water content, soil texture, bulk density, organic matter, and distance between soil surface and the sensor (Corwin and Lesch, 2005; Kweon and Maxton, 2013). When the calibration of the sensor values to particular soil properties is challenging, empirical models can be built to predict within-field Yp using the sensor values directly (Kitchen et al., 2003; Moral et al., 2010; Stadler et al., 2015). Past research has shown that machine learning approaches can perform better compared to linear regression type models (Kitchen et al., 2003). Machine learning is related to mathematical optimization that

allows for an incorporation of non-linear relationships with little a priori knowledge of the functional relationship (Kaul et al., 2005). General yield constraints can be also inferred by determining the correlations between the proximal sensing values to measured soil properties (Kitchen et al., 2003; Stadler et al., 2015), whereas explicit soil analyses may be necessary when the empirical model fails to predict the Yp.

This study was conducted on 26 arable fields in the Mid-Atlantic region of the US over three physiographic provinces. The objectives were to i) assess the spatial variation of surface and soil profile soil properties and topography across a range of arable fields in the region, ii) assess the correlations between the measured soil properties and proximal sensor information, iii) assess the feasibility of predicting within-field Yp using an empirical model, and iv) explore potential site-specific yield constraints using the empirical model and soil samples.

4.4 MATERIALS AND METHODS

4.4.1 Site description

We selected 26 fields from the states of Delaware, Maryland, Pennsylvania, Virginia, and West Virginia that were under maize (*Zea mays*, L.), soybean (*Glycine max*, L.), and wheat (*Triticum aestivum*. L.) or barley (*Hordeum vulgare*, L.) cropping systems. They were located between 75° 33’ 51” and 77° 54’ 49” W, and between 38° 56’ 10” and 39° 50’ 23” N, a total area of approximately 19,630 km^2. There are two climate regions, warm temperate climates in the southern part and hot summer continental climates in the northern part (Peel et al., 2007).

Three physiographic provinces (Fenneman, 1938) and five distinctively different areas of soil characteristics were present according to the National Cooperative Soil Survey (Soil Survey Staff et al., n.d.). The first physiographic province (Area 1) was the Coastal Plain province, and the latter two were the Piedmont and Blue Ridge provinces (Area 2). Area 1 is mainly associated with Typic Hapludults and predominantly formed on coastal plain deposits (sandy loam; Figure 4.1) below varying depths (40-to-100 cm depth) of aeolian silt deposits that are very acidic (Simonson, 1982). Closer to the Chesapeake Bay on the Eastern Shore of Maryland, the surface silt deposits could reach more than 150 cm deep (Foss et al., 1978). Area 2 had undulating topography with four distinct soil characteristic areas, which are Hagerstown Limestone Valley, Middletown Valley, Western Piedmont, and Piedmont Crystalline Rocks (Figure 4.1). Hagerstown Limestone Valley is mainly associated with Typic Hapludalfs and the soil pH is neutral to slightly acid and has low rock content; Middletown Valley is associated with Typic Hapludults on moderate to strongly acid soil and has high rock contents; Western Piedmont is associated with Ultic Hapludalfs formed on reddish parent materials of Triassic period and is moderately acidic and extremely rocky; and Piedmont Crystalline Rocks is associated with Typic Hapludults and has thin to intermediate depth (< 75 cm) aeolian silt deposits on the surface (Simonson, 1982; Weaver, 1967) and is moderately acidic and rocky.

4.4.2 Yield data acquisition

Detailed descriptions about the yield data acquisition and analyses are described in Chapter 2. In short, well-calibrated yield monitors combined with GPS units were used to collect spatially-referenced yield data. Yield Editor software (Sudduth and Drummond, 2007) was used

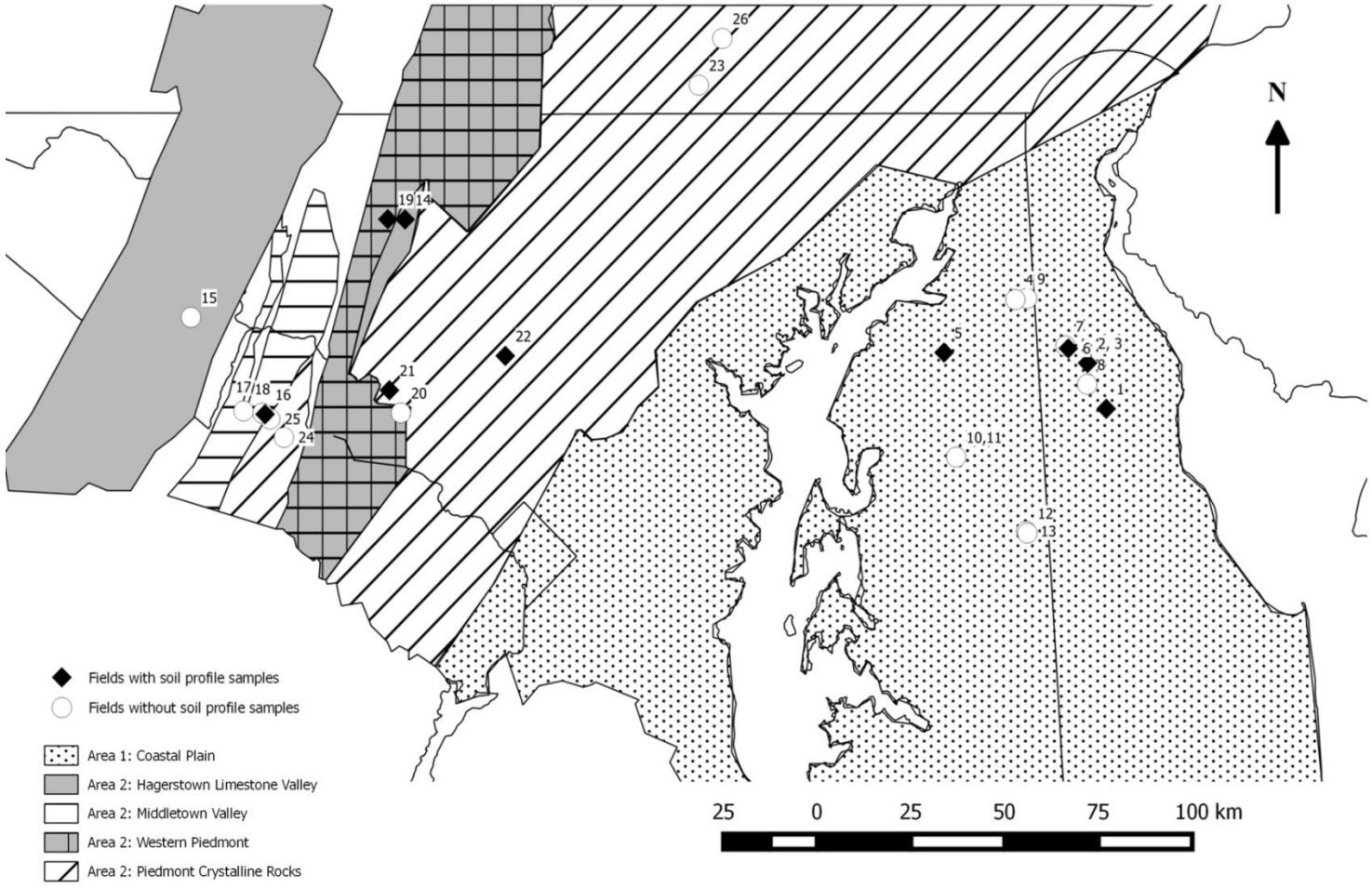

Figure 4.1 Map of the study site with indications of each characteristic soil region and field locations.

to remove data points with potential errors (i.e. flow delays and slow combine velocity at the beginning and end of each pass), and the data were rasterized to 6 m by 6 m grids using the SAGA function (Conrad et al., 2015) within the QGIS environment (QGIS Development Team, 2015). Field borders (18 m) were removed from the analyses where we expected and visually observed unusually low grain yields caused by factors including reduced incident solar radiation by surrounding treelines, soil compaction due to higher traffic, yield monitor recording errors associated with slower tractor velocities, and higher pest pressures.

We selected nine fields where there were three to seven years of yield records available, and undertook stdPCA (Eastman and Filk, 1993), which is a linear multivariate transformation method by retaining the maximum amount of original variance in the first principal component (PC). Major yield patterns are removed in the first PC and subsequent PCs may reveal hidden yield patterns. The transformed values in the PC space are called scores and the varying effects of in-season precipitation conditions to the score values for each study field was unique and described in Chapter 2.

4.4.3 Field and laboratory data collection

4.4.3.1 Apparent electrical conductivity, optic sensors, pH, and elevation

Prior to any soil sampling, proximal sensor information was collected using a Veris MSP-3 unit (Veris Technologies, Salina, KS). This unit was equipped with an RTK-GPS (ParaDyme; 2.54 cm horizontal and 5.08 cm vertical accuracy; Ag Leader Technology, Ames, IA), apparent electrical conductivity (ECa) sensors (≈0-45 cm and ≈0-90 cm depth), optic

sensors (660 nm and 940 nm), and an on-the-go pH sensor. These wavelengths were selected in studies using the Midwest soils to predict soil organic matter (SOM; Kweon and Maxton, 2013). The data collection was carried out between March and November, 2014. All sensors were calibrated before each data collection following the operating instructions (Veris Technologies, 2012) and we undertook the measurement on transacts at 18 m intervals across a field. Recorded ECa, optic sensor, and pH data were error corrected by the Veris Technologies company and interpolated to 6 m by 6 m grids using inverse distance weighting (IDW) and a SAGA function (Conrad et al., 2015) in the QGIS environment (QGIS Development Team, 2015).

Variable field conditions including soil moisture content and soil temperature are known to affect ECa and optic sensor values (Corwin and Lesch, 2005; Kweon and Maxton, 2013), and therefore we collected in-situ soil moisture information at the time of measurement. In addition, we ran the unit on the same field twice, before planting (April 11th) and after harvest (September 23rd) in Field 1 to assess the effects of in-situ soil conditions on the sensor values.

The elevation data were first assessed for erroneous values by searching for significantly different values from the surrounding elevation. Elevation data were then interpolated using a regularized spline function with a tension of 10 (Mitášová and Hofierka, 1993) to 6 m by 6 m grids, which has been shown as a more accurate method compared to IDW (Williams et al., 2011). The interpolated elevation data were used to derive various topographical properties such as slope (SLOPE), aspect (ASP), profile curvature (PROF), tangential curvature (TAN), and mean curvature (CURV). We also derived topographic wetness index (TWI), which generally shows the zones of soil water saturation within a landscape (Moore et al., 1991) using the SAGA GIS function (Conrad et al., 2015):

$$TWI = \ln(\frac{a}{tanb}) \quad (1)$$

where *a* is the specific (contributing) catchment area and *b* is the slope gradient.

Topographic position index (TPI) was also calculated, which is the categorized elevation difference between a point in a landscape and the set size of surrounding cells as neighborhood:

$$TPI = z_0 - \bar{z} \quad (2)$$

where z_0 is the elevation at the central point and $\bar{z}$ is the average elevation within a predetermined radius (Gallant and Wilson, 2000; Weiss, 2001). We determined TPI using two different neighborhood sizes and used the combination of two TPI categories to classify each cell in 10 different landform classes (Weiss, 2001) using a SAGA GIS function (Conrad et al., 2015). We reclassified these 10 classes into four landform classes: swale, flat, slope, and knoll.

4.4.3.2 Soil sampling

In all of the 26 fields, shallow soil samples (0-to-15 cm depth; shallow sample set) were collected at six locations, at each compositing eight push probe (2.06 cm i.d.) draws that were 60 cm apart in the inter-row of the previous crop row adjacent to the optic sensor track. The sampling locations were determined using a combination of collected shallow ECa (0-45 cm depth; ECsh) values and optic sensor information at 940 nm (IR). Each field was divided into six equal-sized grids and two samples were collected from high ECsh and low IR areas; two samples from low ECsh and high IR areas; one sample from high ECsh and high IR area, and one sample from low ECsh and low IR area classified using the Veris SoilViewer software (Veris Technologies, Saline, KS). In total, we had 142 samples in the shallow sample set.

Only selected fields (n = 9) were assessed for soil profile sampling (deep sample set) in this study and we selected at least one field from each of five distinctly different soil

characteristic areas (Figure 4.1). In Area 1, four fields were subjected to full-profile soil sampling. Nine soil samples were collected from the 0-to-90 cm depth using the JMC Environmentalist's Sub-Soil Probe (2.88 cm i.d.; Clements Associates, Newton, IA). Each field was divided in nine equal-sized rectangle areas, and judgment sampling was adopted according to the deep ECa (0-90 cm depth; ECdp) values. We calculated the quantiles of 1/3 and 2/3 of the mean values and took three samples from each ECdp level (low, medium, and high) while maximizing the distance between the sampling grids of the same ECdp level. At each sampling point, 3 subsamples were taken from non-traffic inter-row of the previous crop within a 2 m^2 quadrant. The depth of the borehole and the length of the extracted core were measured and compared for estimation of soil compaction during coring (Schrumpf et al., 2014). Each subsample was cut in increments at 0-15, 15-30, 30-45, 45-60, and 60-90 cm depth and composited once brought back to the laboratory. In Area 2, five fields were subjected for soil sampling, two from the Piedmont Crystalline Rocks and one from each of the three characteristic soil areas. We needed to modify the sampling protocol due to high rock contents except for Field 22. Six soil samples were collected from the 0-to-60 cm depth using a diamond tipped rotary core (9.44 cm i.d.; Lackmond Products, Inc., Marietta, GA) assisted with a gas motor. We modified the judgment sampling scheme by combining ECdp data with available maize yield data. The mean maize yield was categorized in three classes following the same method as the ECdp values using quantiles, while yield stability was classified in two categories (stable and variable) at the median coefficient of variation (CV; Abuzar et al., 2004). One sample was taken from each combination of the ECdp and yield data but only when the same class occupied more than 2023 m^2 (≈ 0.5 acre).

- High ECdp & High Mean Yield & Stable Yield

- Low ECdp & High Mean Yield & Stable Yield
- High ECdp & Low Mean Yield & Stable Yield
- Low ECdp & Low Mean Yield & Stable Yield
- High ECdp & Variable Yield
- Low ECdp & Variable Yield

When no areas larger than 2023 m^2 existed, we took samples from a smaller area. Only one soil core was taken at each location and was cut in increments at 0-15, 15-30, 30-45, and 45-60 cm depth. In total, we had 66 sampling locations and 298 samples in the deep sample set. We recorded the geographical coordinates using a hand-held GPS unit (eTrex Venture HC; Garmin, Schaffhausen, Switzerland). All of the collected soil samples were stored at 4 °C and transferred to the laboratory.

4.4.3.3 In situ soil penetration and moisture measurement

Soil penetration resistance (Pen) was measured using a soil compaction meter at the first 45 cm of the soil profile where a deep sample set was taken using the FieldScout SC900 Soil Compaction Meter (Spectrum Technologies, Aurora, IL). At each soil subsample point, two penetrometer measurements were taken within a 30 cm radius and a total of 6 measurements were averaged per sample point. In-situ soil moisture content was measured using the Aquaterr moisture meter (model EC-350, Aquaterr Instruments, Fremont, CA) in Area 1 at 7.5, 22.5, 37.5, and 52.5 cm depth, which measures the dielectric constant of the soil-air-water combination (Sánchez-Gómez et al., 2006). The measured values were converted to volumetric moisture content using equations developed by Proulx (2001). In Area 2, the FieldScout TDR100 Soil

Moisture Meter (Spectrum Technologies, Aurora, IL) was used due to the high rock content and the inability of the Aquaterr Moisture Meter to penetrate the soil profile. The Field Scout TDR 100 was used after excavating the 1 m^2 area to the depth of 50 cm and it directly estimates volumetric water content. The moisture measurement was made at three sides of the soil pit at equivalent depth increments as Area 1 and averaged.

4.4.3.4 Soil analyses

Whole soil samples were first weighed at field moisture state and then air-dried by forced air until the soil sample weight was stabilized. The soil sample was then crushed using a metal ring and passed through a 2-mm sieve. Any non-organic materials that were bigger than 2-mm were retained, washed, oven dried at 105 °C, and then weighed to determine the mass of coarse rock fragments. The mass was then converted to volume and used to estimate the dry bulk density (ρ_b), which was calculated using the known volume of the soil sampling probe. The field soil moisture content could not be determined for the deep sample set in Area 1 due to the soil sampling equipment used.

For all soil samples, soil texture was assessed using a rapid quantitative method developed by Kettler et al. (2001). The soil sample was dispersed with 3% sodium hexametaphosphate ($(NaPO_3)_n$) and a combination of sieving and sedimentation steps was used to separate size fractions. Water contents at -10 kPa, -33 kPa, -100 kPa, and -1500 kPa were assessed gravimetrically. Saturated soil samples were equilibrated to each pressure on ceramic high pressure plates (Topp et al., 1993). Dry bulk density was used to convert the gravimetric water content to volumetric water content.

The shallow sample set, the first increment of the deep sample set, and all increments of the deep sample set from Field 1 and 22 were subjected to soil organic matter (SOM), soil pH, and soil nutrient analyses by Spectrum Analytic Inc. (Washington Court House, OH), and water stable aggregation (WSA) at the Cornell Soil Health Laboratory (Ithaca, NY). Soil organic matter was analyzed by mass loss on ignition in a muffle furnace at 360°C for two hours. Soil pH was measured in 1:1 water slurry. Other soil nutrients, including P, K, Mg, Ca, Zn, Cu, S, and Al were extracted using Mehlich III extraction and quantified by inductively coupled plasma optical emission spectrometry. Water stable aggregation was assessed using a rainfall simulator (Ogden et al., 1997) that allows the soil particles to receive the impacts of known rainfall energy, applying 2.5 J of energy for 300 s on aggregates (0.25–2 mm) placed on a 0.25-mm mesh sieve. The fraction of soil aggregates remaining on the sieve, corrected for stones >0.25 mm, was regarded as the percent WSA after drying at 105°C (Gugino et al., 2009).

4.4.4 Data analysis

4.4.4.1 Exploratory data analysis

We first assessed a scatter matrix of the shallow sample set and found non-linear associations among the measured values. Therefore, Spearman's rank correlation coefficients were calculated to assess the relationships among the measured properties. We then assessed the correlations of each proximal sensor value to measured soil properties using the Pearson correlation after verifying the linear associations. For the optic sensors, we used both the surface (0-15 cm depth) increment of the deep sample set and the shallow sample set, while the ECa

sensor values were assessed against the deep sample set only to match the depths of the measurements. The on-the-go pH sensor measured values were validated against the laboratory measured soil pH values using both the surface (0-15 cm depth) increment of the deep sample set and the shallow sample set.

4.4.4.2 Random Forest

Random Forest (RF) was used to predict within-field variable PC scores, proxy for Yp using the measured spatially continuous variables including topographical properties and proximal sensor data. The method has been utilized to predict various soil properties at different geographical scales (Grimm et al., 2008; Hengl et al., 2015; Kinoshita et al., 2016) but has not been applied widely for the estimation of grain yields. It is an extension of randomized classification and regression trees (CART; Breiman et al., 1983) with bagging of a set number of trees (Breiman, 1996). This method has several advantages over other statistical models (Breiman, 2001; James et al., 2013) such as its ability to model high dimensional non-linear relationships in mixed categorical and continuous predictors (Grimm et al., 2008). It is also resistant to overfitting and internally calculates unbiased measures of estimation error (James et al., 2013), which essentially eliminates the need to undertake a separate cross validation (CV).

First, we developed a model for all the fields in the region with available stdPCA scores (n = 9), and then we built physiographic province-specific models (i.e., global rainfed maize model, Area 1 rainfed maize model, and Area 2 rainfed maize model). Before assessing the predictability for each field, there was a need to select the most appropriate model. We first fit each RF model to the full dataset and assessed the % variance explained by the model.

Subsequently, we undertook 3-fold cross-validation and calculated R^2 and Root Mean Square Error (RMSE) for each model:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(v_{pred.i} - v_{meas.i})^2}{n-1}} \quad (3)$$

where $v_{pred.i}$ is the score value predicted by the RF model at each 6 m by 6 m cell and $v_{meas.i}$ is the calculated score value using measured yield data in Chapter 2.

We used the *randomForest* package (Liaw and Wiener, 2002) in the R statistical computing environment. RF can restrict the number of predictors to utilize at each node when a split in tree is considered (m_{try}) in an effort to create uncorrelated trees to maximize variance reduction (James et al., 2013). We decided to use 400 trees and a third of the number of variables as m_{try}.

4.4.4.3 Model validation

We validated each RF model using full-site cross-validation and independent-site validation. The cross-validation was carried out by removing one field as a validation field and iterated the procedure until all fields were used for both calibration and validation. Independent-site validation was undertaken using the fields that were not utilized to calculate stdPCA scores (n = 11). The numbers of yield data were not sufficient to calculate stdPCA scores at the independent validation sites, therefore we used mean yield values from raw yield data for independent-site validation. In Chapter 2, it was apparent that stdPCA scores for maize in Area 1 represent the yield patterns under moisture-limited growing conditions for maize (abundant and well-distributed rainfall < 75; Chapter 2), whereas the scores represented yield patterns for all growing seasons in Area 2. Therefore, the mean yield values were calculated using only

moisture-limited years in Area 1, whereas all yield records were used for calculating the mean yield values in Area 2.

For cross-validation, we calculated R^2 and Root Mean Square Error (RMSE) for each model.

For both cross-validation and independent-site validation, pattern similarity distance (d; Davis, 1986) was calculated, which allows for quantification of the similarity of two spatial patterns. This method uses a moving window of a set size that relocates across a field and fits a trend surface at each step for each of the two spatial patterns. We then assessed the differences of the model coefficients of each spatial pattern at each cell. First, we standardized both the measured and predicted data to standard scores:

$$z_{\pi,xy} = \frac{v_{\pi,xy} - \bar{v}_{\pi}}{s_{\pi}} \quad (4)$$

where $\bar{v}_{\pi}$ and s_{π} are mean and standard deviation of the measured and predicted values in pattern π. We used the window size of 42 m by 42 m which is approximately equivalent to 0.5 acres. This was the minimum size of an area for which farmers can effectively undertake site-specific management (M. Twining, personal communication.). We then fit a third-degree two-dimensional polynomial function to each window (Van Uffelen et al., 1997):

$$Y(X,Y) = b_{\pi,0} + b_{\pi,1}(X) + b_{\pi,2}(Y) + b_{\pi,3}(X^2) + b_{\pi,4}(Y^2) + b_{\pi,5}(XY) + b_{\pi,6}(X^3) + b_{\pi,7}(Y^3) + b_{\pi,8}(X^2Y) + b_{\pi,9}(XY^2) + \varepsilon \quad (5)$$

using a least square regression fit. Then, d was determined by:

$$d = \sqrt{\sum_{i=0}^{p-1} \frac{(b_{1,i} - b_{2,i})^2}{p}} \quad (6)$$

where p is the number of polynomials (p = 10), and leading numbers (1 and 2) indicate the two patterns for comparison. The d was determined for each 6 m by 6 m cell and thus can be

mapped to assess locations of the agreement of two spatial patterns. Perfect matching of the RF modeled spatial pattern will indicate the d of zero. There is no statistical assessment for threshold d values, but past studies relied on visual assessment and concluded that it depends on the scope of the problem (Van Uffelen et al., 1997). In order to compare the model performance among fields, we calculated the mean d value ($\bar{d}$) for each field.

4.5 RESULTS AND DISCUSSION

4.5.1 Measured soil and topographical properties

We measured a range of soil physical, chemical, and biological properties at the surface and subsoil horizons that revealed complex relationships among the soil properties as well as their strong associations with soil forming factors.

The areas 1 and 2 showed distinct differences in surface soil texture where Area 1 showed a large amount of samples in sandy loam and loamy sand categories, while Area 2 showed a higher amount of samples in loam category (Figure 4.2) due to higher clay content (Table 4.1; Appendix A). The characteristic loess silt deposition (Simonson, 1982) was observed in Area 1, especially in Field 5 where the silt content remained high to a depth of 90 cm (Appendix B). In other fields in Area 1, the loess silt deposit (% silt > 30) was observed to the depth of 60 cm. In Area 2, silt content was high throughout the soil profile and presented high rock contents as described in the Official Soil Description (Soil Survey Staff et al., n.d.). Total AWC from 0-to-60 cm depth for fields in Area 2 were higher (0.130 m m^{-1}) compared to Area 1

(0.113 m m^{-1}) except for Field 5 (0.137 m m^{-1}) where the loess silt deposition was substantial. Among the different soil water potential points used to assess water retention, we found water content held at -100 kPa to be most correlated to SOM (r = 0.64; Table 4.2) compared to other potential points, which was also found in the study undertaken in Chazy, NY (Chapter 6). Bulk density was high in Field 16, exceeding 1.69 g cm^{-3} at the soil depth of 30 cm or more (Appendix B), which is the value that can reduce the root growth of soybean by 50 % (Rosolem and Takahashi, 1998). Also, root density was observed to be 20% of the maximum density with a ρ_b of 1.67 g cm^{-3} at the 30- to-61 cm depth while the maximum root density was found at a ρ_b of 1.32 g cm^{-3} (Grimes et al., 1975). In general, the ρ_b values of the top 30 cm were below 1.50 g cm^{-3} apart from Field 19, where the ρ_b at 15-to-30 cm was as high as 1.70 g cm^{-3} in certain soil sampling locations. Water stable aggregation was only measured in the top 0-to-15 cm in the deep sample set and showed the highest correlation to Al content (r = 0.74) compared to SOM (r = 0.33) and clay (r = 0.25). It has been identified that Al to play a major role in aggregation and SOM accumulation in tropical low activity clays by the Al substitution in goethite and hematite that increase specific surface area (Barthès et al., 2008). This makes the use of WSA to assess the effects of management feasible only within a small area where the levels of Al are relatively stable. Soil pH in the shallow sample set showed similar ranges in both of the areas but there were some samples with values lower than 6.0 (Table 4.1). The CV of the pH values within each field ranged from 1.14 to 10.98 % and shows the possibility of localized effects of acidic soil conditions to crop growth (Cambardella et al., 1994). Soil pH was higher in lower horizons in Field 22, whereas this was not observed in Field 1 and the lowest pH was found at the 60-to-90 cm depth. Phosphorus concentrations were higher in Area 1 where there were manure

applications and also in Field 26 but the levels were within a “medium” index according to the Animal Feeding

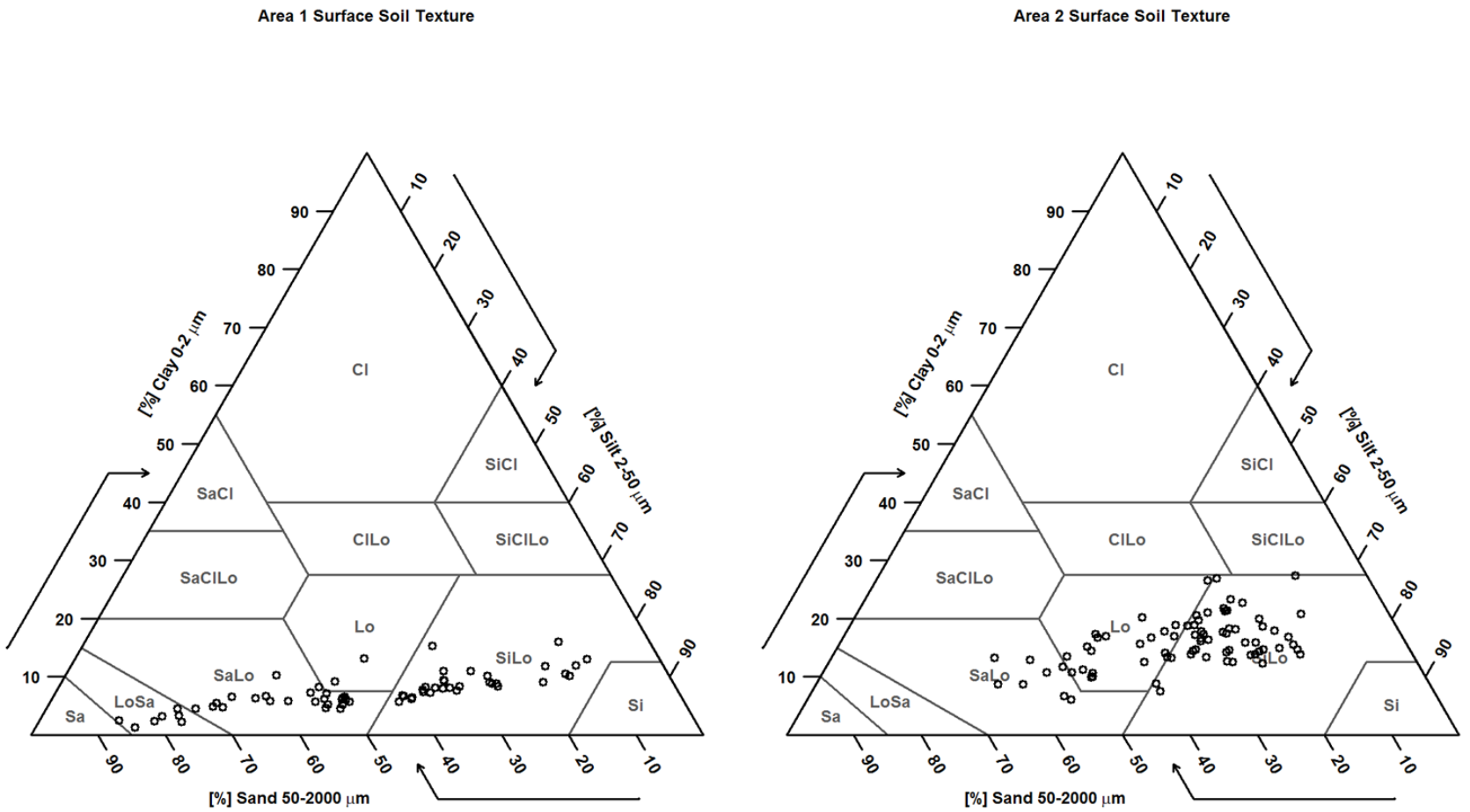

Figure 4.2 Soil textural classess of the surface horizon (0-to-15 cm depth) collected in Area 1 (Coastal Plain province) and Area 2 (Piedmont province).

Table 4.1 Summary statistics of the shallow sample set (0-to-15 cm depth)

	Area 1 (n = 65) †			Area 2 (n =78)		
	Mean	Range	CV (%) ‡	Mean	Range	CV (%)
SOM (%) §	1.11	0.300-2.20	37.1	1.54	0.40-3.30	41.9
pH	6.32	5.40-7.10	5.30	6.36	5.10-7.20	7.10
P (ppm)	81.1	13.0-226	68.8	59.5	7.00-269	93.9
K (ppm)	130	58.0-269	33.3	135	45.0-343	58.1
Mg (ppm)	120	67.0-193	25.6	204	77.0-409	35.2
Ca (ppm)	685	322-1542	37.8	1148	334-2298	35.9
Al (ppm)	806	472-1232	21.7	762	505-1076	15.9
CEC (cmol kg-1)	4.51	2.10-9.90	37.3	7.69	3.60-14.3	33.9
Clay (%)	7.43	1.27-16.1	40.1	15.8	6.09-27.4	27.7
Silt (%)	46.3	11.9-76.2	36.5	50.9	24.3-69.4	20.8
Sand (%)	46.3	10.8-85.6	41.7	33.3	10.7-64.2	38.4
ρb (g cm-3) §	1.29	0.900-1.63	11.2	1.13	0.886-1.52	13.2
- 10 kPa (m m-1)	0.243	0.115-0.355	25.5	0.335	0.203-0.475	17.0
- 33 kPa (m m-1)	0.173	0.0680-0.313	33.5	0.25	0.137-0.370	20.8
-100 kPa (m m-1)	0.127	0.0550-0.239	28.9	0.205	0.121-0.315	21.8
-1500 kPa (m m-1)	0.0502	0.0170-0.115	37.1	0.11	0.0580-0.194	23.9
AWC (m m-1)	0.193	0.0970-0.280	24.2	0.225	0.133-0.309	17.1

† Area 1= the Coastal Plain province; Area 2 = the Piedmont province

‡ CV = coefficient of variation

§ SOM = soil organic matter; CEC = cation exchange capacity; ρ_b = dry bulk density; -10 kPa = water content at -10 kPa; -33 kPa = water content at -33 kPa; -100kPa = water content at -100 kPa; -1500 kPa = water content at -1500 kPa; AWC = available water capacity

Table 4.2 Spearman's rank correlation coefficients of the shallow sample set (n = 142).

	SOM†	pH	P	K	Mg	Ca	Al	CEC	Clay	Silt	Sand	BD	-10 kPa	-33 kPa	-100 kPa	-1500 kPa	AWC
SOM	1.00																
pH	na	1.00															
P	na	na	1.00														
K	0.17 *	na	0.37 ***	1.00													
Mg	0.47 ***	0.46 ***	-0.25 **	na	1.00												
Ca	0.66 ***	0.47 ***	na	na	0.68 ***	1.00											
Al	0.34 ***	-0.19 *	0.36 ***	0.52 ***	na	na	1.00										
CEC	0.66 ***	0.28 ***	na	na	0.63 ***	0.84 ***	0.20 *	1.00									
Clay	0.45 ***	na	-0.24 **	na	0.63 ***	0.55 ***	na	0.60 ***	1.00								
Silt	0.53 ***	na	na	0.24 **	0.23 **	0.49 ***	0.43 ***	0.55 ***	0.43 ***	1.00							
Sand	-0.58 ***	na	0.19 *	-0.20 *	-0.42 ***	-0.60 ***	-0.34 ***	-0.65 ***	-0.71 ***	-0.92 ***	1.00						
ρ_b	-0.18 *	na	na	na	-0.32 ***	-0.30 ***	na	-0.36 ***	-0.36 ***	na	0.24 **	1.00					
-10 kPa	0.58 ***	na	-0.24 **	na	0.63 ***	0.65 ***	na	0.74 ***	0.79 ***	0.68 ***	-0.83 ***	-0.46 ***	1.00				
-33 kPa	0.59 ***	na	na	na	0.53 ***	0.62 ***	0.29 ***	0.70 ***	0.83 ***	0.69 ***	-0.86 ***	-0.43 ***	0.89 ***	1.00			
-100 kPa	0.64 ***	na	na	na	0.68 ***	0.67 ***	0.24 **	0.76 ***	0.89 ***	0.60 ***	-0.79 ***	-0.42 ***	0.93 ***	0.91 ***	1.00		
-1500 kPa	0.57 ***	na	-0.22 **	na	0.69 ***	0.68 ***	na	0.75 ***	0.92 ***	0.49 ***	-0.73 ***	-0.52 ***	0.89 ***	0.90 ***	0.95 ***	1.00	
AWC	0.48 ***	0.17 *	-0.23 **	na	0.47 ***	0.51 ***	na	0.59 ***	0.50 ***	0.72 ***	-0.74 ***	-0.34 ***	0.89 ***	0.70 ***	0.72 ***	0.61 ***	1.00

† SOM = soil organic matter; CEC = cation exchange capacity; ρ_b = dry bulk density; -10 kPa = water content at -10 kPa; -33 kPa = water content at -33 kPa; -100kPa = water content at -100 kPa; -1500 kPa = water content at -1500 kPa; AWC = available water capacity

Operations strategy (Beegle and Sharpley, 1999). The majority of lower CEC (< 5.0 cmol kg^{-1}) were found in Area 1 and higher CEC (> 10.0 cmol kg^{-1}) were found in Area 2. The Spearman's rank correlation for the shallow sample set showed CEC to have a higher correlation to SOM (ρ = 0.66; Table 4.2) compared to clay content (ρ = 0.45). Surface SOM was higher in Area 2 (mean = 1.54 %; Table 4.1) compared to Area 1 (mean = 1.11 %) and the highest correlation was found with Ca and CEC (ρ = 0.66; Table 4.2) and showed a negative correlation to sand content (ρ = -0.58).

The elevation of the Area 1 fields ranged from 15.6 to 28.0 m and from 98.6 to 283 m in the Area 2 fields (Appendix C). Within-field elevation differences in each field ranged from 2.6 to 7.1 m in Area 1 and from 6 to 19 m in Area 2.

4.5.2 Exploratory data analysis of the proximal sensors

We determined Pearson correlations between each proximal sensor and the measured soil properties. For the optic sensors, we found substantially less correlation to SOM compared to the past study conducted by Kweon and Maxton (2013).They showed the R^2 for SOM of 0.79 across six fields in central Kansas in Mollisols, which could be caused by the soil conditions at the time of measurement including soil moisture and temperature (Kweon et al., 2013; Kweon and Maxton, 2013), as well as the presence of a wider range of soil colors from multiple soil orders. The correlations of the optic sensors to other soil properties were also low (r < 0.51; Table 4.3). In Area 1, P and Al values showed the highest correlations, whereas CEC, Ca, and clay content had the highest correlations in Area 2. In the Coastal Plain province, Page (1974) assessed the predictability of SOM using a color-difference meter and showed an r of 0.89 but the samples were air-dried and ground thoroughly in a laboratory.

Table 4.3 Pearson correlation coefficients of optic sensor values and measured soil properties of the shallow sample set and the surface increment (0-to-15 cm depth) of the deep sample set.

		Red†		IR		OMR	
Area	n	*r*	property	*r*	Property	*r*	Property
1	100	-0.51/-0.47/-0.27	P/Ca/pH	-0.44/0.30/0.30	Al/Sand/SOM	0.43/0.31/-0.28	P/Ca/Al
2	108	-0.49/-0.49/0.40	CEC/Ca/Clay	-0.47/0.44/-0.40	Ca/Clay/CEC	-0.36/-0.35/-0.33	BD/P/K
All	208	-0.41/-0.35/-0.23	Ca/CEC/SOM	0.30/-0.27/-0.21	Clay/Ca/CEC	-0.22/0.20/0.19	Al/Ca/CEC

† Red = reflectance at 660 nm; IR = infra-red reflectance at 940 nm; OMR = the ratio of Red and IR

Table 4.4 Pearson correlation coefficients of apparent electrical conductivity values and measured soil samples of the deep sample set.

		ECsh†		ECdp		ECR	
ID	n	*r*	Property	*r*	Property	*r*	Property
1	9	0.95	-100 kPa‡ at 30-45cm	0.92	-100 kPa at 0-90cm	-0.85	K at 0-15 cm
		0.94	Clay at 30-45cm	0.90	Clay at 30-45cm	-0.83	Al at 0-15 cm
		0.92	-100 kPa at 0-90 cm	0.90	-100 kPa at 60-90cm	0.79	ρ_b at 45- 60cm
2	9	0.93	-100 kPa at 15-30cm	0.89	-10 kPa at 15-30cm	0.89	Silt at 45-60cm
		0.90	Ca at 0-15 cm	0.88	-100 kPa at 15-30 cm	0.87	Silt at 0-90cm
		0.89	Clay at 0-60cm	0.88	Clay at 0-60cm	-0.87	Sand at 45-60cm
5	9	0.92	-100 kPa at 0-60cm	0.91	-100 kPa at 0-90cm	-0.80	Clay at 0-15cm
		0.84	-100 kPa at 0-90 cm	0.89	-100 kPa at 0-60 cm	0.65	Silt at 0-30cm
		0.83	-1500 kPa at 45-60cm	0.84	-100 kPa at 45-60cm	0.65	Silt at 0-45cm
6	9	0.91	Ave Clay at 0-45cm	0.89	Clay at 30-45 cm	-0.81	Clay at 15-30cm
		0.89	Ave Clay at 45-60cm	0.89	Clay at 0-60cm	-0.73	-1500 kPa at 15-30cm
		0.89	Clay at 30-45cm	0.88	-1500 kPa at 30-45 cm	-0.71	Clay 0-30cm
14	6	-0.97	P at 0-15 cm	-0.98	P at 0-15 cm	0.83	AWC at 45-60cm
		0.96	-1500 kPa at 30-45cm	0.96	ρ_b at 0-30 cm	-0.80	K at 0-15 cm
		-0.96	Al at 0-15 cm	0.94	-33 kPa at 30-45cm	0.76	AWC 0-60cm
16	6	0.99	-10 kPa at 0-15cm	0.95	Clay at 30-45cm	0.72	-1500 kPa at 30-45cm
		0.98	CEC at 0-15cm	0.92	-100 kPa at 30-45cm	0.72	Clay at 30-45cm
		0.96	-100 kPa 0-45cm	0.91	-1500 kPa at 30-45cm	-0.68	P at 0-15 cm
19	6	0.92	-10 kPa at 30-45cm	0.98	-10 kPa at 30-45cm	0.88	SOM at 0-15 cm
		0.91	Clay at 15-30cm	0.94	Clay at 0-60cm	-0.83	ρ_b at 0-15cm
		0.87	Soil moisture at 0-15cm	0.93	Clay at 15-30cm	-0.79	Pen 0-30cm
21	6	0.89	Soil moisture at 0-15 cm	0.93	Soil moisture at 0-15cm	-0.86	-1500 kPa at 45-60cm
		-0.69	Pen at 30-45cm	-0.85	Pen at 30-45cm	-0.78	Clay at 45-60cm
		0.67	Mg at 0-15 cm	0.69	Mg at 0-15 cm	-0.76	-33 kPa at 45-60cm
22	6	0.86	-33 kPa at 0-15cm	0.79	-1500 kPa at 0-15 cm	0.89	Pen at 0-45cm
		0.84	-1500 kPa at 0-15 cm	-0.78	K at 0-15 cm	-0.85	Al at 0-15 cm
		0.84	-33 kPa at 45-60cm	0.70	-33 kPa at 45-60cm	-0.85	Pen 0-30cm

† ECsh = shallow apparent electrical conductivity at 0-to-45 cm; ECdp = deep apparent electrical conductivity at 0-to-90 cm; ECR = the ratio of ECsh and ECdp

‡ SOM = soil organic matter; CEC = cation exchange capacity; ρ_b = dry bulk density; -10 kPa = water content at -10 kPa; -33 kPa = water content at -33 kPa; -100kPa = water content at -100 kPa; -1500 kPa = water content at -1500 kPa; AWC = available water capacity

Table 4.4 (Continued)

		ECsh†		ECdp		ECR	
ID	n	*r*	Property	*r*	Property	*r*	Property
Area 1	36	0.78	Clay at 0-45cm	0.76	Clay at 30-45 cm	0.77	-100 kPa at 60-90cm
		0.78	Clay at 45-60cm	0.76	-1500 kPa at 30-45cm	-0.76	Sand at 60-90cm
		0.77	Clay at 15-30cm	0.75	Clay 0-60cm	0.74	Clay at 60-90cm
Area 2	30	0.70	Clay at 30-45 cm	0.65	-1500 kPa at 30-45 cm	0.33	ρ_b at 30-45 cm
		0.67	Clay at 0-45 cm	0.64	-33 kPa at 30-45 cm	0.32	Ca at 0-15 cm
		-0.67	Sand 0-45 cm	0.63	Ca at 0-15 cm	0.32	ρ_b at 0-45 cm
All	66	0.77	Clay at 0-45 cm	0.75	-1500 kPa at 30-45 cm	0.31	Silt at 30-45 cm
		0.75	Clay at 30-45 cm	0.71	Clay at 30-45 cm	0.29	Silt at 0-45 cm
		0.74	-1500 kPa‡ at 30-45 cm	0.70	-33 kPa at 30-45 cm	0.28	Silt at 15-30 cm

† ECsh = shallow apparent electrical conductivity at 0-to-45 cm; ECdp = deep apparent electrical conductivity at 0-to-90 cm; ECR = the ratio of ECsh and ECdp

‡ SOM = soil organic matter; CEC = cation exchange capacity; ρ_b = dry bulk density; -10 kPa = water content at -10 kPa; -33 kPa = water content at -33 kPa; -100kPa = water content at -100 kPa; -1500 kPa = water content at -1500 kPa; AWC = available water capacity

For ECa sensor values, we assessed the correlations to soil properties using only the deep sample set (n = 66) across nine fields. For an individual field, the correlations, *r*, were relatively high and ranged from 0.72 to 0.99 (Table 4.4). Both ECsh and ECdp sensors were most correlated to clay content and water retention values measured at various pressure points (Table 4.4). The latter were also in turn well correlated to clay content ($0.79 < \rho < 0.92$; Table 4.2) and SOM ($0.57 < \rho < 0.64$). We found good correlations of ECR for Area 1 especially with soil texture and water retention at -100 kPa ($r \approx 0.77$; Table 4.4), but this was not observed for Area 2 with the best correlation with ρ_b at 30-to-45 cm ($r = 0.33$). However, ECR was well correlated in individual fields in Area 2, but the properties correlated to the signal varied widely (Table 4.4). Past research has shown ECa signals to correlate the most to soil properties such as clay, P ($r = 0.50$ and -0.58, respectively; Johnson et al., 2001), silt ($r = -0.74$; Jung et al., 2005), and CEC ($r = 0.88$) within a single field (Moral et al., 2010). In addition, Mueller et al. (2003) assessed the correlations of ECsh to soil properties across four fields with loess silt layers overlaying limestone residuum and found the correlations to elevation ($r = -0.65$) and clay ($r = 0.63$) for ECsh, and Ca ($r = 0.58$) and Mg ($r = 0.48$) for ECdp. However, in all of these studies, samples were only collected for the surface soil (≈ 0-to-15 cm) and they did not match the depths of the reach of the sensor signals.

In this study, on-the-go pH measurement was possible in Area 1 and only selected fields in Area 2 due to high surface rock contents. We undertook a linear regression analysis to predict the laboratory measured soil pH but found low predictability ($R^2 = 0.28$; Figure 4.3). Soil samples were taken near the optic sensor tract but the on-the-go pH measurement was only taken every 30 m, thus the distance between the soil samples and the on-the-go pH measurement can be as far as 15 m.

4.5.3 In-situ soil conditions on the proximal sensors

In this study, we undertook proximal sensor data collection in both spring and fall seasons in Field 1. The effects of in-situ soil conditions are known to affect sensor values for both ECa (Corwin and Lesch, 2005) and optic sensors (Kweon and Maxton, 2013). The effects of in-situ soil moisture is an area of interest for optic sensors and related reflectance spectroscopy for predicting various soil properties (Gobrecht et al., 2014; Minasny et al., 2011). The in-situ soil moisture content was higher in the spring (mean = 19.1 %) compared to the fall (mean = 15.8 %) in the shallow sample set. We assessed the predictability of the fall measured values using the spring measured values and found the predictability to be in order of ECdp > ECsh > IR > Red (Figure 4.4). We found the repeatability of the optic sensors extremely low in this study. This field was irrigated during the growing season and the different water management inside and outside the pivot may be affecting the sensor values. We assessed the predictability outside the pivot and found the R^2 for ECsh and ECdp improved to 0.49 and 0.71 respectively. We therefore recommend carrying out ECa measurements in irrigated fields in the spring and also mean center the ECa sensor values in each field due to the large distance between the regression line and the 1:1 line. Overall, the repeatability of the optic sensor values were low.

4.5.4 Random Forest

Random Forest was used to utilize proximal sensor information for prediction of within-field variable Yp within the region, and to identify area-specific yield constraints. The

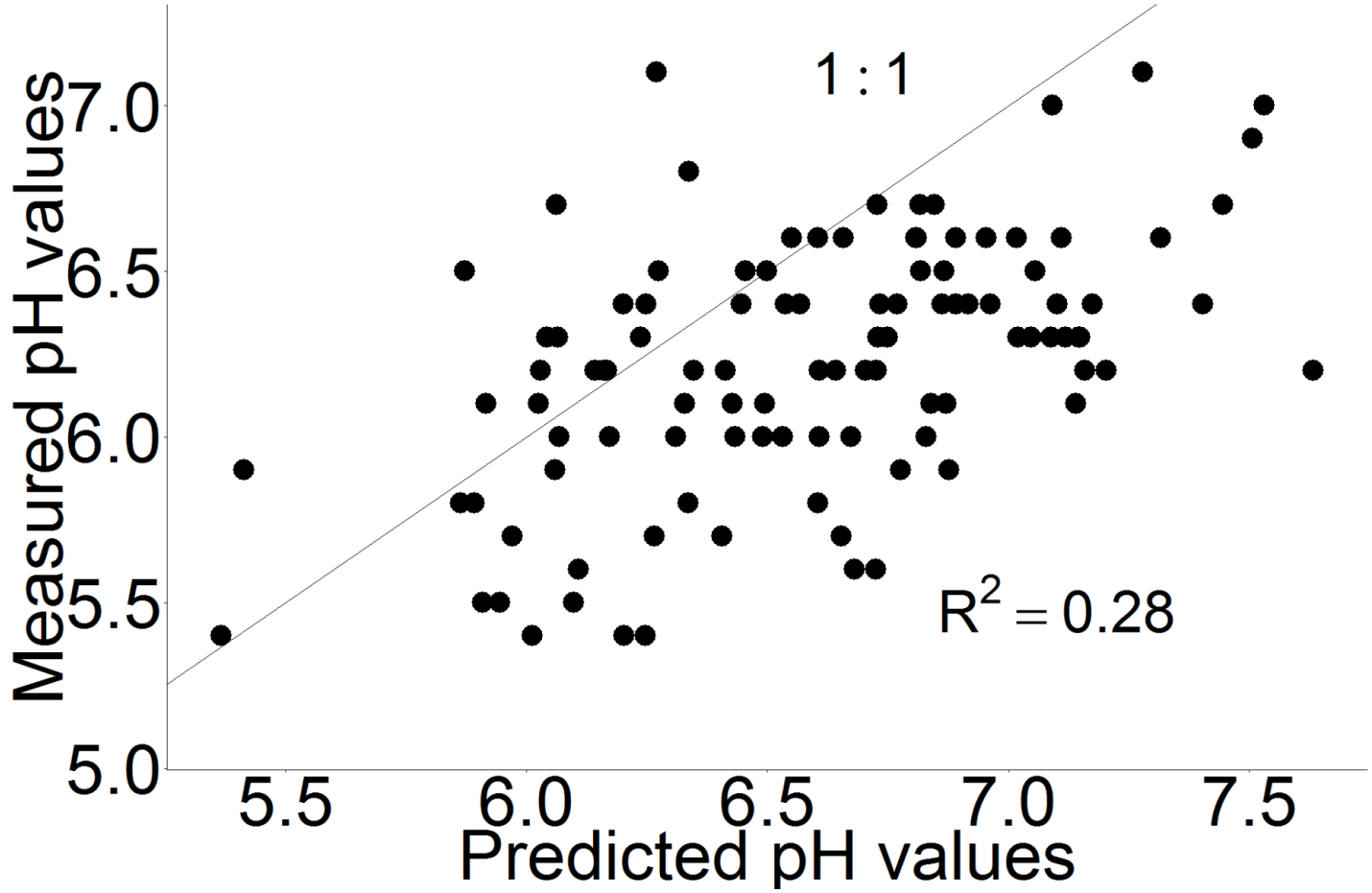

Figure 4.3 Scatter plot of soil pH. The predicated pH values were obtained using the on-the-go pH sensor and the measured values by a bench-top pH meter in a laboratory.

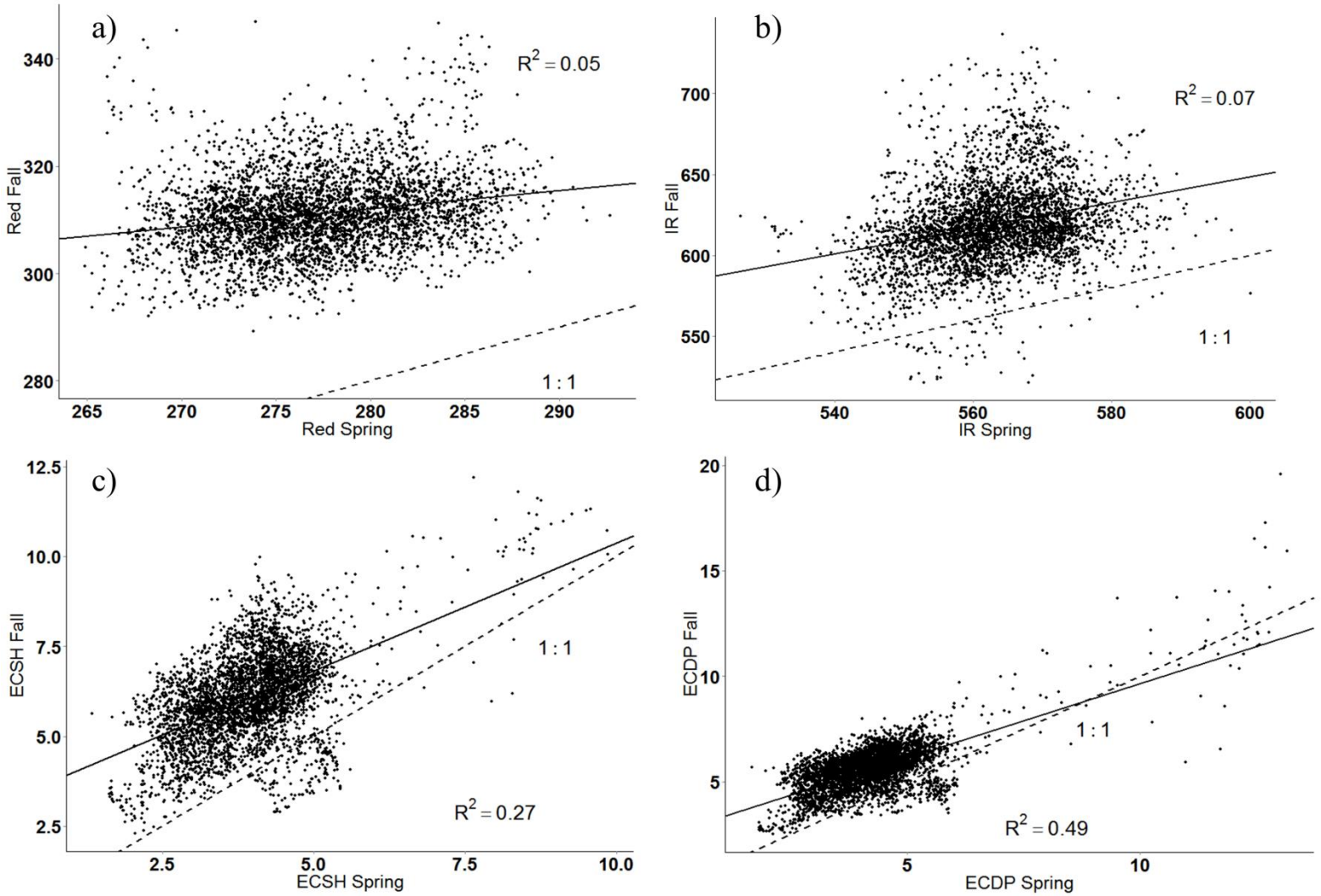

Figure 4.4 Scatter plot of proximal sensing information collected in the spring and the fall of 2014. a) red sensor (Red; 660 nm), b) infra-red (IR; 940 nm), c) shallow apparent electrical conductivity at 0-to-45 cm (ECSH), and d) deep apparent electrical conductivity at 0-to-90 cm (ECDP)

Table 4.5 Statistical results of the Random Forest models for the whole region (Global rainfed maize), the Coastal Plain province (Area 1 rainfed maize), and the Piedmont province (Area 2 rainfed maize) with topographical properties as predictors or the combination of topographical properties and proximal sensing information.

	Topographical properties only					Topographical properties + proximal sensing				
ID	% variance explained	Important variables	%IncMSE†	R^2	RMSE	% variance explained	Important variables	%IncMSE	R^2	RMSE
Global rainfeld maize	60.4	ASP‡	176	0.59	0.845	75.5	ECR	129	0.74	0.697
		TWI	131				ASP	120		
		Slope	108				OMR	109		
Area 1 rainfed maize	61	ASP	172	0.60	0.885	80	ECR	160	0.78	0.68
		TWI	113				pH	110		
		Slope	97.9				TWI	102		
Area 2 rainfed maize	69.8	ASP	119	0.68	0.67	78.1	ASP	108	0.76	0.59
		Slope	102				Slope	96.2		
		PROF	88.9				ECR	72.2		

† %IncMSE = percentage increase in mean square error by dropping one of the important variables; RMSE = root mean square error of prediction

‡ ASP = aspect; TWI = topographic wetness index; PROF = profile curvature; ECR = ratio of shallow and deep apparent electrical conductivity ; OMR = ratio of the reflectance at 660 and 940 nm

predictability was higher when the proximal sensor information was combined with topographical information (Table 4.5), and the improvement in predictability was the highest for Area 1. We found ECR, pH, and TWI as the three most important predictors in the area. The improvement was the smallest for Area 2 and the two most important predictors were identical (ASP and SLOPE; Table 4.5). Therefore, we selected models with proximal sensing information that were specific to a physiographic province as the most appropriate prediction model. Due to the nature of empirical models, the use of the RF model outside of the physiographic province is challenging (Kaul et al., 2005). It will be necessary to verify the similarity of the new field to the existing physiographic province specific model before use.

We undertook full-site cross validation and found the predictability to be higher for Area 1 ($0.07 < R^2 < 0.28$) compared to Area 2 ($0.01 < R^2 < 0.28$). We found the $\bar{d}$ value in Area 1 of 0.249 compared to 0.272 in Area 2, which indicated that the RF models perform better in Area 1. We found the lowest $\bar{d}$ value in Field 3 and the highest in Field 20 (Table 6) among the cross-validated fields. Van Uffelen et al. (1997) stated that the threshold d value for adequate pattern similarity depends upon the scope of the work. In this study, we undertook a visual assessment and determined the d value of 0.3 to be adequate, which was within the range used by Gandah et al. (2000). In independent-site validation, good predictability was observed for Field 6 and 7, but not in Field 5 (Table 4.6; Figure 4.5), where the loess silt deposition was more significant compared to the other two fields. Field 4 shared the same soil series, Metapeake and Mattapex as Field 5, but the predictability was low. There are several possibilities for the mismatch of the model predicted values including: i) errors in the measured yield data (i.e. error associated with yield monitor; Arslan and Colvin, 2002), ii) errors associated with the elevation and proximal sensor measurement (Erskine et al., 2007; Kweon et al., 2013), and iii) yield constraints that

Table 4.6 Statistical results of Random Forest model prediction of yield potential using full-site cross validation and independent site validation.

	Cross-validation			Independent-validation			
Field ID	R^2	RMSE†	$\bar{d}$	n	R^2	RMSE	$\bar{d}$
1	0.17	1.09	0.257	na	na	na	na
2	0.07	1.21	0.281	na	na	na	na
3	0.28	0.974	0.209	na	na	na	na
4	0.15	1.11	0.250	na	na	na	na
5	na	na	na	1	0.02	1.32	0.292
6	na	na	na	2	0.29	0.957	0.196
7	na	na	na	2	0.22	1.03	0.214
14	na	na	na	1	0.17	1.08	0.209
15	na	na	na	1	<0.01	1.45	0.292
16	na	na	na	2	<0.01	1.43	0.289
17	na	na	na	2	0.01	1.35	0.227
18	na	na	na	1	0.11	1.16	0.197
19	na	na	na	1	0.01	1.34	0.269
20	0.02	1.3	0.300	na	na	na	na
21	0.03	1.28	0.278	na	na	na	na
22	0.09	1.18	0.280	na	na	na	na
23	<0.01	1.38	0.300	na	na	na	na
24	0.28	0.97	0.200	na	na	na	na
25	na	na	na	2	0.34	1.28	0.261
26	na	na	na	1	0.10	1.17	0.214

† RMSE = root mean square error of prediction; $\bar{d}$ = mean pattern similarity distance; n = number of yield data available

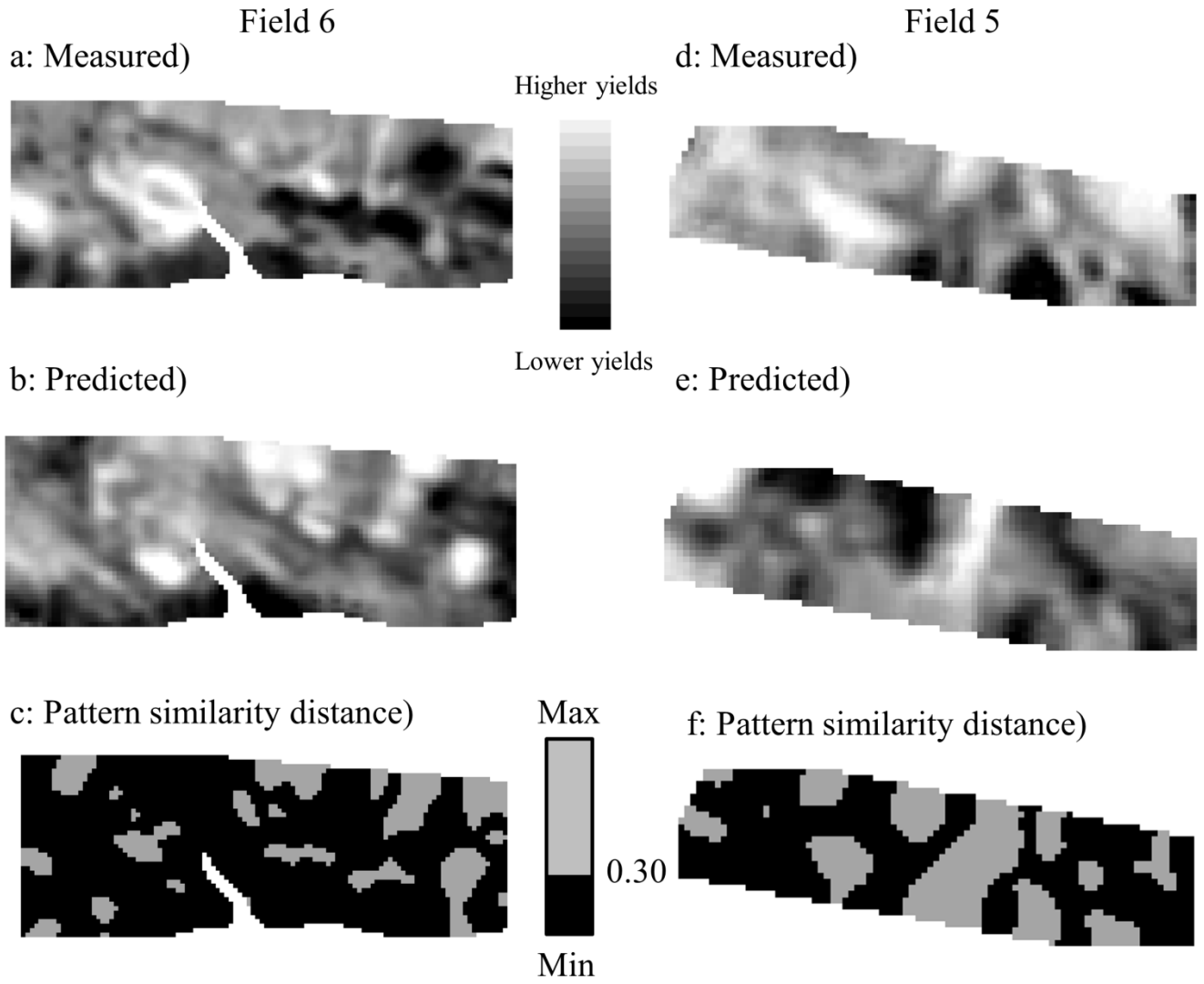

Figure 4.5 Maps of a and d) measured yield data, b and e) stdPCA score values predicted using Random Forest models, and c and f) pattern similarity distance indicating the spatial pattern similarity of the measured yield data and predicted stdPCA scores.

were not represented by the predictors (van Ittersum et al., 2013). The potential errors associated with the yield monitor were minimized using the Yield Editor software and removal of the 18 m field borders. The errors associated with the elevation and proximal sensing data collection were partly addressed by using the mean centered data for the proximal sensor values and calculating both the ECR and OMR on the mean centered data. Nevertheless, three dimensional variations in soil moisture in each field could affect the results (Corwin and Lesch, 2003). The yield constraints rather than the factors accounted by the predictors including pests and diseases could affect the results and could be significant because of the small number of yield data used as validation data for independent-site validation.

In Area 2, the $\bar{d}$ value ranged from 0.197 to 0.292 in independent-site validation and the predictability was variable depending on the field. This could be partly affected by the low correlation between ECR and particular soil properties in the area (r = 0.33; Table 4.4) compared to Area 1 (r = 0.77). Among the deep sampled fields, Field 14 had the lowest $\bar{d}$ value ($\bar{d}$ = 0.209) and found ECR was most correlated to AWC at 45-to-60 cm depth (Table 4.4). The other deep sampled fields had a $\bar{d}$ values of more than 0.250 and the ECR values were correlated to factors such as water retention at -1500 kPa, SOM, and soil penetration resistance.

The identification of location specific yield constraint is important for management decision making because they provide information on i) whether the constraint is ameliorable (Oliver et al., 2010), ii) selection of specific management (e.g. soil amelioration, hybrid selection), and iii) the estimates of the cost of management (Plant, 2001; Chapter 3). In this study, we attempted to utilize various methods for the identification of yield constraints. First, stdPCA allowed determination of the spatial pattern of Yp under moisture-limited growing seasons (Area 1) or Yp of all seasons (Area 2). The extraction of the spatial patterns of moisture-limited Yp is

important because the relationship between soil and topographic properties are known to be more significant under extreme weather conditions (Jiang and Thelen, 2004; Kravchenko and Bullock, 2000). Second, the verification of the correlation between each proximal sensor value and measured soil properties infers the underlying yield constraints within the areas where the RF models perform adequately. In Area 1, ECR was the most important predictor in the RF model (Table 4.5) and it was correlated to the change in soil texture in the 60-to-90 cm depth (Table 4.4). Therefore, the within-field Yp patterns under moisture-limited growing seasons are presumed to be related to subsoil texture change along with surface pH variation, and TWI (Table 4.5). The measurement of subsoil textural change is extremely costly and difficult and the proximal sensing method allowed for a rapid estimation with high spatial resolution. Subsoil texture and TWI are not changeable by management but site-specific crop management such as fertilizer and crop variety selection could utilize this information (Chen et al., 2011). In Area 2, the identification of underlying yield constraints appears be more challenging due to a lower RF model fit compared to Area 1 (Table 4.5). Area 2 included four areas of distinctly different soil characteristics and had higher variations in various measured soil properties including clay, SOM, pH, P and K (Table 4.1). The correlations between ECR and measured soil properties were also lower compared to Area 1 (Table 4.4) with ρ_b at 30-to-45 cm the most correlated property (r = 0.33). Topographical properties (ASP and SLOPE) were more important predictors in this area (Table 4.5) over the proximal sensor information. In both of the areas, the calculation of the d-value allowed assignment and visual presentation where the RF model successfully predicted the spatial pattern of Yp (Figure 4.5c, 4.5f, 4.6c, and 4.6f) and this required the availability of at least one year of yield data. This is significantly less challenging compared to collecting a few

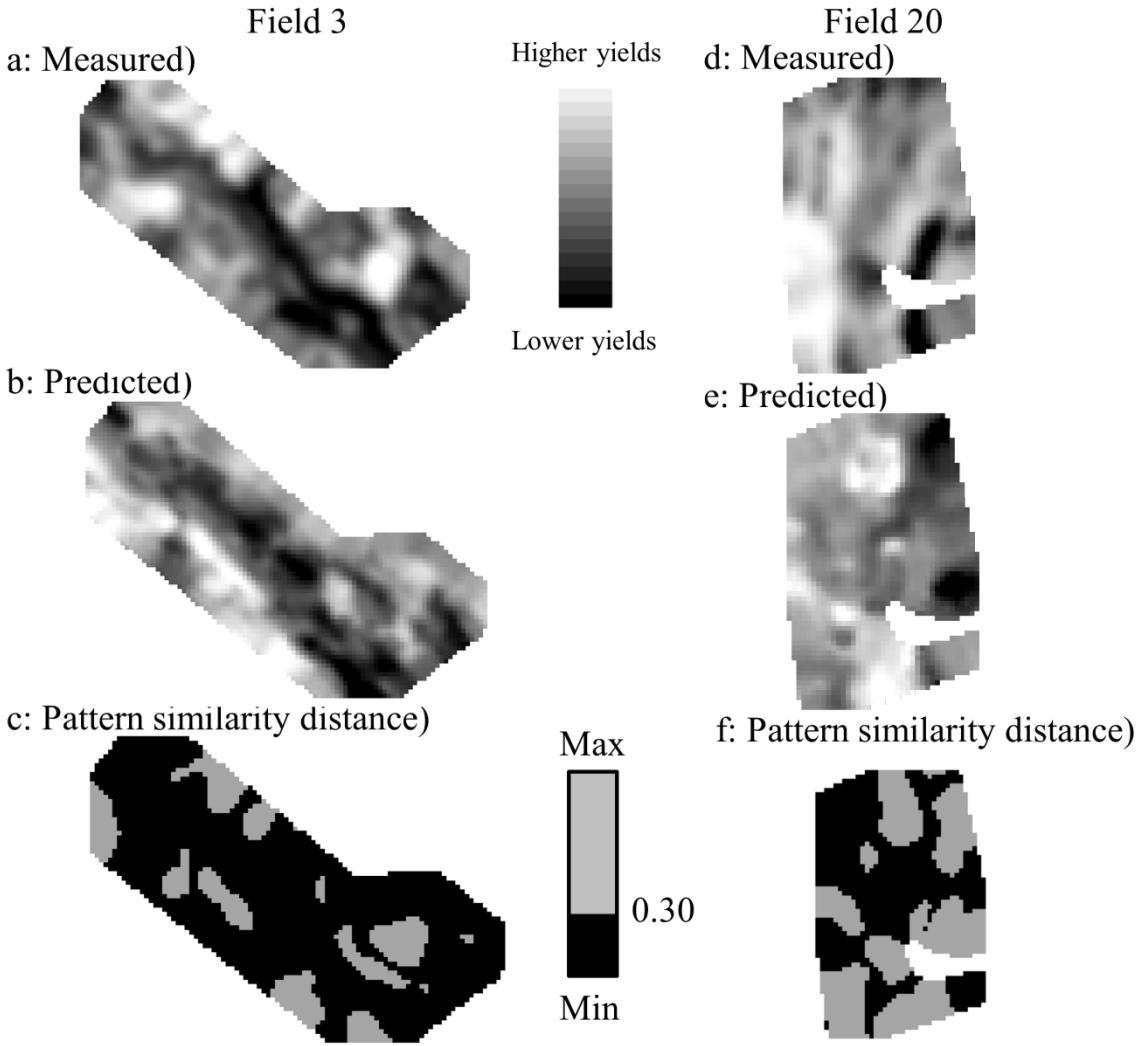

Figure 4.6 Maps of a and d) standardized principal component analysis (stdPCA) derived score values from measured yield data, b and e) stdPCA score values predicted using Random Forest models, and c and f) pattern similarity distance indicating the spatial pattern similarity of the measured and predicted stdPCA scores.

years of yield data of a single crop. The locations where the d-values were low had a successful RF model estimation of the spatial pattern of Yp and the areas where the d-values were high require further assessment. In order to identify specific yield constraints, we can undertake judgment sampling in the locations where the d-values were high and the measured yields were low. It is also important to utilize farmer knowledge (Calviño et al., 2003; Oliver et al., 2010) to confirm the measured low yields are not affected by pests and diseases. The collected soil samples need to be assessed for a wide array of soil properties including soil physical, chemical, and biological properties that allow for an identification of soil constraints other than conventional soil nutrient values (Idowu et al., 2008; Sojka et al., 2003).

4.6 CONCLUSIONS

In this study, we utilized a combination of proximal sensor information and machine learning to predict the spatial patterns of yield potential (Yp) and assessment of site-specific yield constraints in multiple fields across the Mid-Atlantic US. This study revealed i) higher predictability of Yp on the Coastal Plain province using a Random Forest (RF) model and ii) subsoil textural variation, surface pH, and topographic wetness index (TWI) as the regional yield constraints on the Coastal Plain. The predictability of Yp and also the correlations of particular soil properties to the EC ratio (ECR) was lower in the Piedmont province. In both provinces, we found the use of pattern similarity distance (d) a useful tool to identify areas where properties represented by the RF model control the spatial patterns of Yp. Also, areas with higher d-values can be used as targeted soil sampling locations for further assessment of potential edaphic yield constraints. Finally, the inclusion of more fields for calibration of RF has a potential in further

improving the predictability of within-field spatial patterns of Yp given the quality of data are assured.

4.7 ACKNOWLEDGEMENTS

I am grateful to all the participated growers in the Mid-Atlantic region of the US. Also, Mike Twining, Dave Yannacci, Chris Atkinson, Jamie Kimbles , Nelson Oberholzer, and David Hertel from Willard Agri-Service for their support in providing data and field work. I also would like to thank John Dantinne for the invaluable suggestions on data analysis.

CHAPTER 5: MULTIPLE SOIL PROPERTY PREDICTIONS USING VNIR SPECTROSCOPY AND SPECTRAL STRATIFICATION

5.1 ABSTRACT

The assessment of soil properties beyond soil nutrient concentrations is important, and the incorporation of soil physical and biological tests has been encouraged, but low-cost rapid assessment options still need to be explored. In this study, we determined effective uses of visible and near-infrared reflectance spectroscopy (VNIRS; 350-2500 nm) to predict 10 soil physical, chemical, and biological properties of 1977 oven-dried soil samples from the US Northeast and the Midwest. Partial least squares regression (PLSR) and boosted regression trees (BRT) were validated using the full-site cross validation method. Water stable aggregation (WSA) and pH were best predicted among the soil properties (R^2 > 0.50; ratio of performance to interquartile range > 2) using BRT but the overall predictability was low in general. We created stratified prediction models using k-means clustering, which improved the model predictability for sand, silt, soil organic matter (SOM), and permanganate oxidizable carbon in a sample cluster with largely Mollisols samples (n = 115). In this cluster, we found the absence of Fe-oxide related absorption bands, which is known to be an important property for controlling WSA and SOM levels. Therefore, k-means clustering was a promising approach to stratify soil samples according to their clay content and mineralogy, and allowed for direct quantifications using VNIRS.

5.2 KEY WORDS

k-means clustering; PLSR; Soil assessment; Soil health; VNIR spectroscopy

5.3 INTRODUCTION

Soil nutrient assessment has been widely adopted in agricultural management, especially when building a soil fertility program (Cornell University Cooperative Extension, 2012; Havlin, 2005). In the last 20 years, the efforts of introducing other disciplines of edaphology, especially soil physical and biological components in soil management, have led to the development of the soil quality (SQ) and soil health (SH) concepts (Karlen et al., 1997; Sojka and Upchurch, 1999) and their assessment frameworks (Andrews et al., 2004; Idowu et al., 2008). Mausbach and Tugel (1995) defined SQ as "the capacity of a specific kind of soil to function within natural or managed ecosystem boundaries, to sustain plant and animal productivity, maintain or enhance water and air quality, and support human health and habitation" and SH as " the ability of the soil to perform according to its potential". The assessment has been undertaken using selected SH indicators (Gugino et al., 2009; Idowu et al., 2008), which include soil texture, water stable aggregation (WSA), available water capacity (AWC), penetration resistance, soil organic matter (SOM), permanganate oxidizable carbon (POXC), potentially mineralizable nitrogen, root health, pH, and plant available soil nutrients (Gugino et al., 2009). The measured values are then interpreted based on soil texture-specific scoring functions to identify SH constraints (Andrews et al., 2004; Moebius-Clune, 2010). Having successfully introduced and measured these indicators in a soil testing laboratory as a

routine assessment, we are interested in exploring an alternative rapid and cost-effective approach for soil measurements particularly for soil physical and biological properties.

Visible and near-infrared reflectance spectroscopy (VNIRS) is a nondestructive technique to analyze soil properties through multivariate statistical or data mining techniques by relating spectra to directly-measured soil property values (Dalal and Henry, 1986). The spectra contain the information on the overtones of stretching and bending vibrations in molecular bonds such as C–C, C–H, N–H, and O–H in the mid-infrared region as well as the absorption characteristics caused by the size, the shape, the color, and the arrangement of soil particles (Gobrecht et al., 2014; Stenberg and Viscarra Rossel, 2010). Visible and near-infrared reflectance spectroscopy has been utilized to predict various soil properties including soil carbon (C), nitrogen (N), texture (Chang et al., 2001), potentially mineralizable N (Morón and Cozzolino, 2002; Reeves and Van Kessel, 1999), water stable aggregation (WSA; Cañasveras et al., 2010), heavy metals, micronutrients (Cozzolino and Morón, 2003; Kooistra et al., 2001; Udelhoven et al., 2003), and soil enzymatic activities (Dick et al., 2013). In a previous study, Kinoshita et al. (2012) have assessed the feasibility of using VNIRS for predicting various soil properties selected as SH indicators within one soil order (i.e. Ultisols) in Kenya. They showed successful predictions ($R^2 > 0.8$; ratio of performance to deviation (RPD) > 2.0) of soil properties including SOM, POXC, water content at field capacity (θ_{fc}), water content at permanent wilting point (θ_{pwp}), and cation exchange capacity (CEC). However in a soil testing laboratory, soil samples from various soil orders are submitted and the prediction needs to be made for those samples. Past studies have shown calibrations of VNIRS prediction models challenging when the samples come from a large geographical area (Sankey et al., 2008). For soil organic carbon (SOC), the predictability was substantially lower ($0.39 \leq R^2 \leq 0.85$) when independent sites

validation was performed across three soil orders compared to non-independent site validation ($0.74 \leq R^2 \leq 0.86$; Brown et al., 2005). The primary reasons for the low predictability were suggested to be caused by i) the prediction models calibrated based on indirect correlations of the spectra to the property of interest, or ii) variable soil constituents especially Fe-oxide and secondary clay minerals masking the absorption features of the property of interest. Several approaches have been proposed to improve the predictability for a heterogeneous sample set such as i) building a large global spectral library with diverse soil sample set (Brown et al., 2006), ii) creating a small local spectral library (Wetterlind et al., 2010), or iii) including a small number of local samples in a global spectral library (i.e. spiking; Guerrero et al., 2010, 2014). However, the use of the latter two approaches are challenging in a soil testing laboratory where soil samples can be submitted from any geographical locations, and the information about the origin of each sample is scarce. Furthermore, SH assessment undertakes soil property measurements that may not be included in large spectral libraries, which have been primarily focusing on CEC, SOC, soil texture, exchangeable cations, and pH (Brown et al., 2006; ICRAF and ISRIC - World Soil Information, 2010).

Spectral stratification is a technique to group soil samples according to the spectral characteristics, and allows reducing the within-group variation of spectral variations. This technique has been recently applied to classify topsoil soil samples in Australia into corresponding soil orders using VNIRS through canonical variate analyses (Viscarra Rossel and Webster, 2011), while Knadel et al. (2013) assessed topsoil samples from Denmark and identified historical glacial boundaries through the combination of principal component analysis (PCA), k-means clustering, and kriging. In soil classifications, important soil properties have been identified as clay mineralogical constituents and organic characteristics (Viscarra Rossel

and Webster, 2011). However, there is lack of evidence whether unsupervised classifications of the spectra can subsequently improve model predictability of soil properties, and also what soil constituents are responsible for the classifications.

Therefore, this study was conducted utilizing soil samples collected from agricultural and woodland sites in the US Northeast (NE) and the Midwest (MW). The main objectives of this study were to i) assess the feasibility of VNIRS in predicting SH indicators, ii) assess any improvements in predictability using sample stratification through k-means clustering, and iii) determine the mechanisms of predictability improvements using k-means clustering.

5.4 MATERIALS AND METHODS

5.4.1 Sampling design

We compiled soil property data of the samples that were submitted to the Cornell Soil Health Analysis Laboratory (Ithaca, NY) from the NE and the MW in 2011 and 2012. Only the samples collected from agricultural or woodland sites were retained for this study. Additionally, soil samples with SOM contents of more than 10% were removed, which were mainly collected from soils under high tunnels and other non-agricultural sites. In total, 1671 samples were retained from the NE and 306 samples from the MW. The soil samples were collected from an individual field with 10 subsamples composited from the 0-to-15 cm depth using a spade (Gugino et al., 2009).

5.4.2 Soil assessment

Soil texture was assessed using a rapid quantitative method developed by Kettler et al. (2001). The soil sample was dispersed with 3% sodium hexametaphosphate ($(NaPO_3)_n$) and a combination of sieving and sedimentation steps was used to separate size fractions. Water stable aggregation was assessed using a rainfall simulator (1997) that allows the soil particles to slake under known rainfall energy, applying 2.5 J of energy for 300 s on aggregates (0.25–2 mm) placed on a 0.25-mm mesh sieve. The fraction of soil aggregates remaining on the sieve, corrected for stones >0.25 mm, was regarded as the percent WSA after drying at 105°C (Gugino et al., 2009). Water content at field capacity (θ_{fc}) and permanent wilting point (θ_{pwp}) were assessed gravimetrically. Saturated soil samples were equilibrated to pressures of –10 kPa and –1500 kPa on two ceramic high pressure plates (Topp et al., 1993). The difference in gravimetric moisture content between these two pressure points was considered as the available water capacity. The SOM was analyzed by loss on ignition in a muffle furnace at 500°C for two hours. Permanganate oxidizable carbon was measured by oxidizing samples with dilute potassium permanganate ($KMnO_4$) followed by measuring absorbance at 550 nm using a hand-held colorimeter (Hach, Loveland, CO). A 1:1 soil and water suspension was assessed for pH. Soil nutrients, including P, K, Mg, Fe, Mn, and Zn were extracted using Modified Morgan, an ammonium acetate solution, buffered at pH 4.8 (McIntosh, 1969), and analyzed using inductively coupled plasma optical emission spectrometer.

5.4.3 Visible and near-infrared reflectance spectroscopy

The reflectance of soil samples was determined in both the visible and near-infrared spectral regions between the spectra of 350 nm and 2500 nm at 10 nm spectral resolutions using a Fieldspec Pro hyperspectral sensor (Analytical Spectral Devices, INC., Boulder, CO).

Air-dried and 2-mm sieved samples were placed in a 4 cm diameter optical quality Petri dish, and spectral reflectance was collected through the glass bottom at a constant angle (55 degrees from horizontal) from a distance of 4 cm, in an enclosed box.

5.4.4 Visible and near-infrared reflectance spectroscopy modeling

Predictive models were constructed using partial least squares regression (PLSR) and Boosted Regression Trees (BRT), which have been applied in VNIRS studies for predicting various soil properties, primarily soil C (Brown et al., 2006; Kinoshita et al., 2012). For PLSR, Unscrambler 10.3 (CAMO software, Oslo, Norway, 2013) was used for the analysis. First, we transferred the reflectance data using the first-order derivatives of the Savitsky-Golay transformation function, which is recognized as a suitable method for VNIRS applications because it removes noise from the data while enhancing the absorption features of the original spectra with minimum distortion (Ruffin and King, 1999). The method relates two variables, X (spectral readings) and Y (measured soil property values), by a linear multivariate model. Orthogonal and weighted linear combinations of the spectral readings are used for predicting each Y variable. Partial least squares regression is suited to handle data with strong collinearity in the (X) variables, which are usually more numerous than the observations (Y) that they predict. The normality of Y variables are ideal to ensure the normality and homoscedasticity of

the residuals; however, we avoided transforming the Y variables due to unknown departures from the physical relationships between the values of the soil property of interest and spectral absorption characteristics (Gobrecht et al., 2014). The selection of the number of factors to include was critical to avoid overfitting or underfitting of the model to the data. Overfitting of calibration models reduces their ability to predict soil property of new unknown samples. In this study, additional factors were only added if they reduced the total residual Y-variance at cross validation by the percentage equivalent to the number of factors (Kinoshita et al., 2012). Also, the maximum number of factors was restricted to 25 to prevent overfitting. Nevertheless, the linearity between the spectral absorption characteristics and the properties of interest do not hold in particulates like soils because of light scattering (Ciani et al., 2005), and also interactions among various soil constituents interfere.

The non-linear distribution of the Y-variable and the non-linear relationship between the spectral absorption characteristics and the soil properties have led to applying non-linear prediction models. Boosted regression trees is an extension of randomized classification and regression trees (Breiman et al., 1983) that allows for bagging of a selected number of tree sizes (Breiman, 1996). It combines boosting that adds reweighted calibration datasets, which grants higher weights where there were higher residuals. Detailed information on the method can be found in Brown et al. (2006). The *dismo* package (Hijmans et al., 2013) in the R computing environment was used with the tree size decided by fitting a range of tree sizes to the model and determining the lowest prediction error in 10-fold cross validation with a maximum of 3000 trees in each sample set.

5.4.5 Prediction accuracy

Geographically independent validation is important when assessing the predictability for unknown future samples (Brown et al., 2005). We validated the PLSR and BRT models of each sample set using a full-site cross validation technique where each fold of the cross validation contains all the samples from a single farm or a woodland site (Kinoshita et al., 2012). The largest sample size from a single field was 189 samples, which was 9.56% of the entire sample set in this study.

The accuracy of the PLSR and BRT models for each SQ indicator was evaluated using the coefficient of determination (R^2) of the measured and the model-predicted values. Root mean square error of prediction (RMSE) was also determined, which is a measurement of accuracy calculated as the differences between model predicted values and measured values:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_{pred.i} - Y_{meas.i})^2}{n-1}} \quad (1)$$

The RMSE is a combination of bias and residual variance, the latter being the deviation from the 1:1 line of predicted values against measured values (Bellon-Maurel et al., 2010):

$$RMSE = \sqrt{Bias^2 + MSEc} \quad (2)$$

where MSEc is the mean square error of prediction corrected for bias.

Bias can be calculated as:

$$Bias = \bar{Y}_{pred.} - \bar{Y}_{meas.} \quad (3)$$

which is the difference between the mean of predicted values and the mean of measured values. Bias can be caused by systematic errors due to instrumental and methodological errors while the residual variance relates to random errors that can be improved by averaging the

predicted values (Bellon-Maurel et al., 2010). The decomposition of RMSE is therefore important to understanding the origin of errors.

The ratio of performance to interquartile distance (RPIQ) is a useful performance measure for comparing prediction models from different datasets even when the observations are not normally distributed (Bellon-Maurel et al., 2010):

$$RPIQ = \frac{(Q3-Q1)}{RMSE} \quad (4)$$

where Q1 and Q3 are the 1st and the 3rd quartiles respectively.

The ratio of performance to deviation was also derived, which is a widely used measure of the predictability of a VNIRS-based model (Williams and Sobering, 1993). Although this measure is known to be susceptible to compare among different datasets when they are non-normally distributed (Bellon-Maurel et al., 2010), but included to be compatible with past studies:

$$RPD = \frac{SD}{RMSE} \quad (5)$$

where SD is the standard deviation of the dataset.

5.4.6 K-means clustering

In order to stratify the soil sample set according to the spectral similarity, we performed k-means clustering (MacQueen, 1967) using the R statistical environment (R Core Team, 2014). It is an unsupervised machine learning technique where the algorithm starts with k random clusters and then moves samples in order to maximize the distance between groups while minimizing the intra-group variability.

In each cluster, we undertook an exploratory data analysis of both the spectra as well as the measured soil property values. The spectra were assessed using the PCA, and the loading values were evaluated to assess important wavelengths for each cluster. The mean difference of measured soil property values between each cluster was assessed using a combination of Welch's t-test and the Games-Howell post hoc test, which are suitable for unbalanced and unequal variance comparisons.

The predictability of each soil property values was assessed for each cluster using PLSR and the model statistics were calculated. In order to verify whether the PLSR model prediction relies on direct or indirect correlations between the spectra and the property of interest, we assessed the Pearson correlation coefficients of the regression coefficients in each cluster.

5.5 RESULTS AND DISCUSSION

5.5.1 Prediction of soil properties

In all of the SH indicator predictions, BRT was better compared to PLSR (Table 5.1), which was analogous to the findings by Brown et al. (2006). The non-linear nature of the relationship between the soil reflectance spectra and the soil property of interest due to light scattering (Ciani et al., 2005) make linear prediction models inferior to non-linear prediction models. Also, our heterogeneous sample set makes the spectral absorption features not unique to a particular soil property of interest, and therefore, the incorporation of multiple, high-level interactions by BRT improved the predictability compared to PLSR (Brown et al., 2006). The

Table 5.1 Statistical results of cross-validation based on visible and near-infrared reflectance spectroscopy with partial least squares regression and boosted regression tree for the global sample set (n = 1977), Cluster 1 (n = 506), Cluster 2 (n = 115), and Cluster 3 (n = 1356).

	PLSR						BRT					
	F‡	R^2	RMSE	Bias	RPIQ	RPD	R^2 ‡	RMSE	Bias	RPIQ	RPD	RI (%)
WSA, % †	8	0.47	18.1	0.586	2.28 §	1.36	0.51	17.2	0.598	**2.40 §**	1.43	5.26
AWC, $m^3\ m^{-3}$	15	0.23	0.0500	< 0.0001	1.47	1.19	0.32	0.0490	0.0033	**1.50**	1.21	2.04
Θ_{fc}	15	0.31	0.0669	-0.0008	1.45	1.30	0.41	0.0665	0.0038	**1.46**	1.31	0.69
Θ_{pwp}	10	0.42	0.0283	-0.0007	1.64	1.44	0.55	0.0273	0.0003	**1.70**	1.49	3.66
Sand, %	12	0.53	13.4	-0.0930	1.82	1.43	0.56	12.7	-0.131	**1.92**	1.51	5.49
Clay, %	5	0.48	4.51	0.0448	1.85	1.38	0.51	4.37	0.299	**1.91**	1.43	3.24
Silt, %	12	0.45	12.5	<0.0001	1.63	1.27	0.45	11.77	0.144	**1.73**	1.35	6.13
SOM , $g\ kg^{-1}$	7	0.47	12.1	0.0029	1.76	1.48	0.60	11.3	0.0332	**1.89**	1.58	7.39
POXC, $mg\ kg^{-1}$	9	0.52	122	1.21	1.93	1.50	0.56	121	0.894	**1.95**	1.51	1.04
pH	11	0.42	0.505	-0.0138	2.10	1.48	0.61	0.470	-0.0268	**2.26**	1.60	7.62

†WSA = water stable aggregation; AWC = available water capacity calculated by the difference between -10 kPa and -1500 kPa; Θ_{fc} = water content at -10 kPa; Θ_{pwp} = water content at -1500 kPa; SOM = soil organic matter; POXC = permanganate oxidizable carbon.
‡ F = number of factors used in PLSR; R^2 = coefficient of determination; RMSE = root mean square error of cross validation; RPIQ = ratio of performance to interquartile distance.
§ Bold types show the highest RPIQ value for the property of interest among the sample clusters.

predictability exceeded the RPIQ of 2 for WSA and pH for both PLSR and BRT and the relative improvements were high for pH, SOM, silt, and sand (Table 5.1). Nevertheless, the predictability was much lower compared to similar studies undertaken previously using one soil order where they gained the RPD above 2 for clay, SOM, POXC, CEC, Ca, and Cu (Kinoshita et al., 2012). The nature of our full-site validation method also made the apparent predictability lower. In this study, we utilized a soil sample set that existed in a soil analysis laboratory where some of the soil samples were collected from a single site. The largest sample size from a single field was 189 samples (Oxyaquic Hapludalfs), and the predictability of some of the cross validation fold may be poor if the spectral characteristics are not represented in the calibration model. We also assessed the model predictability of PLSR using leave-one-out cross validation and found the R^2 to be improved from 0.47 to 0.60 and the RMSE from 12.1 to 11.3 for SOM.

5.5.2 K-means clustering

We performed the k-means cluster analysis on a range of cluster sizes and identified three clusters as optimum through a visual assessment of the scree plot of a within-group sum of squares. In total, 506 out of 1977 samples were classified as Cluster1 and all but two samples were from the NE (Figure 5.1). The largest number of samples came from Glossic Hapludalfs and Typic Dystrudepts, which were affected by Pleistocene glaciers. There were 115 samples classified as Cluster 2, and all except four samples were from the MW (Figure 5.1). Those four NE soil samples had high SOM contents (mean = 70.00 g kg^{-1}) compared to the global mean SOM (mean = 39.15 g kg^{-1}), which were mainly from vegetable production fields. The most dominant soil type of this cluster was Aquic Argiudolls, which is known to have dark color and are rich in bases. Cluster 3 contained 1356 samples of which 193 samples were from the MW

Figure 5.1 Biplot of PC1 and PC2 with each sample categorized in the respective geographical region and k-means cluster. MW and NE are the Midwestern and Northeastern USA, respectively.

and 1163 samples from the NE. Within the MW sample set the dominant soil types were Typic Hapludalfs, Alfic Haplorthods, and Alfic Fragiorthods. For the NE, the dominant soil types were Oxyaquic Hapludalfs, Glossic Hapludalfs, and Typic Dystrudepts. There were significant differences in some of the measured soil property values. Water stable aggregation and sand content were the highest in Cluster 1 and the lowest in Cluster 2 (Table 5.2). Conversely, Cluster 2 had the highest mean values for AWC, θ_{fc}, θ_{pwp}, clay and silt content. Permanganate oxidizable carbon and pH were the highest in Cluster 3.

We performed the k-means cluster analysis on the first derivative reflectance but present the mean reflectance of each cluster on the raw reflectance basis (Figure 5.2). The shape of the mean spectra was distinctly different for Cluster 2 but similar for Cluster 1 and 3. Cluster 2 showed the lowest reflectance in the visible range (435-750 nm) and the near infrared (750-1134 nm). There were common absorption bands at 1414, 1913, and 2203 nm, which are features of molecular H_2O in the samples and O-H stretch vibration (Bishop et al., 1994; Hunt, 1977; Oinuma and Hayashi, 1965).

Principal component analysis on the spectra showed that the first three PCs explained 75 to 88 % of the total variance. Their loadings (Figure 5.3a, b and c) revealed less distinct contributions of absorption bands related to Fe-oxide for Cluster 2 (515-535 nm ; Sherman and Waite, 1985; Scheinost et al., 1998), but higher loadings for smectite and other 2:1 clays at 1893 nm in PC2 (1893 nm; Figure 5.3b; Bishop et al., 1994). In this study, Fe-oxide was not directly measured but Fe was measured by extracting the samples with the Modified Morgan method. It is known that the main source of plant available Fe is derived from Fe-oxide in most soils (Schwertmann, 1991). The mean Fe content was higher for Clusters 1 (9.82 ppm) and 3 (6.36 ppm) compared to Cluster 2 (1.48 ppm). The majority of Cluster 2 samples were associated with

Table 5.2 Descriptive statistics of the measured soil properties for each cluster. Statistical differences among the means were assessed using Welch's t-test combined with the Games-Howell post hoc test.

	Cluster	N ‡	Min	Max	Mean	SD	Skewness	Kurtosis
WSA, % †	1	506	3.45	97.55	51.02 a §	24.44	0.07	-1.21
	2	115	13.78	94.25	32.16 c	14.94	1.43	2.56
	3	1356	3.13	97.13	45.70 b	24.74	0.38	-1.05
AWC, $m^3 m^{-3}$	1	506	0.02	0.39	0.18 a	0.06	0.09	-0.23
	2	115	0.07	0.29	0.19 a	0.04	-0.27	0.41
	3	1356	0.01	0.55	0.16 b	0.06	0.51	1.89
Θ_{fc}	1	506	0.03	0.50	0.27 b	0.09	-0.16	-0.45
	2	115	0.12	0.43	0.32 a	0.06	-0.67	0.90
	3	1356	0.02	0.67	0.27 b	0.09	0.08	0.70
Θ_{pwp}	1	506	0.01	0.26	0.09 c	0.04	0.93	1.53
	2	115	0.04	0.21	0.13 a	0.03	0.07	1.11
	3	1356	0.01	0.26	0.10 b	0.04	0.28	0.77
Sand, %	1	506	7.23	96.15	50.31 a	17.56	0.20	-0.46
	2	115	1.57	84.04	20.81 c	18.97	1.05	0.65
	3	1356	1.85	97.00	45.31 b	18.26	0.58	-0.24
Clay, %	1	506	1.57	41.36	8.87 c	5.77	2.55	8.93
	2	115	4.07	41.33	20.71 a	5.31	0.17	2.34
	3	1356	0.28	54.07	11.48 b	5.68	0.94	3.74
Silt, %	1	506	2.28	84.69	40.82 c	15.16	-0.18	-0.37
	2	115	7.31	76.90	58.47 a	16.70	-0.68	-0.42
	3	1356	1.71	84.18	43.22 b	15.34	-0.38	-0.27
SOM , $g kg^{-1}$	1	506	4.70	99.30	39.70 a	17.80	0.52	0.10
	2	115	15.5	80.70	39.70 a	13.30	0.91	0.32
	3	1356	3.00	97.50	38.90 a	18.20	0.61	0.24
POXC, $mg kg^{-1}$	1	506	17.77	1083.76	466.97 b	182.02	0.21	0.16
	2	115	208.20	1062.97	460.28 ab	163.99	1.02	1.47
	3	1356	24.66	1144.00	490.53 a	183.89	0.29	0.02
pH	1	506	4.31	8.45	6.20 b	0.66	-0.02	0.18
	2	115	5.08	7.67	6.32 b	0.53	0.70	-0.18
	3	1356	3.18	8.75	6.51 a	0.78	-0.03	-0.48

†WSA = water stable aggregation; AWC = available water capacity calculated by the difference between -10 kPa and -1500 kPa; Θ_{fc} = water content at -10 kPa; Θ_{pwp} = water content at -1500 kPa; SOM = soil organic matter; POXC = permanganate oxidizable carbon.

‡ N = number of samples in each cluster; Min = minimum; Max = maximum; SD = standard deviation

§ Means of each yield level followed by an identical lowercase alphabet are not significantly different at the $\alpha = 0.05$.

Figure 5.2 Mean raw reflectance spectra of the three clusters defined using k-means clustering.

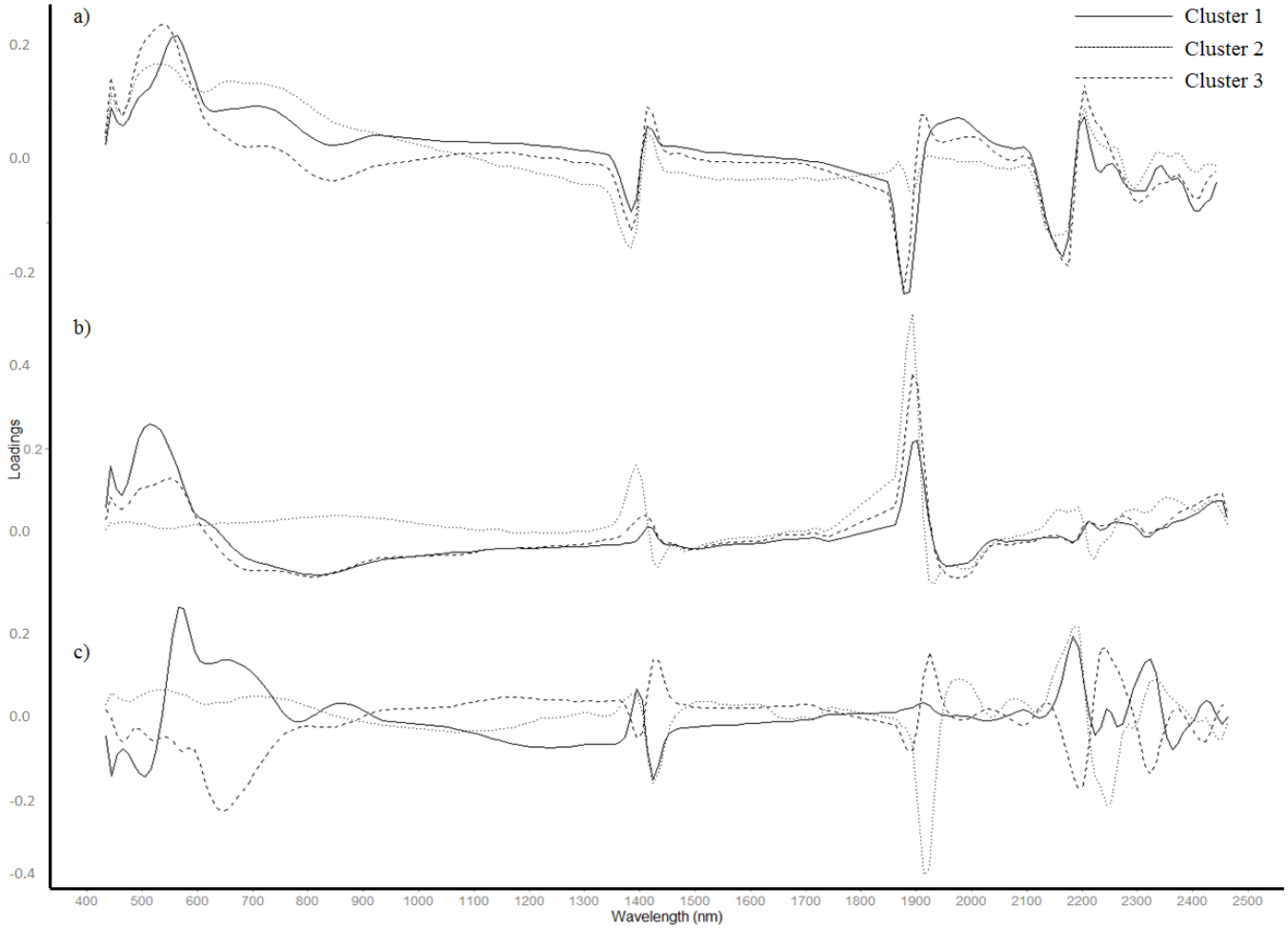

Figure 5.3 Loading plots of each k-means cluster showing contributing wavelengths.

Mollisols that are known to contain marginal amounts of Fe-oxide and are dominated with 2:1 clay such as illite, smectite, and vermiculite (Allen and Fanning, 1983; Brady and Weil, 2008). Stoner and Baumgardner (1981) described the importance of Fe-oxides on spectral reflectance, and showed a characteristic concave shape of the low Fe-oxide containing Mollisols in reflectance between 500 and 1300 nm, analogues to the mean reflectance for Cluster 2 (Figure 5.2).

5.5.3 Prediction models for each cluster

The k-means cluster analysis improved the predictability for different soil properties in each cluster. The improvements were observed for AWC, θ_{fc}, and SOM for Cluster 1; sand, silt, SOM, and POXC for Cluster 2; and pH for Cluster 3 (Table 5.3).

Brown et al. (2005) argued that the low predictability of VNIRS models in a heterogeneous soil sample set is due to indirect correlation of the spectra to the soil property of interest and/or the existence of Fe-oxide and clay mineral absorption bands masking important spectral features. The k-means clustering grouped the soil samples by soil orders, in particular for Cluster 2, which helps to reduce the variation of clay mineralogy (Brady and Weil, 2008). Cluster 2 had the highest RPIQ values for POXC, SOM, sand, and silt (RPIQ > 2) compared to other clusters (Table 5.3).

We assessed the regression coefficients of each PLSR model to identify important wavelengths used for each prediction. In Cluster 2, the wavelengths at 2123 nm (organic matter, cellulose, glucan, pectin, and polysaccharides), 2263 nm (biotite/gibbsite), and 2293 nm (Fe-smectite) were important for predicting POXC (Ben-Dor et al., 1997; Clark et al., 1990; Post and Noble, 1993). The correlation between the regression coefficients of the PLSR model for POXC

Table 5.3 Statistical results of cross-validation based on visible and near-infrared reflectance spectroscopy with partial least squares regression for Cluster 1 (n = 506), Cluster 2 (n = 115), and Cluster 3 (n = 1356).

	Cluster 1					Cluster 2						Cluster 3					
	R^2	RMSE	Bias	RPIQ	RPD	F	R^2	RMSE	Bias	RPIQ	RPD	F	R^2	RMSE	Bias	RPIQ	RPD
WSA % †	0.43	19.2	0.198	2.25	1.27	3	0.38	13.4	1.22	1.36	1.11	8	0.46	18.8	0.315	2.19	1.32
AWC $m^3\ m^{-3}$	0.20	0.0578	< 0.0001	**1.61**	1.04	9	0.29	0.0401	0.0007	1.16	1.00	11	0.21	0.0467	0.0012	1.37	1.28
Θ_{fc}	0.22	0.0789	< 0.0001	**1.66**	1.14	9	0.22	0.0547	0.0008	1.19	1.10	13	0.29	0.0634	0.0012	1.31	1.42
Θ_{pwp}	0.37	0.0309	< 0.0001	1.61	1.29	8	0.24	0.0271	0.0005	1.18	1.11	13	0.40	0.0264	< 0.0001	1.45	1.52
Sand %	0.16	16.8	0.182	1.36	1.05	9	0.60	12.8	2.56	**2.16**	1.48	11	0.56	13.4	-0.543	1.72	1.36
Clay %	0.39	4.21	-0.158	1.23	1.37	9	0.53	4.11	0.404	1.11	1.29	5	0.39	4.51	0.0485	1.76	1.26
Silt %	0.14	14.8	-0.135	1.33	1.02	3	0.49	12.0	-0.465	**2.40**	1.39	11	0.51	12.6	0.449	1.50	1.22
SOM $g\ kg^{-1}$	0.49	12.7	0.0171	**1.92**	1.40	8	0.61	8.13	-0.114	**2.54**	1.64	11	0.52	10.9	0.0320	1.85	1.67
POXC $mg\ kg^{-1}$	0.49	125	1.83	1.92	1.46	11	0.75	92.2	11.1	**2.67**	1.78	10	0.56	121	-0.808	1.92	1.52
pH	0.33	0.506	-0.0044	1.70	1.30	15	0.52	0.393	0.0400	1.74	1.35	11	0.50	0.47	-0.0212	**2.47**	1.66

†WSA = water stable aggregation; AWC = available water capacity calculated by the difference between -10 kPa and -1500 kPa; Θ_{fc} = water content at -10 kPa; Θ_{pwp} = water content at -1500 kPa; SOM = soil organic matter; POXC = permanganate oxidizable carbon.
‡ F = number of factors used in PLSR; R^2 = coefficient of determination; RMSE = root mean square error of cross validation; RPIQ = ratio of performance to interquartile distance.
§ Bold types show the highest RPIQ value for the property of interest among the sample clusters.

Table 5.4 Pearson correlation coefficients of the regression coefficients in the partial least squares regression for each cluster.

	WSA	AWC	Θ_{fc}	Θ_{pwp}	Sand	Clay	Silt	SOM	POXC	pH
		0.48 ‡	**0.68**	**0.51**	ns	ns	ns	**0.94**	**0.78**	**-0.17**
		ns	ns	ns	**0.23**	ns	**-0.95**	ns	ns	ns
WSA †	1.00	**0.26**	**0.27**	**0.27**	ns	ns	ns	**0.52**	**0.46**	**-0.47**
			0.89	**0.48**	**-0.53**	**0.16**	**0.63**	**0.56**	**0.60**	ns
			0.93	**0.36**	**-0.65**	**0.57**	**0.15**	**0.35**	**0.34**	ns
AWC		1.00	**0.95**	**0.75**	**-0.71**	**0.20**	**0.70**	**0.80**	**0.72**	ns
				0.79	**-0.62**	**0.32**	**0.43**	**0.80**	**0.78**	ns
				0.66	**-0.79**	**0.73**	**0.16**	**0.47**	**0.48**	ns
Θ_{fc}			1.00	**0.90**	**-0.76**	**0.24**	**0.74**	**0.81**	**0.65**	ns
					-0.74	**0.57**	**0.18**	**0.69**	**0.65**	ns
					-0.61	**0.67**	ns	**0.52**	**0.47**	ns
Θ_{pwp}				1.00	**-0.77**	**0.30**	**0.73**	**0.81**	**0.60**	**-0.21**
						-0.61	**-0.61**	**-0.18**	**-0.22**	ns
						-0.70	**-0.32**	**-0.51**	**-0.43**	ns
Sand					1.00	**-0.35**	**-0.98**	**-0.58**	**-0.54**	ns
							0.18	ns	ns	ns
							0.15	**0.58**	**0.61**	ns
Clay						1.00	**0.25**	**0.24**	**0.24**	ns
								ns	-0.12	ns
								ns	ns	ns
Silt							1.00	**0.54**	**0.51**	ns
									0.92	ns
									0.45	ns
SOM								1.00	**0.81**	**-0.31**
										ns
										ns
POXC									1.00	ns
pH										1.00

†WSA = water stable aggregation; AWC = available water capacity calculated by the difference between -10 kPa and -1500 kPa; Θ_{fc} = water content at -10 kPa; Θ_{pwp} = water content at -1500 kPa; SOM = soil organic matter; POXC = permanganate oxidizable carbon.
‡ Bold types are significant for $p < 0.05$ and the non-bold types are significant at $p < 0.1$.

and SOM was relatively low (r = 0.45) compared to Cluster 1 and 3 (r = 0.92 and 0.81; Table 5.4). For SOM, the important wavelengths were at 2243 nm (protein), 2293 nm (Fe-smectite), and 2393 nm (nontronite or protein; Post and Noble, 1993; Fourty et al., 1996; Ben-Dor et al., 1997). The highest correlation of the regression coefficients of POXC was with clay content (r = 0.61; Table 5.4) but the predictability for clay was low (R^2 = 0.53; RPIQ = 1.11; Table 5.3). Past study has shown that the prediction of clay content can have a moderately successful R^2 of around 0.76 compiled from studies between 1986 and 2006 (Stenberg and Viscarra Rossel, 2010). The correlation between the regression coefficients of sand and silt was low (r = -0.32; Table 5.4). For sand, there were numerous peaks in the regression coefficients; however, the most prominent ones were found at 1434 nm (O-H), 2213 nm (Al-OH bend with O-H stretch combinations), 2293 nm (Fe-smectite), and 2393 nm (nontronite or protein; Oinuma and Hayashi, 1965; Clark et al., 1990; Post and Noble, 1993; Ben-Dor et al., 1997). For silt, they were more distinct and identified at 1913 nm (molecular water), 2193 nm (protein), and 2253 nm (biotite) in this cluster (Ben-Dor et al., 1997; Hunt, 1977; Post and Noble, 1993).

In contrast, the predictability for WSA was substantially higher for Clusters 1 and 3 compared to Cluster 2 (Table 5.3). In Cluster 1 and 3, high regression coefficients were found for goethite and SOM related wavelengths in the visible range between 575 and 645 nm (Galvao and Vitorello, 1998; Stenberg et al., 2010) as well as protein and biotite/gibbsite absorption bands in the SWIR range of 2033, 2193, 2253, and 2383 nm (Ben-Dor et al., 1997; Clark, 1999; Clark et al., 1990; Post and Noble, 1993). Also, the correlation between the regression coefficients of WSA and SOM was the highest for Cluster 1 (r = 0.94; Table 5.4) and lower for Cluster 3 (r = 0.52) and 2 (r = 0.11). Duiker et al. (2003) have shown amorphous Fe-oxide to correlate with WSA, measured by the wet-sieving method. Positive correlations between Fe-oxide and SOM

have been reported by numerous studies, including the Ap horizon of Inceptisols and Alfisols, which showed Fe-oxide to be a major contributor to the specific surface area of clay minerals and contribute to both aggregation and higher SOM (Pronk et al., 2011). For Cluster 2, high regression coefficients for SOM were only found in the SWIR range for protein (2193 and 2243 nm) but not in the visible range (Ben-Dor et al., 1997; Fourty et al., 1996). The physical basis of the relationship among Fe-oxide, SOM, and WSA (Duiker et al., 2003; Pronk et al., 2011) as well as the varying spectral importance of the visible spectral range suggest i) the necessity of the presence of Fe-oxide related SOM for the WSA to be predicted by VNIRS and ii) the possible requirement for separate interpretations for WSA levels when Fe-oxide is present. Further study is necessary to assess the feasibility of utilizing VNIRS to assess the relationship among SOM, WSA, and Fe-oxide.

5.5.4 The role of VNIRS in soil health assessment

In the soil science community, the use of the SH concept has been much debated (Karlen et al., 2001; Sojka et al., 2003), but one of the identified issues was the interpretation of the measured values using a global soil scoring function often calibrated for Mollisols or Alfisols that has a larger emphasis on the managed soils for plant productivity to define overall SH index. This also contradicts the definition for SH that a soil should be assessed relative to its potential (Mausback and Tugel, 1995), which is known to vary substantially due to, but not limited by, its characteristics influenced by genetic pedological processes (Carter, 2002; Richter, 1987). We found the unsupervised machine learning technique of k-means clustering on the VNIRS spectral data to be able to stratify soil sample set according to Fe-oxide signatures and soil texture. Past research has also shown spectral signatures of soil mineralogy such as hematite at 490 nm

(Sherman and Waite, 1985) or goethite at 620 nm (Stenberg et al., 2010) as well as in the near-infrared range for smectite, kaolinate, or illite (Knadel et al., 2013). Clay mineralogy and Fe-oxide contents are partly linked with the inhere differences in soil potential (Mausback and Tugel, 1995) because they control soil nutrient and water holding capacity (Brady and Weil, 2008; Hillel, 1980) and aggregation (Duiker et al., 2003; Jozefaciuk and Czachor, 2014). Further research is needed to determine the inherent potential of different soil types and to evaluate the current condition of SH from the measured values. Here, VNIRS can be a promising candidate to rapidly gain information about the clay mineralogy at relative low running cost.

5.6 CONCLUSIONS

We assessed the possibility of using visible and near-infrared reflectance spectroscopy (VNIRS) to i) predict a range of soil physical, chemical, and biological properties utilizing linear or nonlinear multivariate statistical models and ii) spectrally stratify soil samples to improve the model predictability. The application of the simple k-means clustering approach revealed the importance of spectral signatures related to Fe-oxide in sample stratification and its importance in the prediction of water stable aggregation (WSA), soil organic matter (SOM), and permanganate oxidizable carbon (POXC). The predictability was higher for WSA with the presence of Fe-oxide related spectral signatures whereas SOM and POXC were better predicted when it was absent. Past research has identified the importance of Fe-oxide for affecting the spectra of the above three components, and its strong signatures are promising for utilizing VNIRS as a way to approximate inherent soil health (SH) differences. The predictability of the VNIRS for any soil property was low in this study, although it was improved after the removal of

the Fe-oxide effect for POXC, SOM, sand, and silt contents through k-means clustering. In general, a nonlinear prediction model of BRT had higher predictability compared to PLSR, but the predictability was lower compared to other studies partly due to the cross-validation strategy utilized for this study. Further study is necessary to effectively integrate VNIRS in improving the interpretation of the measured SH indicator values incorporating the information regarding clay mineralogical differences.

5.7 ACKNOWLEDGEMENTS

This work was in part supported through a USDA-NIFA Special Grant on Computational Agriculture. I am grateful to Dean Hively for his help on setting up the VNIRS equipment.

CHAPTER 6: QUANTITAITVE SOIL PROFILE-SCALE ASSESSMENT OF THE SUSTAINABILITY OF LONG-TERM MAIZE RESIDUE AND TILLAGE MANAGEMENT

6.1 ABSTRACT

Both surface and subsoil layers can be a significant source of soil moisture and nutrients for crop growth but the changes in subsoil properties due to management are rarely assessed. This study was conducted to determine tillage and residue management effects on the vertical distribution of soil nutrients as well as soil biological and physical properties within an entire rooted profile. We utilized an experiment with 40-year long continuous maize (*Zea mays* L.) cropping under crossed plow-till (PT) vs. no-till (NT) and residue removed (Harv) vs. residue returned (Ret) treatments on a silt loam soil in Chazy, NY. We assessed soil properties (texture, bulk density (BD), water stable aggregation (WSA), available water capacity (AWC), soil organic matter (SOM), permanganate oxidizable carbon, mineralizable carbon, soil protein, pH, and plant available nutrients) at five depth increments to 60 cm depth. We found that the residue and tillage treatments strongly impacted SOM-related properties (marginal R^2 values, R^2m, from 0.66 to 0.91), whereas AWC and unfertilized nutrients were less affected by the treatments ($0.11 < R^2m < 0.65$). NT-Ret showed the highest availability of majority of the measured soil nutrients but the availability of soil moisture did not change significantly among the treatments. The accessibility of subsoil nutrients and moisture were indirectly assessed using BD and WSA, and NT-Ret showed better soil conditions at the transition layer (18-to-30 cm depth), the layer important for assessing root growth potential. The PT treatments showed the absence of SOM

transfer across the transition layer, whereas NT-Harv showed nutrient depletion at the transition and subsoil layers. These emphasized the importance of surface residue return combined with no-tillage to 1) provide better soil physical conditions for root growth and 2) maintain adequate nutrient concentrations across a soil profile especially when considering subsoil properties. Assessment of the sustainability of residue removal needs to include soil profile-scale evaluation of soil biological, chemical, and physical properties.

6.2 KEY WORDS

Crop residue, Maize, Soil Health, Subsoil, Tillage

6.3 INTRODUCTION

The health of soils impacts their ability to perform critical functions, including the support of crop growth. In rainfed agriculture, limited or excessive amounts of soil moisture during critical growth stages are important regulators for yield levels and yield stability (Boyer et al., 1990; Timlin et al., 2001), which are important factors for reducing unstable food supplies as well as income risks for growers (Gilbert and Morgan, 2010; Osborne and Wheeler, 2013). Subsoil layers (> 30 cm depth) have been identified as an important source of soil moisture (Ewing et al., 1991; Gaiser et al., 2012; Kirkegaard et al., 2007) and nutrients (Carter and Gregorich, 2010; Gransee and Merbach, 2000; Heming, 2004) for crop growth especially under moisture-limited conditions. Distinct soil microbial communities may also be present in subsoil layers compared to surface layers due to unique nutrient dynamics, soil physical properties, and

redox potential (Fischer et al., 2013; Leininger et al., 2006), and can be a sink for a large amount of soil organic carbon (SOC; Batjes, 1996). However, limited attention has been paid to the effects of land management on subsoil soil properties (Baker et al., 2007; Rumpel and Kögel-Knabner, 2010), and traditional soil testing on growers' fields has been limited to topsoil soil nutrients (typically 0-to-15 cm depth; O.J. Idowu et al., 2008). Shallow soil sampling has been justified due to i) the difficulties of sampling to deeper depths, ii) the relative importance of surface soil layer when adequate growing conditions are met, and iii) the assumptions that available nutrient levels are relatively constant for most soil types and do not change over time (Beatty and Corey, 1962; Kautz et al., 2013; Kirkegaard et al., 2007).

Variable effects of tillage across a soil profile has been shown primarily for SOC contents in relation to climate change though the concentration of SOC does not fully address the changes in soil conditions for plant growth nor does higher SOC necessarily mean higher crop productivity (Sojka et al., 2003). Past study has shown that no-till (NT) treatments to have higher SOC stocks in the surface layer (0-10 cm) while moldboard plow (PT) treatments to havehigher stocks in the deeper layers (20-40 cm) across eight sites in Canada (Angers et al., 1997). For crop growth, the interface between the plow layer (cultivated soil layer) and the subsoil has been shown important, and this layer is referred to as "transition layer" (Peigné et al., 2013). It corresponds to the soil depth where a plow pan could form (Spoor et al., 2003), which causes compaction and reduced root growth (Alakukku, 2000; Van den Akker et al., 2003). Transition layer conditions may also be important for the maintenance of soil nutrients and soil organic matter (SOM) levels in subsoil layers, which have been found to rely on the exchanges to and from topsoil via plant root systems and soil fauna, especially earthworms (Kautz et al., 2013), and dissolved SOM by preferential flow (Rumpel and Kögel-Knabner, 2010). The evaluation of

soil physical conditions in transition and subsoil layers has often been carried out by visual and qualitative assessments (Peigné et al., 2013; Roger-Estrade et al., 2004), and quantitative assessment of the effects of crop and soil management on soil conditions at this layer or subsoil through biological, chemical, and physical indicators has been limited.

In recent years, combinations of soil measurements including i) soil biological assessment of total and labile components of SOM, ii) soil physical assessment of water stable aggregation (WSA), available water capacity (AWC) and soil strength, and iii) soil nutrient and pH indicators have been shown important to determining yield constraints and have been utilized as a soil health test (Idowu et al., 2008; Karlen et al., 2001; Schindelbeck et al., 2008). Such a set of measurements has been successfully applied to detect aspects of soil degradation caused by tillage (Moebius-Clune et al., 2008; Van Eerd et al., 2014) and land use change (Moebius-Clune et al., 2011). The test also allows for investigation of the interactions among soil biological, chemical, and physical properties, but has mainly been applied for surface soil layers. There is a need to assess whether it is also relevant to quantitatively assessing the effects of soil and crop management in subsoil layers.

This study was conducted on 40-year continuous maize experimental plots with tillage and maize residue management treatments. Our hypothesis is that PT creates a root growth restricting layer that does not allow the effective movement of residue-derived organic materials and nutrients to the subsoil. Also, we hypothesize that the absence of residue return causes unfertilized nutrients to get depleted especially from the deeper soil layers where the amount of root residue is lower.

The objective of this study was to investigate the degree of impacts of crop and soil management on surface as well as subsurface layer soil conditions using soil physical, chemical, and biological indicators.

6.4 MATERIALS AND METHODS

6.4.1 Study site

The study site is located in Chazy, NY (44°53'N, 73°28'W). There are 16 plots with four treatments of crossed tillage (plow till vs. no-till) and maize stover removal (residue returned vs. residue harvested) treatments: No-till/residue harvested (NT-Harv), plow-till/residue harvested (PT-Harv), no-till/residue returned (NT-Ret), plow-till/residue returned (PT-Ret; Figure 6.1), as well as continuous mixed grass sod (SOD) in the surrounding area. All the experimental plots share one soil series: Roundabout silt loam (Aeric Endoaquept: coarse-silty, mixed, active, nonacid, frigid). The soil was formed from medium-textured glaciolacustrine and glaciomarine deposits of Wisconsin Age on the Lake Champlain Plain, near Plattsburgh, NY. According to the Official Series Description, surface 18 cm is in Ap horizon, 18-43 cm in Bw, 43-66 cm in Bg, 66-76 cm in BCg, and 76-165 cm in C horizon (Soil Survey Staff et al., n.d.). The soil is poorly to somewhat poorly drained, and the plots are tile drained to the depth of 100 cm. The experiment was established in a randomized complete block design in 1973 after many years of SOD. The PT plots were moldboard plowed and disked annually in the fall, and maize was planted in the spring, while the NT plots were not tilled and planted with a NT planter

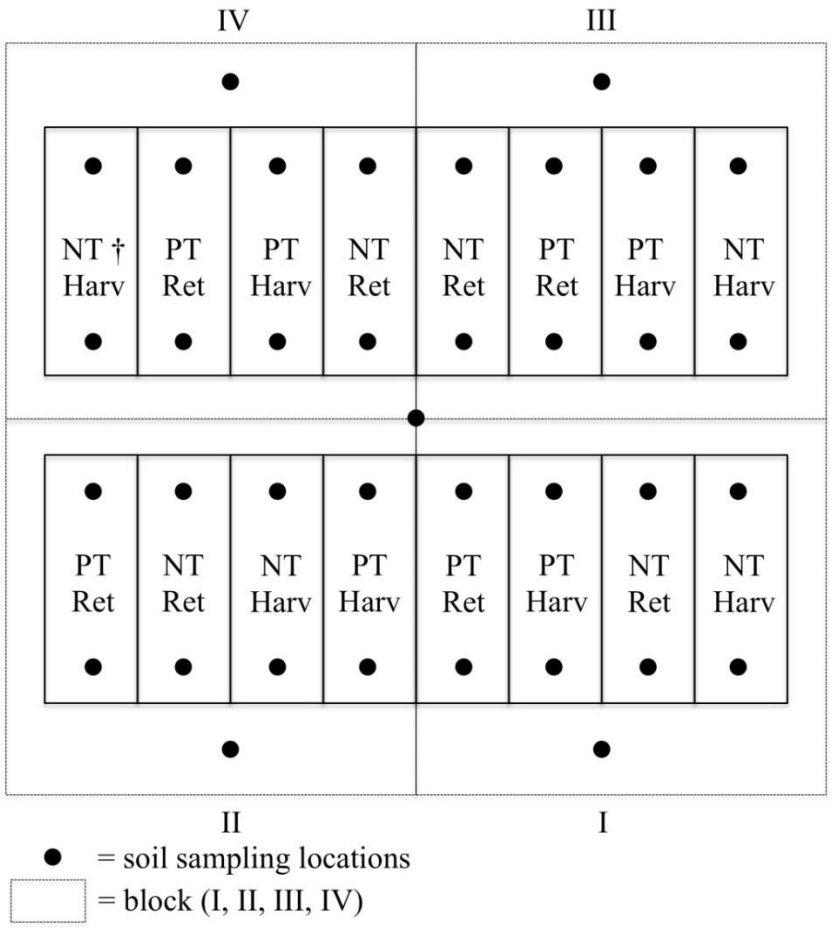

Figure 6.1 Experimental layout with no-till (NT), plow till (PT), maize residue returned (Ret), and residue removed (Harv) established in 1973 in Chazy, NY.

(Idowu et al., 2009). In general, a maize hybrid with maturity class of 85 to 90 days was planted. Fertilizer management of the field consisted of banded application of 17 kg N ha^{-1}, 67 kg P_2O_5 ha^{-1}, and 67 kg K_2O ha^{-1} at the time of planting. In addition, a side-dress application of 140 kg N ha^{-1} was added when the maize plants were between 30 and 45 cm tall. Weed management in the recent years consisted of pre-emergence herbicide applications of an S-metolachlor, atrazine, and mesotrione mixture followed by glyphosate early in the growing season depending on the level of weed pressure. The plow depth was between 15 and 20 cm (Ramsey, 1984).

6.4.2 Soil sampling

Soil sampling was undertaken in July 2013 using a 3.81 cm diameter soil sampling probe. Two subsamples at two locations within each experimental plot were taken in non-traffic inter-rows of maize away from field edges, and were then composited. The soil samples were cut in 0-6, 6-18, 18-30, 30-45, and 45-60 cm increments. The first two increments were generally in the Ap horizon, the third and the fourth in the Bw horizon, and the last increment in the Bg horizon. Five additional samples were taken in SOD using the same equipment (Figure 6.1). The soil samples were kept at 4 °C until analysis.

6.4.3 Soil analysis

Whole soil samples were weighed to estimate percent field soil moisture content and a subsample was subsequently weighed again, after being oven-dried at 105 °C, to determine dry sample weight. Dry soil bulk density (BD) was then determined using the volume of the

sampling core and the dry sample weight, with rock fragments subtracted, based on the standard rock density to convert the rocks from mass to volume.

Soil texture was assessed using a rapid quantitative method developed by Kettler et al. (2001). The soil sample was dispersed with 3% sodium hexametaphosphate ($(NaPO_3)_n$). A combination of sieving and sedimentation steps was used to separate size fractions. Water stable aggregation was assessed using a rainfall simulator (Ogden et al., 1997) that allows the soil particles to receive the impacts of known rainfall energy, applying 2.5 J of energy for 300 s on aggregates (0.25–2 mm) placed on a 0.25-mm mesh sieve. The fraction of soil aggregates remaining on the sieve, corrected for stones >0.25 mm, was regarded as the percent WSA after drying at 105°C (Gugino et al., 2009). Water retention at -10 kPa, -33 kPa, -100 kPa, and -1500 kPa were assessed gravimetrically. Saturated soil samples were equilibrated to each pressure listed above on ceramic high pressure plates (Dane and Hopmans, 2002). Volumetric water content was derived using the measured BD multiplied by the measured gravimetric water contents at each pressure point. -10 kPa is often regarded as field capacity in coarse textured soils whereas -33 kPa is used for medium- to fine-textured soils (Hudson, 1994). In this study, the difference between -10 kPa and -1500 kPa was described as AWC to be consistent with a previous study (Moebius-Clune et al., 2008).

Various fractions of SOM were quantified to assess both the quantity and the quality of SOM. Total SOM content was analyzed by mass loss on ignition in a muffle furnace at 500 °C for two hours. Labile component of SOM was estimated as permanganate oxidizable carbon (POXC) using dilute potassium permanganate ($KMnO_4$), which is an effective method to quantify easily oxidizable carbon by measuring absorbance at 550 nm using a hand-held colorimeter (Weil et al., 2003). Preserved field moist samples were used for C-min. C-min was

determined using sealed chamber alkali trap respirometers using capillary rewetted soil (Haney and Haney, 2010), and it measures the metabolic activity of the soil microbial community (D. Moebius-Clune et al., 2014). Carbon dioxide evolved from rewetted soils over a 4-day room temperature incubation was trapped in KOH and quantified by conductivity change in an alkali trap. The autoclaved citrate extractable fraction of soil proteins and protein-like substances (Protein) was measured, which is a proxy measurement of the large fraction of organically bound nitrogen (N) in total SOM (D. Moebius-Clune et al., 2014). The extraction was with 0.02 M sodium citrate at pH7, and the extract was then quantified by bicinchoninic acid assay against a bovine serum albumin standard curve for soil protein concentration after a sequence of centrifugation and autoclaving steps (Walker, 2002; Wright and Upadhyaya, 1996). The ratio of Protein to SOM (Protein:SOM) was calculated as an indicator of the relative quality of the SOM. A lower ratio indicates a relative richness of organically bound N in the SOM, and it relates to potential N availability through mineralization (D. Moebius-Clune et al., 2014).

Soil pH was measured in 1:1 water slurry. Plant available soil nutrient concentrations were measured by extracting the nutrients using Modified Morgan, an ammonium acetate solution, buffered at pH 4.8 (McIntosh, 1969), and analyzed using inductively coupled plasma emission spectroscopy.

6.4.4 Data analysis

We assessed a scatter matrix of the dataset to confirm linear associations among the measured soil properties. We calculated Pearson correlation coefficients to assess the relationships among the soil properties within the experimental plots (n = 80).

The effects of fixed factors (residue, tillage, depth, residue × tillage, and residue × tillage × depth) and random factors (block, replicate, and their interactions) on both grain yields and soil properties were assessed using a mixed model (SAS Institute Inc., 2015). Ca values were log transformed because of the identified unequal variances. The variance explained by the fixed factors, and the variance explained by both the fixed and random factors, were determined using marginal R^2 (R^2m) and conditional R^2 (R^2c), respectively (Nakagawa and Schielzeth, 2013; Vonesh et al., 1996). They were calculated using the *MuMIn* package (Bartoń, 2015) in the R statistical computing environment. Post hoc tests were carried out to compare the means of measured soil properties in each fixed factor treatment at $p \leq 0.05$ using Tukey's method. The soil test results from SOD were included as references to assess the changes in soil conditions but were not used for statistical comparisons due to a lack of randomization with the tillage and residue treatments.

In order to illustrate the overall soil condition, we scored each treatment at each depth increment as Soil Health Score using 12 selected indicators (SOM, C-min, Protein, Protein:SOM, P, K, Ca, Mg, BD, WSA, water content at -100kPa, and AWC) to represent soil biological, chemical, and physical properties uniformly. Apart from BD, each indicator was ranked from the 1st to the 4th as "more is better" for all measured values, except BD which was ranked as "less is better". We assigned four points to the 1st, three points to the 2nd, two points to the 3rd, and one point to the 4th rank.

6.5 RESULTS AND DISCUSSION

6.5.1 Within-site variation of inherent soil properties

We evaluated the magnitude of tillage and residue management effects on the measured soil properties using two classes of pseudo R^2 values calculated for each fitted mixed model (Nakagawa and Schielzeth, 2013), which was found to be extremely useful to determine the interactions between the fixed effects and spatial differences in inherent soil properties. The R^2m indicated higher variance explained for soil biological properties ($0.66 \leq R^2\text{m} \leq 0.91$; Table 6.1) compared to soil physical ($0.33 \leq R^2\text{m} \leq 0.85$) and chemical properties ($0.11 \leq R^2\text{m} \leq 0.88$). These suggest that tillage and residue management most strongly determine the variation of soil biological properties within this experimental site but they are less influential on soil physical properties especially water retention related values and soil chemical properties of Ca, Zn, and S. With the incorporation of random factors, the R^2c values were higher than that of R^2m for all soil properties but the increase was especially large for S, water content at -1500 kPa and -33 kPa, and Zn (Table 6.1) indicating significant interactions among the fixed factors and inherent spatial soil variations. S and Zn were not applied through fertilizer application. The correlations between S and other soil properties were in general low with the highest correlation with Ca (r = 0.49; Table 6.5). Soil water retained at higher pressures (< -100 kPa) is thought to be primarily regulated by soil texture since the water is retained by adsorption instead of capillary action (Hillel, 1980).

6.5.2 Surface (0-to-18 cm) soil properties

Surface layer is the most frequently tested zone for soil properties such as for fertilizer recommendations. The majority of the properties with significant treatment effects showed the

order of NT-Ret>NT-Harv>PT-Ret>PT-Harv (Figure 6.3; Table 6.2, 6.3, and 6.4). Each soil biological property test aims to represent different fractions of SOM, trying to explain the quality that relates to agronomically important processes (e.g. N mineralization). POXC and Protein are chemically extracted fractions of SOM while C-min is mediated by soil microbial biomass and activity. C-min has been suggested to be more closely linked with residue quality than POXC (Culman et al., 2013). In this layer, all measured soil biological properties were highly correlated to each other ($r > 0.83$; Table 6.5). They all followed the same trend of NT-Ret being the highest and PT-Harv being the lowest. We calculated the ratio of Protein:SOM to represent the relative abundance of soil protein and protein-like materials in the total SOM. At this layer, there was no significant difference among the treatments (Table 6.2). The mean pH value was significantly lower for NT-Ret compared to PT treatments (Table 6.3) and it was negatively correlated to SOM in this layer (Table 6.5). Acidification of surface soils under NT management has been shown previously, and suggested to be caused by the nitrification of surface applied N fertilizer (Blevins et al., 1977). P, K, and Zn contents were significantly higher for NT-Ret and were significantly positively correlated to all soil biological properties (Table 6.5). This demonstrates the presence of positive benefits from NT and Ret management on soil nutrient availability, which is analogues to previous studies that indicated positive effects of surface enrichment of P, K, Zn, and Mn under NT-Ret (Franzluebbers and Hons, 1996). However, Ca and S concentrations did not have correlations to soil biological properties (Table 6.5), and soil test S values are often not adequate to address plant available S, since a large fraction of plant available S becomes available through mineralization. The P contents were in “High” category, K contents in “Medium”, and Mg contents in “High” for all treatments (Jokela et al., 2004).

BD showed a significant treatment effect only in the first 6 cm layer and it was

Table 6.1 Marginal (R^2m) and conditional (R^2c) coefficient of determination for each mixed model result

	R^2m‡	R^2c	Difference
BD†	0.85	0.88	0.03
WSA	0.73	0.82	0.09
−10 kPa	0.33	0.59	0.26
−33 kPa	0.53	0.88	0.35
−100 kPa	0.39	0.58	0.19
−1500 kPa	0.38	0.76	0.38
AWC	0.41	0.70	0.29
−33 ~ −100 kPa	0.53	0.91	0.38
−33 ~ −1500 kPa	0.65	0.91	0.26
−100 ~ −1500 kPa	0.53	0.75	0.22
SOM	0.91	0.94	0.03
POXC	0.66	0.93	0.27
C-min	0.90	0.92	0.02
Protein	0.91	0.95	0.04
pH	0.88	0.92	0.04
P	0.88	0.89	0.01
K	0.71	0.72	0.01
Ca	0.38	0.68	0.30
Mg	0.49	0.72	0.23
Zn	0.33	0.62	0.29
S	0.11	0.62	0.51

† BD = dry bulk density; WSA = water stable aggregation; AWC = available water capacity; SOM = soil organic matter; POXC = permanganate oxidizable carbon; C-min = mineralizable carbon

‡ R^2m = marginal coefficient of determination; R^2c = conditional coefficient of determination

Table 6.2 Means for soil biological properties

Soil depth	Tillage	Residue	SOM ‡	POXC	C-min	Protein	Protein:SOM
cm			g kg^{-1}	ppm	mg CO2 g^{-1} day^{-1}	mg g^{-1}	
0-6	NT†	Ret	36.9 a §	674 a	0.257 a	8.33 a	0.226a
	NT	Harv	29.9 b	487 b	0.219 b	6.10 b	0.204a
	PT	Ret	20.4 c	340 c	0.190 bc	4.35 c	0.213a
	PT	Harv	20.4 c	299 c	0.157 c	4.02 c	0.197a
	NT Mean		33.4 A	580 A	0.238 A	7.22 A	0.215A
	PT Mean		20.4 B	320 B	0.173 B	4.18 B	0.205A
	Ret Mean		28.6 A	507 A	0.224 A	6.34 A	0.219A
	Harv Mean		25.1 B	393 B	0.188 B	5.06 B	0.200A
	SOD		58.9	869	0.670	13.60	0.233
6-18	NT	Ret	32.0 a	496 a	0.216 a	6.78 a	0.212a
	NT	Harv	28.6 a	429 a	0.203 a	5.35 b	0.188a
	PT	Ret	20.2 b	320 b	0.180 ab	4.43 bc	0.219a
	PT	Harv	18.9 b	285 b	0.149 b	3.94 c	0.208a
	NT Mean		30.3 A	463 A	0.209 A	6.06 A	0.200A
	PT Mean		19.6 B	303 B	0.164 B	4.19 B	0.213A
	Ret Mean		26.1 A	408 A	0.198 A	5.61 A	0.216A
	Harv Mean		23.8 A	357 A	0.176 B	4.64 B	0.198A
	SOD		35.9	514	0.276	7.56	0.212
18-30	NT	Ret	19.7 a	256 a	0.164 a	3.80 a	0.192ab
	NT	Harv	14.1 b	130 b	0.133 a	1.99 b	0.138b
	PT	Ret	17.6 ab	276 a	0.158 a	3.81 a	0.218a
	PT	Harv	15.2 b	209 ab	0.131 a	2.99 ab	0.196a
	NT Mean		16.9 A	221 A	0.148 A	2.89 A	0.165B
	PT Mean		16.4 A	242 A	0.145 A	3.40 A	0.207A
	Ret Mean		18.6 A	266 A	0.161 A	3.80 A	0.205A
	Harv Mean		14.6 B	194 B	0.132 B	2.49 B	0.167B
	SOD		26.6	324	0.219	4.64	0.175
30-45	NT	Ret	7.3 a	na	0.082 a	0.98 a	0.134a
	NT	Harv	6.4 a	na	0.078 a	0.59 a	0.092a
	PT	Ret	4.4 a	na	0.060 a	0.43 a	0.100a
	PT	Harv	4.0 a	na	0.060 a	0.37 a	0.095a
	NT Mean		6.9 A	na	0.080 A	0.79 A	0.113A
	PT Mean		4.2 B	na	0.060 B	0.40 A	0.098A
	Ret Mean		5.8 A	na	0.071 A	0.70 A	0.117A
	Harv Mean		5.2 A	na	0.069 A	0.48 A	0.094A
	SOD		9.2	53.4	0.107	1.29	0.140
45-60	NT	Ret	6.1 a	na	0.067 a	0.68 a	0.100a
	NT	Harv	5.6 a	na	0.065 a	0.45 a	0.072ab
	PT	Ret	3.3 a	na	0.049 a	0.18 a	0.052ab
	PT	Harv	3.2 a	na	0.055 a	0.04 a	0.014b
	NT Mean		5.8 A	na	0.066 A	0.56 A	0.086A
	PT Mean		3.3 B	na	0.052 A	0.11 A	0.033B
	Ret Mean		4.7 A	na	0.058 A	0.43 A	0.076A
	Harv Mean		4.4 A	na	0.060 A	0.24 A	0.043B
	SOD		6.3	4.93	0.066	0.67	0.106

† NT = no-till; PT = plow-till; Harv = residue removed; Ret = residue returned

‡ SOM = soil organic matter; POXC = permanganate oxidizable carbon; C-min = mineralizable carbon

§ Means of each property followed by an identical lowercase alphabet are not significantly different at the $\alpha = 0.05$.Capital letters show an overall significance of tillage and residue effects.

Table 6.3 Means for soil chemical properties

Soil depth	Tillage	Residue	pH	P	K	Ca	Mg	Zn	S
cm				mg kg^{-1}	mg kg^{-1}	mg kg^{-1}	mg kg^{-1}	mg kg^{-1}	mg kg^{-1}
0-6	NT †	Ret	6.8 c §	14.81 a	132.0 a	3575 a	239 a	0.763 a	8.01 a
	NT	Harv	6.9 bc	12.15 a	66.3 b	3648 a	179 a	0.623 ab	8.63 a
	PT	Ret	7.1 ab	8.38 b	82.9 b	4601 a	165 a	0.625 ab	9.39 a
	PT	Harv	7.1 a	8.76 b	57.9 b	4286 a	143 a	0.438 b	8.74 a
	NT Mean		6.8 B	13.48 A	99.2 A	3611 A	209 A	0.693 A	8.32 A
	PT Mean		7.1 A	8.57 B	70.4 B	4443 A	154 A	0.531 A	9.07 A
	Ret Mean		6.9 B	11.59 A	107.5 A	4088 A	202 A	0.694 A	8.70 A
	Harv Mean		7.0 A	10.45 A	62.1 B	3967 A	161 A	0.530 B	8.69 A
	SOD		6.7	13.59	133.9	3648	307	1.120	12.20
6-18	NT	Ret	6.8 c	10.74 a	65.8 a	4787 a	203 a	0.620 a	6.92 a
	NT	Harv	6.8 bc	5.83 b	44.7 a	4882 a	156 a	0.605 a	7.66 a
	PT	Ret	7.0 b	5.96 b	69.5 a	4656 a	164 a	0.473 a	7.46 a
	PT	Harv	7.2 a	7.10 b	47.0 a	4531 a	147 a	0.383 a	6.90 a
	NT Mean		6.8 B	8.29 A	55.3 A	4834 A	180 A	0.613 A	7.29 A
	PT Mean		7.1 A	6.53 B	58.3 A	4593 A	156 A	0.428 B	7.18 A
	Ret Mean		6.9 B	8.35 A	67.7 A	4721 A	184 A	0.546 A	7.19 A
	Harv Mean		7.0 A	6.47 B	45.9 B	4707 A	152 A	0.494 A	7.28 A
	SOD		6.8	9.14	69.1	4277	224	0.616	8.95
18-30	NT	Ret	7.1 ab	4.13 ab	73.3 a	10483 a	274 a	0.823 a	11.86 a
	NT	Harv	7.2 a	1.92 b	43.2 c	7869 a	148 b	0.383 b	12.81 a
	PT	Ret	7.0 b	4.60 a	71.3 ab	5400 a	181 ab	0.478 b	7.30 a
	PT	Harv	7.1 ab	4.52 ab	45.5 bc	5741 a	171 ab	0.543 ab	7.20 a
	NT Mean		7.1 A	3.02 B	58.2 A	9176 A	211 A	0.603 A	12.33 A
	PT Mean		7.0 B	4.56 A	58.4 A	5570 A	176 A	0.510 A	7.25 B
	Ret Mean		7.0 A	4.37 A	72.3 A	7942 A	227 A	0.650 A	9.58 A
	Harv Mean		7.1 A	3.22 A	44.3 B	6805 A	159 B	0.463 B	10.00 A
	SOD		7.1	5.40	57.2	5850	176	0.534	9.38
30-45	NT	Ret	7.4 a	1.71 a	38.2 a	11046 ab	263 ab	0.633 a	8.16 a
	NT	Harv	7.3 a	1.11 a	32.8 a	6746 b	155 b	0.333 a	7.25 a
	PT	Ret	7.4 a	1.22 a	38.7 a	9618 ab	287 a	0.415 a	7.18 a
	PT	Harv	7.5 a	1.48 a	33.7 a	14790 a	294 a	0.403 a	9.30 a
	NT Mean		7.3 B	1.41 A	35.5 A	8896 A	209 B	0.483 A	7.70 A
	PT Mean		7.5 A	1.35 A	36.2 A	12204 A	291 A	0.409 A	8.24 A
	Ret Mean		7.4 A	1.47 A	38.5 A	10332 A	275 A	0.524 A	7.67 A
	Harv Mean		7.4 A	1.29 A	33.3 A	10768 A	225 A	0.368 A	8.27 A
	SOD		7.4	1.20	40.7	3136	139	0.278	5.57
45-60	NT	Ret	7.6 a	1.87 a	38.5 a	16318 a	298 a	0.528 a	9.07 a
	NT	Harv	7.6 a	1.20 a	30.4 a	8235 b	219 a	0.408 a	6.72 a
	PT	Ret	7.7 a	1.24 a	39.5 a	10811 ab	320 a	0.533 a	7.06 a
	PT	Harv	7.7 a	1.42 a	39.3 a	15542 ab	318 a	0.418 a	9.41 a
	NT Mean		7.6 A	1.53 A	34.4 A	12276 A	258 A	0.468 A	7.89 A
	PT Mean		7.7 A	1.33 A	39.4 A	13177 A	319 A	0.475 A	8.24 A
	Ret Mean		7.7 A	1.55 A	39.0 A	13565 A	309 A	0.530 A	8.06 A
	Harv Mean		7.6 A	1.31 A	34.8 A	11889 A	268 A	0.413 A	8.06 A
	SOD		7.5	1.63	33.6	5273	203	0.372	4.54

† NT = no-till; PT = plow-till; Harv = residue removed; Ret = residue returned

§ Means of each property followed by an identical lowercase alphabet are not significantly different at the $\alpha = 0.05$.Capital letters show an overall significance of tillage and residue effects.

Table 6.4 Means for soil physical properties

Soil depth	Tillage	Residue	BD ‡	WSA	−10 kPa	−33 kPa	−100 kPa	−1500 kPa
cm			Mg m^{-3}	%	m m^{-3}	m m^{-3}	m m^{-3}	m m^{-3}
0-6	NT †	Ret	1.27 b §	49 a	0.415 a	0.343 a	0.251 a	0.117 a
	NT	Harv	1.32 ab	39 a	0.428 a	0.348 a	0.238 ab	0.105 a
	PT	Ret	1.35 ab	24 b	0.432 a	0.353 a	0.230 ab	0.101 a
	PT	Harv	1.38 a	21 b	0.429 a	0.342 a	0.221 b	0.100 a
	NT Mean		1.29 B	44 A	0.422 A	0.346 A	0.245 A	0.111 A
	PT Mean		1.37 A	23 B	0.431 A	0.348 A	0.226 B	0.101 A
	Ret Mean		1.31 A	37 A	0.424 A	0.348 A	0.241 A	0.109 A
	Harv Mean		1.35 A	30 A	0.428 A	0.345 A	0.230 A	0.102 A
	SOD		1.13	70	0.487	0.343	0.258	0.134
6-18	NT	Ret	1.36 a	75 a	0.426 a	0.339 a	0.263 ab	0.123 a
	NT	Harv	1.41 a	69 a	0.443 a	0.356 a	0.268 a	0.115 a
	PT	Ret	1.39 a	40 b	0.402 a	0.349 a	0.245 ab	0.107 a
	PT	Harv	1.45 a	33 b	0.429 a	0.339 a	0.238 b	0.106 a
	NT Mean		1.39 A	72 A	0.434 A	0.348 A	0.265 A	0.119 A
	PT Mean		1.42 A	37 B	0.416 A	0.344 A	0.241 B	0.107 A
	Ret Mean		1.38 A	58 A	0.414 A	0.344 A	0.254 A	0.115 A
	Harv Mean		1.43 A	51 A	0.436 A	0.347 A	0.253 A	0.111 A
	SOD		1.42	87	0.468	0.340	0.271	0.125
18-30	NT	Ret	1.45 b	47 a	0.422 a	0.346 a	0.242 a	0.104 a
	NT	Harv	1.56 ab	42 ab	0.472 a	0.393 a	0.258 a	0.100 a
	PT	Ret	1.50 ab	36 ab	0.425 a	0.379 a	0.263 a	0.100 a
	PT	Harv	1.57 a	33 b	0.446 a	0.383 a	0.253 a	0.105 a
	NT Mean		1.50 A	45 A	0.447 A	0.370 A	0.250 A	0.102 A
	PT Mean		1.54 A	35 B	0.435 A	0.381 A	0.258 A	0.102 A
	Harv Mean		1.56 A	38 A	0.459 A	0.388 A	0.256 A	0.102 A
	Ret Mean		1.48 B	42 A	0.423 A	0.362 A	0.252 A	0.102 A
	SOD		1.48	65	0.444	0.341	0.270	0.113
30-45	NT	Ret	1.58 b	31 a	0.463 a	0.373 a	0.229 b	0.082 a
	NT	Harv	1.64 ab	33 a	0.476 a	0.417 a	0.262 a	0.089 a
	PT	Ret	1.67 ab	32 a	0.462 a	0.423 a	0.253 ab	0.081 a
	PT	Harv	1.69 a	31 a	0.486 a	0.423 a	0.245 ab	0.085 a
	NT Mean		1.61 B	32 A	0.470 A	0.395 B	0.246 A	0.086 A
	PT Mean		1.68 A	31 A	0.474 A	0.423 A	0.250 A	0.083 A
	Ret Mean		1.63 A	31 A	0.463 A	0.398 A	0.241 A	0.082 A
	Harv Mean		1.66 A	32 A	0.481 A	0.420 A	0.255 A	0.087 A
	SOD		1.61	35	0.441	0.373	0.234	0.088
45-60	NT	Ret	1.67 a	30 a	0.491 a	0.416 a	0.234 b	0.083 a
	NT	Harv	1.67 a	41 a	0.494 a	0.419 a	0.265 a	0.086 a
	PT	Ret	1.71 a	41 a	0.480 a	0.444 a	0.257 ab	0.079 a
	PT	Harv	1.70 a	33 a	0.535 a	0.446 a	0.273 a	0.092 a
	NT Mean		1.67 A	35 A	0.492 A	0.417 B	0.249 B	0.085 A
	PT Mean		1.71 A	37 A	0.508 A	0.445 A	0.265 A	0.086 A
	Ret Mean		1.69 A	36 A	0.486 A	0.430 A	0.245 B	0.081 A
	Harv Mean		1.69 A	37 A	0.514 A	0.432 A	0.269 A	0.089 A
	SOD		1.64	40	0.451	0.398	0.252	0.080

† NT = no-till; PT = plow-till; Harv = residue removed; Ret = residue returned

‡ BD = dry bulk density; WSA = water stable aggregation; −10 kPa = water content at −10 kPa;

§ Means of each property followed by an identical lowercase alphabet are not significantly different at the $\alpha = 0.05$.Capital letters show an overall significance of tillage and residue effects

Table 6.4 (Continued)

Soil depth	Tillage	Residue	AWC ‡	−33 ~ −100 kPa	−33 ~ −1500 kPa	−100 ~ −1500 kPa
cm			m m^{-3}	m m^{-3}	m m^{-3}	m m^{-3}
0-6	NT †	Ret	0.298 a §	0.092 a	0.226 a	0.134 a
	NT	Harv	0.323 a	0.110 a	0.243a	0.133 a
	PT	Ret	0.331 a	0.123 a	0.252 a	0.129 a
	PT	Harv	0.329 a	0.120 a	0.242 a	0.121 a
	NT Mean		0.311 A	0.101 A	0.235 A	0.134 A
	PT Mean		0.330 A	0.122 A	0.247 A	0.125 A
	Ret Mean		0.315 A	0.108 A	0.239 A	0.132 A
	Harv Mean		0.326 A	0.115 A	0.243 A	0.127 A
	SOD		0.353	0.086	0.210	0.124
6-18	NT	Ret	0.303 a	0.077 a	0.216 a	0.140 a
	NT	Harv	0.328 a	0.088 a	0.241 a	0.153 a
	PT	Ret	0.295 a	0.104 a	0.241 a	0.137 a
	PT	Harv	0.323 a	0.101 a	0.233 a	0.132 a
	NT Mean		0.315 A	0.082 A	0.229 A	0.146 A
	PT Mean		0.309 A	0.102 A	0.237 A	0.135 A
	Ret Mean		0.299 A	0.090 A	0.229 A	0.139 A
	Harv Mean		0.326 A	0.095 A	0.237 A	0.142 A
	SOD		0.343	0.068	0.214	0.146
18-30	NT	Ret	0.318 a	0.105 a	0.242 a	0.138 a
	NT	Harv	0.372 a	0.135 a	0.293 a	0.158 a
	PT	Ret	0.325 a	0.116 a	0.279 a	0.163 a
	PT	Harv	0.341 a	0.130 a	0.278 a	0.148 a
	NT Mean		0.345 A	0.120 A	0.268 A	0.148 A
	PT Mean		0.333 A	0.123 A	0.278 A	0.155 A
	Harv Mean		0.356 A	0.133 A	0.286 A	0.153 A
	Ret Mean		0.321 A	0.110 A	0.261 A	0.150 A
	SOD		0.331	0.071	0.228	0.157
30-45	NT	Ret	0.381 a	0.144 a	0.290 a	0.146 a
	NT	Harv	0.387 a	0.154 a	0.327 a	0.173 a
	PT	Ret	0.381 a	0.170 a	0.342 a	0.172 a
	PT	Harv	0.401 a	0.176 a	0.338 a	0.162 a
	NT Mean		0.384 A	0.149 A	0.309 B	0.160 A
	PT Mean		0.391 A	0.173 A	0.340 A	0.167 A
	Ret Mean		0.381 A	0.157 A	0.316 A	0.159 A
	Harv Mean		0.394 A	0.165 A	0.332 A	0.167 A
	SOD		0.353	0.140	0.285	0.145
45-60	NT	Ret	0.408 a	0.181 a	0.333 a	0.151 a
	NT	Harv	0.408 a	0.154 a	0.333 a	0.179 a
	PT	Ret	0.401 a	0.187 a	0.365 a	0.177 a
	PT	Harv	0.443 a	0.172 a	0.353 a	0.181 a
	NT Mean		0.408 A	0.168 A	0.333 A	0.165 A
	PT Mean		0.422 A	0.180 A	0.359 A	0.179 A
	Ret Mean		0.405 A	0.184 A	0.349 A	0.164 A
	Harv Mean		0.425 A	0.163 A	0.343 A	0.180 A
	SOD		0.371	0.146	0.318	0.172

† NT = no-till; PT = plow-till; Harv = residue removed; Ret = residue returned

‡AWC = available water capacity calculated by the difference between −10 kPa and −1500 kPa

§ Means of each property followed by an identical lowercase alphabet are not significantly different at the $\alpha = 0.05$.Capital letters show an overall significance of tillage and residue effects

Table 6.5 Pearson correlation coefficients of measured soil properties at the topsoil (0-to-18 cm depth)

	BD	WSA	−10 kPa	−33 kPa	−100 kPa	−1500 kPa	AWC	−33 ~ −100 kPa	−33 ~ −1500 kPa	−100 ~ −1500 kPa	SOM	POXC	C-min	Protein	pH	P	K	Ca	Mg	Zn	S
BD †	1.00																				
WSA	ns	1.00																			
−10 kPa	ns	ns	1.00																		
−33 kPa	ns	ns	ns	1.00																	
−100 kPa	ns	0.73***	ns	ns	1.00																
−1500 kPa	ns	ns	ns	ns	0.65***	1.00															
AWC	ns	ns	0.93***	ns	ns	−0.42*	1.00														
−33 ~ −100 kPa	ns	ns	ns	0.81***	ns	ns	ns	1.00													
−33 ~ −1500 kPa	0.43*	ns	ns	0.85***	ns	ns	ns	0.88***	1.00												
−100 ~ −1500 kPa	0.38*	0.50**	ns	ns	0.50**	ns	ns	ns	ns	1.00											
SOM	−0.60***	0.53**	ns	ns	0.60***	0.55**	ns	ns	ns	ns	1.00										
POXC	−0.66***	0.38*	ns	ns	0.54**	0.53**	ns	ns	ns	ns	0.96***	1.00									
C−min	−0.61***	0.39*	ns	ns	0.49**	0.43*	ns	ns	ns	ns	0.86***	0.85***	1.00								
Protein	−0.58***	0.46**	ns	ns	0.52**	0.48**	ns	ns	ns	ns	0.96***	0.95***	0.83***	1.00							
pH	0.43*	−0.63***	ns	ns	−0.66***	−0.50**	ns	ns	ns	ns	−0.81***	−0.76***	−0.73***	−0.80***	1.00						
P	−0.50**	ns	ns	ns	ns	ns	ns	ns	ns	ns	0.67***	0.68***	0.61***	0.75***	−0.44*	1.00					
K	−0.54**	ns	ns	ns	ns	ns	ns	ns	ns	ns	0.48**	0.55**	0.62***	0.57***	−0.36*	0.58***	1.00				
Ca	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	1.00			
Mg	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	0.38*	ns	0.54**	0.40**	−0.42*	0.41*	0.59***	ns	1.00		
Zn	−0.55**	ns	ns	ns	0.39*	ns	ns	ns	ns	ns	0.63***	0.64***	0.71***	0.58***	−0.62***	0.46**	0.52**	ns	0.59***	1.00	
S	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	0.66***	ns	ns	1.00

† BD = bulk density; WSA = water stable aggregation; −10 kPa = water content at −10 kPa; AWC = available water capacity; SOM = soil organic matter; POXC = permanganate oxidizable carbon; C−min = carbon mineralization

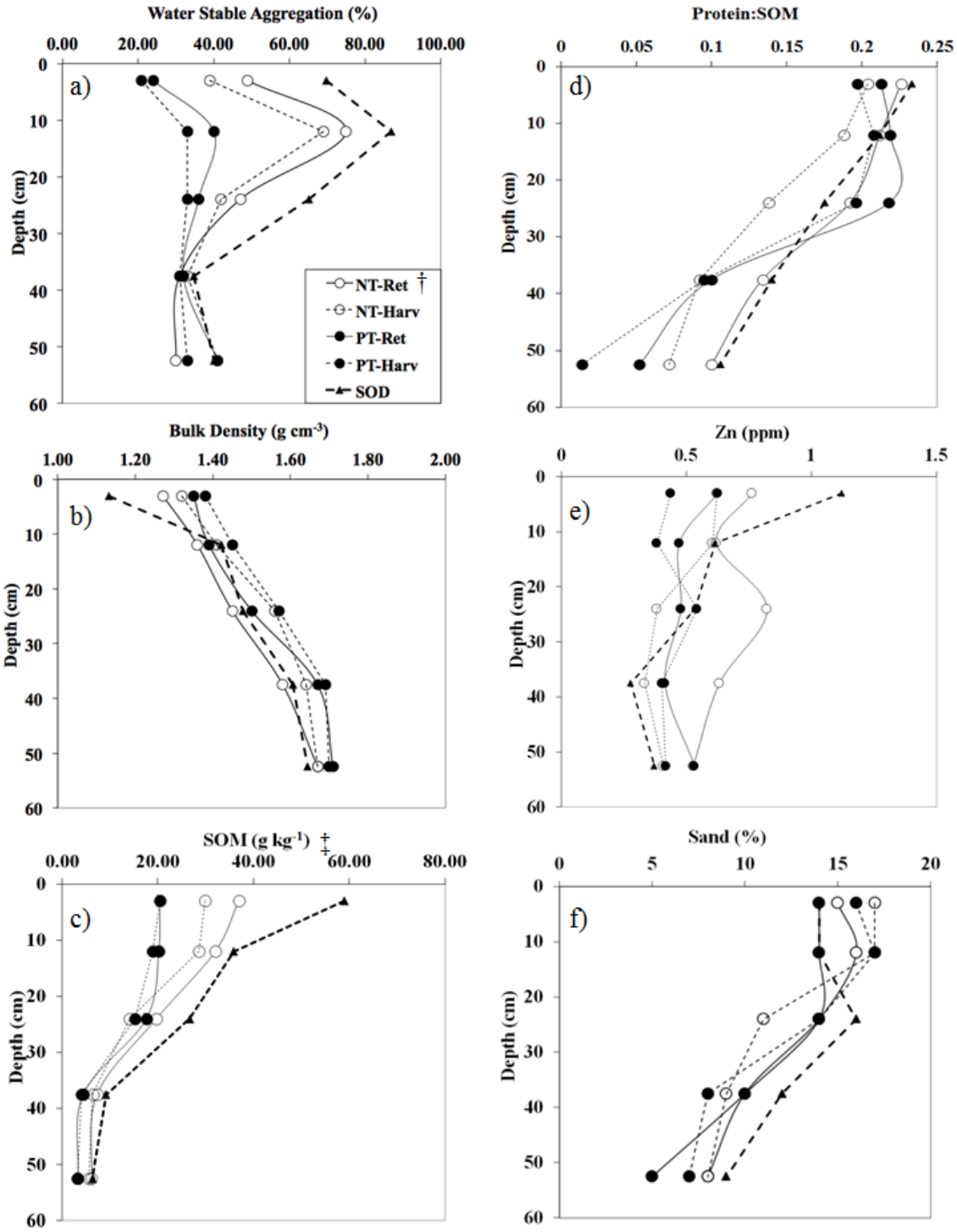

Figure 6.2 Soil profile plots showing the variation of soil properties for no-till residue returned (NT-Ret), no-till residue harvested (NT-Harv), plow-till residue returned (PT-Ret), plow-till residue harvested (PT-Harv), and continuous mixed grass sod (SOD) † NT = no-till; Ret = residue returned; PT = plow-till; Harv = residue harvested
‡ SOM = soil organic matter

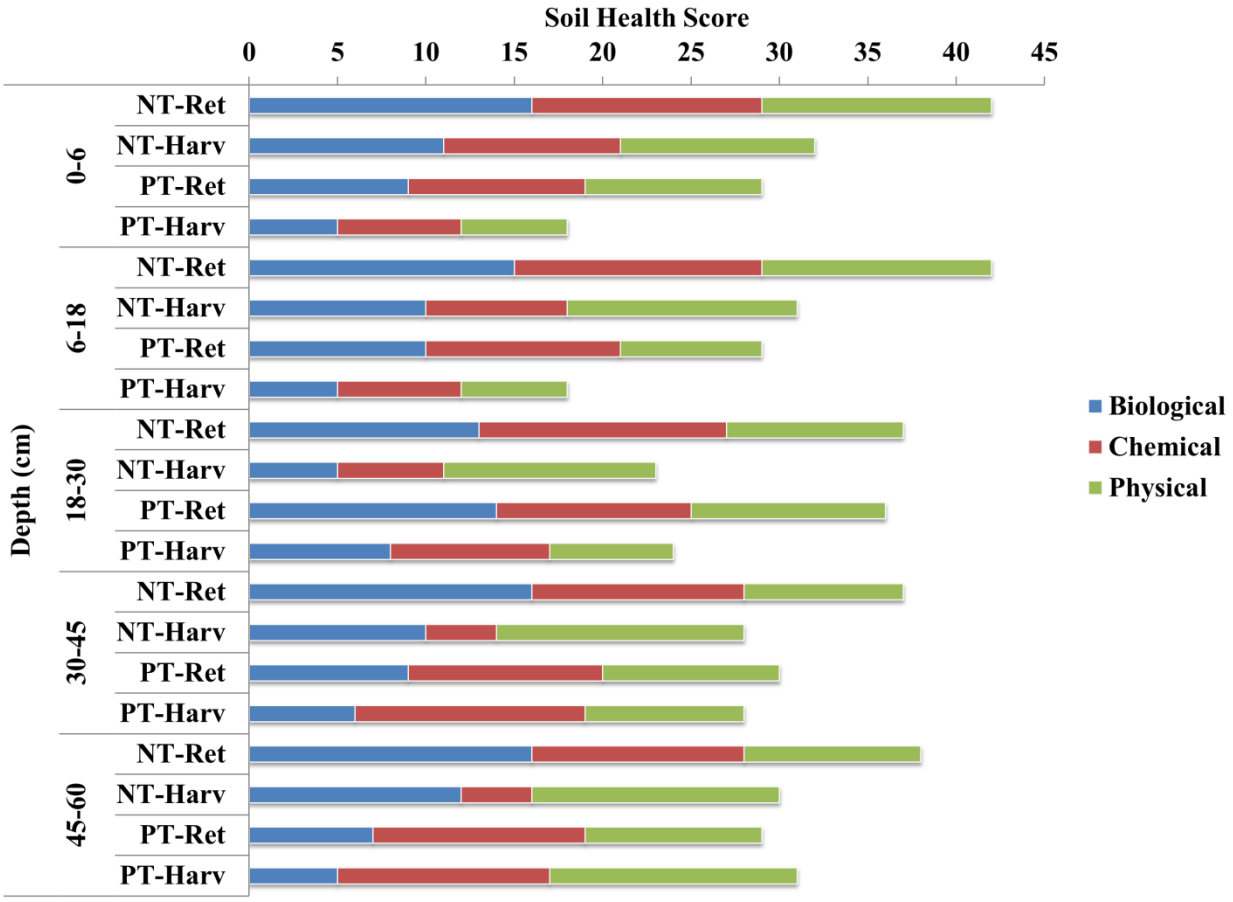

Figure 6.3 Overall profile soil health conditions scored using 12 indicators for no-till residue returned (NT-Ret), no-till residue harvested (NT-Harv), plow-till residue returned (PT-Ret), and plow-till residue harvested (PT-Harv). Simple scores were assigned based on relative ranking of the four treatments for each indicator, where 1st, 2nd, 3rd, and 4th rankings yielded scores of 4, 3, 2, and 1, respectively.

significantly lower in NT-Ret compared to PT-Harv (Table 6.4) in line with a previous study at the same site (Moebius-Clune et al., 2008). We then calculated the optimum BD for rooting and also the critical BD when the rooting approaches 20 % of the maximum using the equation derived by Allan Jones (1983). At 0-to-6 cm depth, all BD values were below the optimum BD (≈1.45), and the PT treatments approached the optimum BD at the 6-to-18 cm depth. Therefore, tillage does not appear to restrict root development at this layer. We measured water retention parameters at four pressure points at -10, -33, -100, and -1500 kPa in this study. At this layer, the treatment effects were only significant for -100 kPa where NT-Ret showed significantly higher values than PT-Harv (Table 6.4). There were no treatment effects for -1500 kPa, which indicates that soil management has a small impact on the amount of small pores, which was also shown by the R^2m value (R^2m = 0.38; Table 6.1). The NT treatment showed significantly higher water content at -100 kPa and this was significantly and positively correlated to WSA but not at other pressure points (Table 6.5). This suggests that the abundance of stable aggregates (> 250 μm) has an influence on soil water retained at the corresponding pore sizes. Traditionally, soil water retained at low pressures (> -100 kPa) was thought to be mainly regulated by soil structure not by soil texture because the water is retained by capillary effect rather than adsorption (Hillel, 1980). However, Guber et al. (2003) found complex relationships between aggregate size distributions and water retention, where they found the effects of aggregates in a wider pressure range from -33 kPa to -1500 kPa.

6.5.3 Transition layer (18-to-30 cm) soil properties

This is the layer comprised of the lower part of topsoil and also the most upper part of subsoil affected by soil management. This layer has also been recognized as the critical zone for assessing root development and water retention (Peigné et al., 2013).

We observed the direct effects of residue return with the PT-Ret treatment at this layer. Some of the indicators showed the order of NT-Ret> PT-Ret> PT-Harv>NT-Harv such as SOM, C-min, K, and Mg, whereas other indicators showed PT-Ret> NT-Ret> PT-Harv>NT-Harv such as POXC and Protein (Figure 6.3; Table 6.2, 6.3, and 6.4). The measurement of different components of SOM revealed the differences in the effects of the treatments at this layer. POXC has been found to be correlated with heavy and small particulate organic carbon, thus representing a relatively stable fraction of labile C (Culman et al., 2012). Harv treatment showed significantly lower POXC content at this layer compared to Ret treatment, and NT-Harv showed the lowest content (Table 6.2). Labile C is an important component of SOM that affects cycling of nutrients, soil aggregation, and water retention (Amézketa, 1999; Culman et al., 2013, 2012; Hudson, 1994), that are thought to improve the critical functions of transition layer including water retention and root development. The ratio of Protein:SOM also showed a significantly lower value for NT-Harv (Figure 6.2d; Table 6.2), which indicates a relatively small fraction of protein rich SOM under this treatment, and suggests a lower potential of mineralizable nutrient pool.

The measured soil chemical properties show the lowest concentrations under NT-Harv for P, K, Mg, and Zn (Table 6.3), and also showed the zone of nutrient depletion at this layer and below (Figure 6.2e; Table 6.3). Active uptake of the nutrients by the maize crop and the lack of surface residue return is likely to cause this depletion, as also shown by the low concentration of

POXC and the lower Protein:SOM value. P was the only soil nutrient that showed significant correlations to SOM, POXC, and Protein (Table 6.6), and had a higher value under PT (Table 6.3). It is interesting to note that significant differences in Mg and Zn contents between NT-Ret and NT-Harv are visible for the first time at this layer (Table 6.3). As discussed earlier, Mg and Zn were not supplied through fertilizer applications, and therefore the nutrients are cycled both vertically and horizontally due to the movement of water, plant uptake, and re-deposition (Kautz et al., 2013). NT-Harv was the only treatment with available P, K, and Zn to be in the "Low" category under the guideline (Jokela et al., 2004).

For soil physical properties, we observed significant treatment effects for WSA, with NT-Ret being the highest value (Table 6.4). It was not correlated to any other measured soil property values (Table 6.6) and showed a unique soil profile distribution (Figure 6.2a) compared to indicators such as BD and SOM (Figure 6.2b and 6.2c). It is an indicator of aeration, water infiltration, and drainage (Kemper and Rosenau, 1986), which is highly important in medium to fine-textured soils because it helps to protect a range of pore sizes (Idowu et al., 2008). It is known that plant roots and hyphae support soil aggregation, and polysaccharides become more important for WSA when SOC is less than 10 g kg^{-1} (Tisdall and Oades, 1982). Available labile-C is thought to be an important component of WSA (Aziz et al., 2013) since it is related to soil microbes that produce organic binding agents for stabilization (Angers et al., 1992), and higher WSA further improves physical protection for labile-C (Boehm and Anderson, 1997). However, the direct impacts of tillage appear more significant at this layer because of the lack of correlations to soil biological indicators, and tillage can disintegrate soil aggregates by exposing them to freeze-thaw and wet-dry cycles (Six et al., 2004), along with higher aeration and faster

Table 6.6 Pearson correlation coefficients of measured soil properties at the transition layer (18-to-30 cm depth)

	BD	WSA	−10 kPa	−33 kPa	−100 kPa	−1500 kPa	AWC	−33 ~ −100 kPa	−33 ~ −1500 kPa	−100 ~ −1500 kPa	SOM	POXC	C-min	Protein	pH	P	K	Ca	Mg	Zn	S
BD †	1.00																				
WSA	ns	1.00																			
−10 kPa	ns	ns	1.00																		
−33 kPa	0.60*	ns	ns	1.00																	
−100 kPa	ns	ns	ns	0.52*	1.00																
−1500 kPa	ns	ns	ns	ns	ns	1.00															
AWC	ns	ns	0.94***	ns	ns	ns	1.00														
−33 ~ −100 kPa	0.62*	ns	ns	0.92***	ns	ns	ns	1.00													
−33 ~ −1500 kPa	0.64**	ns	ns	0.94***	ns	ns	ns	0.87***	1.00												
−100 ~ −1500 kPa	ns	ns	ns	ns	0.71**	−0.68**	ns	ns	0.61*	1.00											
SOM	−0.83***	ns	ns	ns	ns	0.57*	ns	−0.50*	−0.63**	ns	1.00										
POXC	−0.73**	ns	ns	ns	ns	ns	ns	ns	ns	ns	0.86***	1.00									
C−min	−0.72**	ns	ns	ns	ns	ns	ns	ns	ns	ns	0.85***	0.69**	1.00								
Protein	−0.79***	ns	ns	−0.51*	ns	ns	ns	−0.54*	−0.63**	ns	0.90***	0.91***	0.73**	1.00							
pH	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	−0.56*	ns	−0.54*	1.00						
P	ns	ns	ns	ns	ns	ns	−0.61*	ns	ns	ns	0.51*	0.56*	ns	0.59*	ns	1.00					
K	ns	ns	−0.63**	ns	ns	ns	−0.69**	ns	ns	ns	ns	ns	ns	ns	ns	0.71**	1.00				
Ca	ns	ns	ns	ns	ns	ns	−0.51*	ns	ns	ns	ns	ns	ns	ns	ns	ns	0.60*	1.00			
Mg	ns	ns	−0.57*	ns	ns	ns	−0.56*	ns	ns	ns	ns	ns	ns	ns	ns	ns	0.76***	0.84***	1.00		
Zn	−0.56*	ns	−0.60*	ns	ns	ns	−0.71**	ns	ns	ns	ns	ns	ns	ns	ns	0.72**	0.82***	0.73**	0.78***	1.00	
S	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	0.67**	0.65**	0.50*	1.00

† BD = bulk density; WSA = water stable aggregation; −10 kPa = water content at −10 kPa; AWC = available water capacity; SOM = soil organic matter; POXC = permanganate oxidizable carbon; C−min = carbon mineralization

SOM decomposition (Morris et al., 2004). BD was significantly lower for NT-Ret (Table 6.4) and was identical to the SOD treatment at this depth (Figure 6.2b). Higher SOM content showed the strongest negative correlation to BD at this depth (r = -0.83; Table 6.6), which agrees with a past study (Adams, 1973). Based on the measured clay content, the optimum BD for root growth was calculated as 1.45 and the 20 % of the maximum rooting is expected when the BD approaches 1.70 at this depth (Allan Jones, 1983). We found the NT-Ret treatment maintaining the optimum BD while other treatments are exceeding the optimum value (Table 6.4). Water retention parameters showed a complex relationship with BD and SOM at this layer (Table 6.6). Due to the negative relationship between SOM and BD, higher SOM appeared to reduce water content held between -33 and -1500 kPa and a higher BD to increase the water content. However, BD was exceeding the optimum value (Allan Jones, 1983) and a framework such as the least limiting water range (LLWR), which is a value indicating a range within which the plant growth is not limited by moisture stress is necessary (da Silva et al., 1994). It has been shown to effectively incorporate mechanical impedance to infer realistic plant water availability. Archer and Smith (1972) argued that AWC can be managed through cultivation methods by increasing or decreasing BD but it is important to note that the increase in BD significantly reduced the LLWR in similar soil texture (da Silva et al., 1994). Therefore, we need to combine the AWC information with mechanical impedance or alternative information on plant root growth to conclude realistic in-situ plant water availability.

6.5.4 Subsoil (30-to-60 cm) soil properties

At this layer, the trend of the mean values were NT-Ret > NT-Harv > PT-Ret > PT-Harv for all the measured soil biological properties, apart from C-min at 45-to-60 cm depth, analogous

to the surface layer (Figure 6.3; Table 6.2). The higher mean soil biological property values in the transition layer for PT-Ret compared to NT-Harv were, in general, reversed in this layer although some of the indicators were not statistically significant (Table 6.2). Therefore, PT-Ret had a very small zone of high SOM related properties and they were not transported into the deeper layers. Due to the lower available energy source in the subsoil layers, the incorporation of fresh biomass has been identified to stimulate degradation of SOM (Fontaine et al., 2007). The combinations of mechanical disturbance and compaction as well as the priming effect caused by the incorporation of fresh residue at the transition layer after plowing may partly explain the lower SOM contents in the subsoil layers under PT.

There are three major sources of SOM to subsoil layers: i) crop roots and root exudates, ii) bioturbation by soil fauna, and iii) influx of dissolved SOM by preferential flow (Rumpel and Kögel-Knabner, 2010). Although not quantified in this study, there was a higher abundance of biopores observed in the NT plots, and also a significantly higher biomass of earthworms in the NT compared to PT at the same study site (Ramsey, 1984). The higher earthworm population could have contributed to the mixing of the topsoil with the subsoil through bioturbation, or by topsoil washing through continuous biopores into the subsoil layers (Kautz et al., 2013). This was also justified by the higher Protein:SOM value under NT-Ret, which indicates a relatively high content of N rich SOM materials (Table 6.2). The BD values were significantly lower for NT-Ret compared to PT-Harv (1.58 vs. 1.69; Table 6.4) at 30-to-45 cm depth, and the latter was approaching a value of 1.72, which is the calculated BD value at 20 % of rooting potential (Allan Jones, 1983). This further indicates the restrictions of the influx of SOM both by crop roots and also through deep continuous biopores in plowed soil.

All of the measured subsoil nutrients were the lowest for NT-Harv, and statistically significant differences were found for Ca and Mg although the levels were in "High" category (Jokela et al., 2004). They were not applied through fertilizer applications and not correlated to any measured soil biological properties (Table 6.7). Franzluebbers and Hons (1996) found a large decrease in extractable Mg at 30-to-60 cm followed by an increase at 60-to-90 cm depth both under NT and conventional plowing using disking and chisel plowing. They suggested the decrease of extractable Mg to be caused by plant uptake as well as a soil layer with inherently lower Mg content. In this study, we found lower Mg content (< 50 ppm) for NT-Harv in block III, which was located in the northwest corner of the study site (Figure 6.1). The SOD soil sample in block III also showed lower Mg content (< 100 ppm). The lower concentrations of the nutrients under NT-Harv suggests the presence of active uptake of these nutrients by plant roots combined with the absence of nutrient return, and requires further investigation of these nutrient pools for sustainable crop production. Interestingly, Ca, Mg Zn, and S were lower for SOD compared to maize treatments and showed higher uplifting potentials for those nutrients (Figure 6.2e; Table 6.3), and those nutrients except Ca were highly accumulated at the surface layer.

WSA showed relatively high and comparable levels to the topsoil (Table 6.4) although SOM and other soil biological indicator values were significantly lower (Table 6.2). John et al. (2005) found the formation of macroaggregates (> 250 μm) in low soil C soils, which contradicted the previous concept that the formation of macroaggregates can only start after the SOM binding capacity of clay and silt are satisfied (Hassink, 1997; Tisdall and Oades, 1982). Pinheiro-Dick and Schwertmann (1996) found higher proportions of poorly crystalline Fe-oxide in Inceptisol compared to Oxisol, which is known to improve soil aggregation (Duiker et al., 2003). In our study, Fe contents increased at greater depths (data not shown), and could have

Table 6.7 Pearson correlation coefficients of measured soil properties at the subsoil (30-to-60 cm depth)

	BD	WSA	−10 kPa	−33 kPa	−100 kPa	−1500 kPa	AWC	−33 ~ −100 kPa	−33 ~ −1500 kPa	−100 ~ −1500 kPa	SOM	POXC	C-min	Protein	pH	P	K	Ca	Mg	Zn	S
BD †	1.00																				
WSA	ns	1.00																			
−10 kPa	0.42*	0.32*	1.00																		
−33 kPa	0.58***	0.46**	ns	1.00																	
−100 kPa	ns	ns	ns	0.56***	1.00																
−1500 kPa	ns	ns	ns	ns	0.58***	1.00															
AWC	0.41*	0.35*	0.97***	ns	ns	ns	1.00														
−33 ~ −100 kPa	0.52**	0.48**	ns	0.74***	ns	ns	ns	1.00													
−33 ~ −1500 kPa	0.61***	0.54**	0.37*	0.93***	0.35*	ns	0.38*	0.83***	1.00												
−100 ~ −1500 kPa	ns	ns	ns	0.53**	0.85***	ns	ns	ns	0.52**	1.00											
SOM	−0.71***	ns	ns	−0.41*	ns	ns	ns	ns	−0.40*	ns	1.00										
POXC	−0.38*	ns	ns	−0.51**	ns	ns	ns	−0.46**	−0.49**	ns	0.43*	1.00									
C-min	−0.59***	ns	ns	ns	ns	ns	ns	ns	ns	ns	0.71***	ns	1.00								
Protein	−0.82***	ns	ns	−0.54**	ns	ns	ns	−0.45**	−0.54**	ns	0.84***	0.48**	0.63***	1.00							
pH	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	1.00						
P	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	1.00					
K	ns	ns	ns	ns	0.48**	0.38*	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	1.00				
Ca	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	0.81***	0.44*	1.00			
Mg	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	0.48**	0.38*	ns	0.47**	1.00		
Zn	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	0.69***	0.44*	0.54**	0.41*	1.00	
S	ns	−0.44*	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	0.69***	ns	0.77***	0.45**	0.48**	1.00

† BD = bulk density; WSA = water stable aggregation; −10 kPa = water content at −10 kPa; AWC = available water capacity; SOM = soil organic matter; POXC = permanganate oxidizable carbon; C−min = carbon mineralization

encouraged the formation of soil aggregation before the SOM binding capacity of clay and silt were satisfied. Furthermore, subsoil is generally less disturbed by soil management and is known to maintain biopores and soil structure longer compared to the topsoil (Beven and Germann, 1982). In this study, the measurements of WSA appeared not to reflect the effects of management at this depth.

6.5.5 Full profile soil conditions

In order to support optimum crop growth under stress conditions, sufficient availability and accessibility of soil moisture and nutrients are important (Boyer et al., 1990; Timlin et al., 2001). In this study, we determined that soil management practices have variable impacts on different soil layers, with the presence of a clear transition layer separating the topsoil layer and the subsoil layer under PT, and a more fuzzy transition layer under NT (Peigné et al., 2013). Soil management, here represented by tillage and residue return, can affect the soil properties at different depths depending on the combination of the management choice, which can in turn affect the availability and accessibility of soil moisture and nutrients.

The topsoil layer (0-to-18 cm) is the most important reservoir of soil moisture and nutrients as well as oxygen for plant growth. We observed more significant effects of tillage compared to residue management at this depth on soil biological properties, and the residue management effects were minimal under PT at 0-to-6 cm (Table 6.2; Figure 6.3). This is because the residue is diluted across the plow depth under PT, hiding the effects of residue return at the very shallow depth. NT treatments also maintained significantly higher aggregation, which could be a more sensitive indicator of management-induced change in soil physical properties compared to BD (Table 6.4). Soil water retention related properties did not change significantly

at the topsoil layer and across the soil profile, possibly due to the negative relationship between BD and SOM, which was contrary to a past finding (Hudson, 1994).

The transition layer (18-to-30 cm) has been recognized as an important zone for soil physical assessment for root growth and water retention (Peigné et al., 2013), which we confirmed with higher WSA under NT and by significantly lower BD under NT-Ret (Table 6.4). In addition, we found nutrient depletion under NT-Harv for P, K, and Zn (Table 6.3), the levels classified as "Low" (Jokela et al., 2004), which emphasizes the importance of considering the effects of residue management below the conventional soil sampling depth of 0-to-15 cm (Figure 6.3).

The subsoil layer (30-to-60 cm depth) has been recognized as an important reservoir of soil moisture, nutrients, and SOC (Batjes, 1996; Carter and Gregorich, 2010; Ewing et al., 1991; Gaiser et al., 2012; Gransee and Merbach, 2000; Heming, 2004; Kirkegaard et al., 2007). We found fertilizer added nutrients of P and K to be significantly lower in the subsoil for all treatments but they could be utilized if the roots can reach the layer. Other cations of Ca, Mg, and Zn were in high concentration reflecting the inherent soil properties, and Zn was maintained in the "Medium" range throughout the soil profile under NT-Ret (Jokela et al., 2004). Higher SOM content and a higher Protein:SOM value under NT-Ret indicated the importance of SOM redistribution to the subsoil layer from the surface. In the subsoil layer, we believe knowledge of available nutrients affected by soil formation is important, which may allow us to use deep rooting crops to recycle the nutrients at the surface. However, we should also consider the accessibility of the subsoil by roots, which is largely determined by the soil physical conditions at the transition layer (Peigné et al., 2013). Furthermore, the removal of residue at the surface under NT can modify nutrient cycling below the conventional soil sampling depth, and the

depletion of soil nutrients in the transition layer may make the roots concentrate at a shallow depth (Zhang and Barber, 1992). This could potentially reduce the plants' ability to utilize the available soil moisture and nutrients in the subsoil.

Silage and bioenergy production are some of the potential uses of removed maize residues from a farm. When considering the effects of the removal of crop residue, evaluation of soil biological, physical, and chemical properties below the surface layer is critical. We presented the mining of unfertilized nutrients, and also significantly lower concentrations of SOM related properties at soil layers > 18 cm under NT-Harv. Therefore, the removal of the residue may not be justifiable in the long-term when considering the sustainability of this cropping system.

6.6 CONCLUSIONS

This paper presents the importance of surface crop and soil management on surface (0-18 cm), transition (18-to-30 cm depth) and subsoil layer (30-to-60 cm depth) soil biological, chemical, and physical properties. We show that no-till (NT) combined with crop residue return (Ret) maintains soil conditions closest to the original continuous mixed sod, compared to plow till (PT) or residue harvested (Harv) treatments, across the soil profile. Crop residue return was important to avoid the depletion of macro- and micro- nutrients under NT below the surface layer, which emphasized the importance of a full soil profile framework in soil nutrient budgeting. For soil moisture, the accessibility of larger soil volume by crop roots as well as the reduction in evaporation by surface cover appeared more manageable compared to the total quantity of available soil water by tillage and residue management.

We conclude that the integrated assessment of surface, transition and subsoil layer soil conditions is important to understanding the effects of management. The direct impacts of tillage and residue management occur mostly near the soil surface, but have effects on soil properties deep into the profile, where no-tillage and residue return positively influence subsoil conditions. Sustainability of residue harvest needs to be evaluated at soil profile-scale since there may be hidden impacts on soil conditions especially under no-till.

6.7 ACKNOWLEDGMENTS

We are grateful to Daniel Moebius-Clune and Kirsten Kurtz for assistance in laboratory soil assessment, Erika Mudrak for advices in statistical analyses, and Michael Davis for maintaining the long-term field experiments.

CHAPTER 7: LARGE TOPSOIL ORGANIC CARBON VARIABILITY IS CONTROLLED BY ANDISOL PROPERTIES AND EFFECTIVELY ASSESSED BY VNIR SPECTROSCOPY IN A COFFEE AGROFORESTRY SYSTEM OF COSTA RICA

Rintaro Kinoshita, Olivier Roupsard, Tiphaine Chevallier, Alain Albrecht, Simon Taugourdeau, Zia Ahmed, and Harold M. van Es

Published in Geoderma, doi:10.1016/j.geoderma.2015.08.026

7.1 ABSTRACT

Assessing the spatial variability of soil organic carbon (SOC) is crucial for SOC monitoring and comparing management options. Topsoil (0-5 cm) SOC concentrations were surveyed in a coffee agroforestry watershed (0.9 km^2) on Andisols in Costa Rica with uniform farm management. We encountered high values and large spatial variations of SOC, from 48.1 to 172 g kg^{-1} in the dry combustion set (SOC_{ref}; n=72) used for calibrating the visible-near-infrared reflectance spectroscopy (VNIRS) samples (SOC_{VNIRS}; 350 – 2500 nm; n = 520). VNIRS using partial least squares regression was effective in predicting SOC (R^2 = 0.85*;* a root mean square error (RMSE) = 12.3 g kg^{-1}) and proved an effective proxy measurement. We assessed several topographic, vegetation and andic soil property variables, of which only the latter (metal-humus complexes and allophanes) displayed strong correlations with SOC_{ref} concentrations. We

compared Random Forest and three geostatistical approaches for the interpolation of SOC in unsampled locations. Ordinary kriging with SOC_{ref} yielded an RMSE of 28.0 g kg^{-1}. Random Forest was successful in incorporating many weakly and non-linearly correlated covariates with SOC (RMSE = 14.7 g kg^{-1}), provided Al_p (the sodium pyrophosphate extractable aluminum), the best predictor of SOC (*r = 0.85*) but also the most costly variable to acquire. Co-kriging with Al_p also showed high reduction in RMSE (16.0 g kg^{-1}). Co-kriging with SOC_{VNIRS} only showed marginal reduction in RMSE to 24.2 g kg^{-1} due to the presence of a high nugget effect.

Local variability of SOC in this volcanic agroforestry watershed was dominated by andic properties whereas topographic or vegetation variables had very little impact. Estimation of SOC variability is recommended using inexpensive proxy measurements like VNIRS (RMSE = 12.3 g kg^{-1}) rather than spatial interpolation techniques.

7.2 KEY WORDS

Allophane; Agroforestry; Andisols; Co-kriging; Random Forest; Soil organic carbon; VNIR spectroscopy

7.3 INTRODUCTION

Soil organic carbon (SOC) is a fundamental property related to soil physical, chemical and biological quality and is an important component of the global carbon (C) cycle (Magdoff and van Es, 2009). Disruption of sustainable C cycles in agricultural soils has led to diminishing crop yields as well as contributing to further accelerating greenhouse gas (GHG) emissions (Hillel and Rosenzweig, 2010; Lal, 2006; Powlson et al., 2011). At a farm-scale, high spatial

variation of SOC may occur, which causes uncertainty when comparing several management practices or when assessing the effectiveness of various soil conservation measures to restore SOC (Minasny et al., 2013).

There is need for accurate approaches to assess the impact of management on SOC at the farm-scale, whatever the inherent variability. Various biotic and abiotic variables have been identified to correlate with SOC at various spatial scales and soil environment, such as past and present land use (Schulp and Veldkamp, 2008), local terrain (Cambule et al., 2014; Thompson and Kolka, 2005), and vegetation (Bou Kheir et al., 2010; Horwath Burnham and Sletten, 2010; Kunkel et al., 2011; Takata et al., 2007). These correlated variables have been used to predict SOC through various methods such as multiple linear regression (Gessler et al., 2000; Thompson and Kolka, 2005), Random Forest (RF; Grimm et al., 2008), boosted regression tree (Razakamanarivo et al., 2011), co-kriging (Terra et al., 2004) and regression kriging (Bilgili et al., 2011; Hengl et al., 2004; Simbahan et al., 2006). Any suitable approach may need to be determined in each environment according to the availability of information.

In recent years, visible-near-infrared reflectance spectroscopy (VNIRS), a rapid and cost effective proximal soil sensing method (Barthes et al., 2006; Brunet et al., 2007; Kinoshita et al., 2012) has been used to predict SOC (Bellon-Maurel and McBratney, 2011). It can be used in laboratory or in field by first calibrating the reflectance data with analytically measured SOC information through multivariate statistical or data mining techniques, and predicting the SOC content of a new sample only with the spectra (Barthes et al., 2006; Viscarra Rossel et al., 2006). The predicted SOC contents can be used in geostatistical techniques by directly replacing analytically measured SOC information in ordinary kriging (Lamsal, 2009), or as a covariate in

co-kriging and regression kriging as demonstrated in studies using soil organic matter (Bilgili et al., 2011).

Coffee agroforestry is an environmentally and economically important agricultural system in Central America, where Arabica coffee (*Coffea arabica L., var Caturra*) is grown under shade trees on the slopes of volcanoes at high elevations (Somarriba et al., 2012). Shade trees are known to improve the size and the quality of coffee beans by buffering unfavorable climatic conditions (Muschler, 2001) and also result in reduced soil erosion and compaction while increasing SOC (Beer et al., 1998). However, past research has shown extremely localized effect (< 1m) of the shade trees on SOC (Payán et al., 2009), which may not be detectable when soil samples from shaded and non-shaded sites are composited (Noponen et al., 2013).

Many of the coffee growing regions in Central America are associated or dominated by Andisols (USDA-NRCS, 2005), which are often evaluated as the most productive arable soils in the tropics especially where their parent material is basaltic (Shoji et al., 1993). Andisols are known to have substantial SOC sequestration potential, containing 1.8 % of global total soil C stocks while covering only 0.7 % of global ice-free land (Batjes, 1996; Hillel and Rosenzweig, 2010). They contain non-crystalline clay amorphous minerals that originated in volcanic ejecta (Parfitt, 1990) but the mechanisms of SOC stabilization is still debated. Past research has suggested several possible pathways such as Al toxicity against microbial degradation (Boudot, 1992), formation of organo-mineral complexes, physical protection by stabilizing soil aggregates (Huygens et al., 2005) or the fractal structure of allophane (Chevallier et al., 2010). Variable volcanic ash input, mass erosion, soil pH, and moisture regimes cause high spatial variation in the degree of their crystallization and andic soil property distributions (Chesworth, 2008; Nanzyo et al., 1993), leading to high spatial variability of SOC in Andisols. At the landscape scale

(140,000 ha), Powers and Schlesinger (2002) have assessed the relationships between soil C and various biotic and abiotic variables across a large elevation gradient (50-to-750 m). They identified elevation and Normalized Difference Vegetation Index (NDVI) to correlate to soil C across the elevation gradient while andic soil properties (allophane and Al-humus complex) showed strong correlation within a similar elevation site. Nevertheless, farm scale variability of SOC on Andisols has rarely been assessed to date and a need exists to explore correlated biotic and abiotic variables at this scale.

This study was conducted in a micro-watershed on Andisols, covered by Arabica coffee agroforestry, a major land use in Central America (Somarriba et al., 2012). The main objectives of this study are 1) to assess the dominant attributes among topographic, vegetation and andic soil properties influencing farm-scale spatial variability of topsoil SOC on Andisols and 2) to determine optimal feature-space and geostatistical interpolation models to predict SOC concentrations at unsampled sites with available predictors.

7.4 MATERIALS AND METHODS

7.4.1 Site description

The research site is located in the Central-Caribbean area of Costa Rica. It is part of the Reventazón River Basin on the slope of the Turrialba volcano, which eventually drains into the Caribbean Sea. The research site is located within the Aquiares Coffee Farm (6.6 km^2) called the Mejias creek watershed (Figure 7.1a) located between 83° 44’ 39” and 83° 43’ 35”W, and

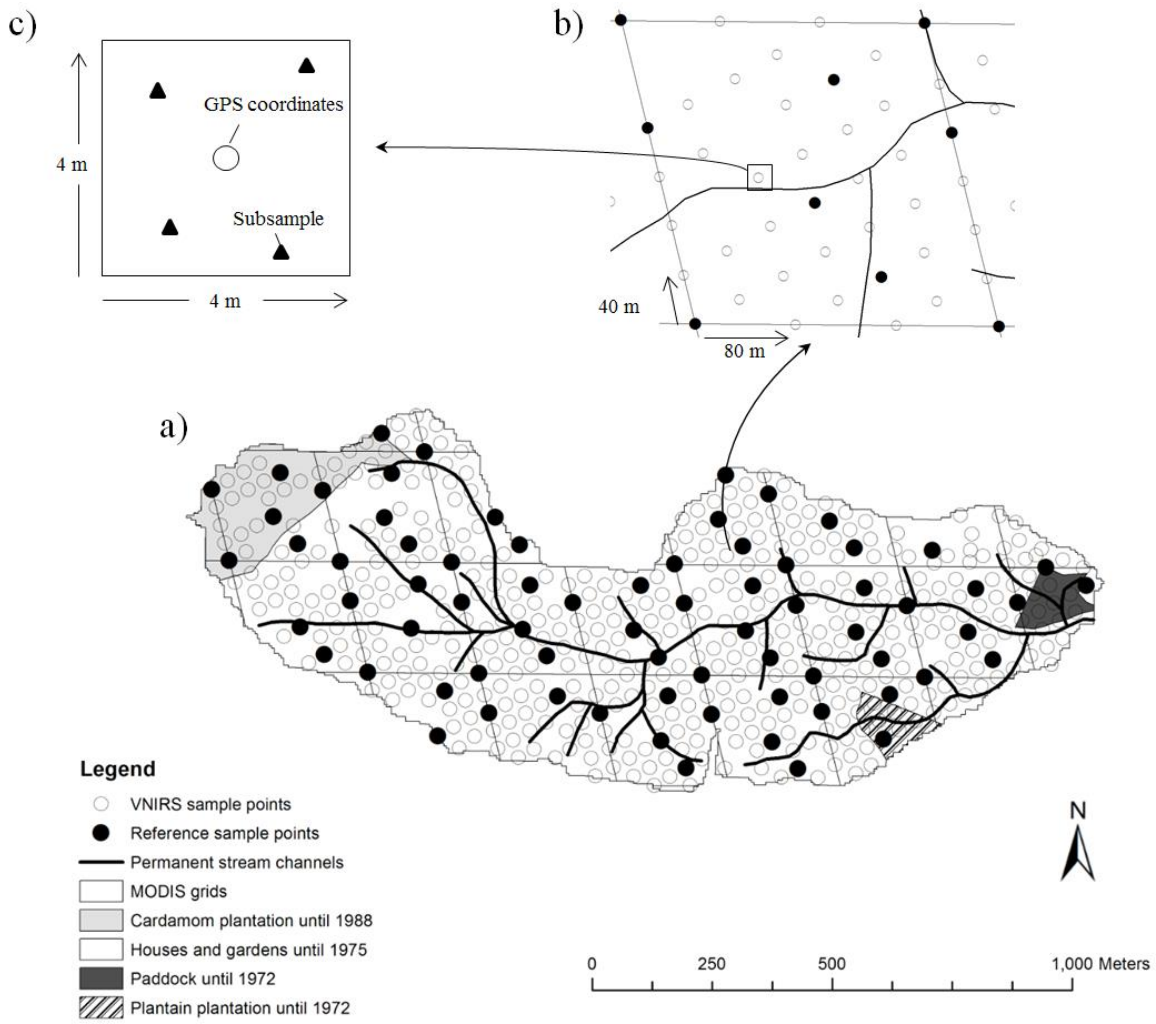

Figure 7.1 Maps of a) the experimental watershed with indications of past land use, b) the sampling locations within each MODIS grid for the reference (n =72) and visible-near-infrared reflectance spectroscopy (VNIRS) sample sets, and c) the configuration of subsamples at each sampling location.

between 9° 56’ 8” and 9° 56’ 35” N. The watershed has an area of 0.9 km^2 with elevation ranging from 1020 to 1280 m a.s.l. and a mean slope of 11.3° (Gómez-Delgado et al., 2011). The climate is tropical humid without a dry season and influenced strongly by the Caribbean Sea (Peel et al., 2007). The mean annual precipitation from 1973 to 2009 was 3014 mm.

According to Mora-Chinchilla (2000), volcanic avalanche deposits form the geology of the area, which was originally produced by the collapse of a 1.3 km wide strip of the southeastern slope of the Turrialba volcano. Indications of lava flows, agglomerates, lahars and ashes are also present. The soils are classified as Andisols and are generally characterized by their high SOC contents, infiltration capacities, and biological activities (Payán et al., 2009). The superficial runoff in the study site is very low even on slopes, and the sediment production by the whole watershed is *ca.* 1 t ha^{-1} yr^{-1} (Gómez-Delgado, 2010; Gómez-Delgado et al., 2011). The steady state infiltrability has been measured as high as *ca.* 1000 mm h^{-1} (Benegas et al., 2014), confirming that the hydrology is dominated by infiltration.

Before the introduction of coffee, the western side of the watershed was under a cardamom (*Elettaria cardamomum*) plantation until 1988 while most of the other parts were under housing and gardens until 1975 (Figure 7.1a). The vegetation is now a mixture of homogeneously plantedcoffee trees and *Erythrina poeppigiana* leguminous shade trees on bare soils. Shade trees have a density of 7.4 trees ha^{-1}, with 15.7 % ± 5.5% canopy projection and an average canopy height of *ca.* 20 m (Taugourdeau et al., 2014). The initial planting density of the coffee trees was 6,300 locations ha^{-1}, which have been selectively pruned over more than 30 years (old exhausted re-sprouts aged > 5-6 years pruned every year in March, representing ca. 15% of total re-sprouts, resulting in minimum leaf area index (LAI) in March). The canopy openness of the coffee is *ca.* 25% and the average canopy height *ca.* 1.2 m. The Aquiares farm is

managed quite intensively (upper conventional mode) in terms of fertilizer application (mean = 214 kg N ha^{-1} yr^{-1}; standard deviation (SD) = 44 kg N ha^{-1} yr^{-1}; 2000-2012), and complies with Rainforest Alliance™ guidelines for its pest and weed management. Weeds are scarce and it is assumed in this research that the soil surface is uncovered except for litter. The yields of green coffee from 1994 to 2011 averaged 1375 kg ha^{-1} yr^{-1} (SD = 341 kg ha^{-1} yr^{-1}).

7.4.2 Soil sampling strategy

In total, 520 soil samples were systematically collected for VNIRS analysis in the 0.9 km^2 watershed (5.8 samples ha^{-1}) in July-August 2010 from the 0-to-5 cm depth using a push probe after the removal of surface residues. We decided to oversample the study site systematically within an achievable timeframe and budget, because no prior knowledge on the correlation between environmental covariates and SOC could guide sampling (Minasny et al., 2013; Walvoort et al., 2010). The watershed was first divided into 23 grids aligned with the MODIS satellite image (20 nearly-full MODIS pixels, Figure 7.1a). Within each MODIS pixel, 6 transects were drawn E-W and 12 N-S (Figure 7.1b) and the sampling points were placed on the intersections, following a triangular grid pattern (Gilbert, 1987). The resulting points were 80 m and 40 m apart in E-W and N-S directions respectively accounting for higher spatial density following larger N-S slopes, following the concave shape of the study site. Some sampling points were not accessible due to watercourses, roads, and dense vegetation along the rivers and omitted from the sample set.

Among the 520 points, every 8^{th} sample was selected systematically and subjected to additional soil tests (Fig. 1a and b; "reference sample points"). We opted for systematic sampling in order to maximize the chance to capture the overall farm-scale spatial variation of SOC *a*

priori. At each sampling point, four subsamples were taken for the VNIRS set (Figure 7.1c), plus four additional subsamples for the reference sample set, as the latter required a higher sample volume for the analytical soil tests. The subsamples were taken randomly within a 4 m by 4 m square centered around each GPS coordinate in the inter-row of coffee plants. The subsamples were thoroughly mixed in a plastic bag. The presence of the shade trees was recorded when the sampling point occurred under the drip-line.

The collected soil samples were oven-dried at 40°C for 72 hours and ground using a mortar and pestle to pass through a 2 mm sieve. Visible roots were removed to avoid any effects on the soil reflectance spectra.

7.4.3 Laboratory analyses

The 72 reference samples were used for the analytical soil assessment. All measured data are reported on a 105 °C oven-dry basis. Total C and N concentrations were analyzed by dry combustion using a LECO TruSpec CHN analyzer (LECO, St Joseph, MI) and the total C concentrations were equal to SOC concentrations due to the absence of carbonates in the samples. pH was assessed in a 1:1 soil and water suspension using a bench-top pH meter (Thermo Fisher Scientific, Beverly, MA). Acid-oxalate extractable Al (Al_o), Fe (Fe_o), and Si (Si_o) were determined after the extraction of soil nutrients in the dark using a 0.2 M ammonium oxalate solution followed by the measurement using inductively coupled plasma emission spectrometry (ICP-ES; Pansu and Gautheyrou, 2006; Soil Survey Staff, 2014). This extractant is effective in dissolving non-crystalline clay minerals such as allophane, imogolite, ferrihydrite as well as metal-humus complexes. Sodium pyrophosphate extractable Al (Al_p) and Fe (Fe_p) were also measured by ICP-ES after extraction with a 0.1 M sodium pyrophosphate solution. The extracted

solution (1 ml) was acid digested on a hotplate with 1 ml 40 % nitric acid and 3ml 60 % perchloric acid in order to get clear extracts. The acid digestion was a more convenient method compared to the standard high-speed centrifugation and micropore filtration while still gaining an almost perfect one-to-one relationship (R^2 = 0.98; RMSE = 0.14 %) to the standard procedure. Sodium pyrophosphate extraction is only effective in dissolving metal-humus complexes and it indicates the varying effects of metals in stabilizing soil C.

Andisols are characterized by the high concentrations of Al_o, which generally represent poorly crystalline aluminosilicates. The value of Al_o+0.5Fe_o has also been used as a criterion to distinguish andic properties in soil taxonomy ($Al_o + 0.5Fe_o > 20$ g kg^{-1}; (ISSS-ISRIC-FAO, 1998). Allophane content was calculated using the formula by Mizota and Van Reewijk (1989):

$$\text{allophane (\%)} = 100\text{Sio}/[23.4-5.1(\text{Alo}-\text{Alp})/\text{Sio}] \qquad (1)$$

7.4.4 Digital elevation model and leaf area index

Topographic properties and LAI (m^2_{leaf} m^{-2}_{soil}) were obtained for the entire study site using remotely sensed data. The digital elevation model (DEM) was created by digitizing the 5 m elevation contour interval and 2 m horizontal resolution map from the TERRA-1998 project (scale ≈ 1: 25000; CENIGA, 1998). This was used to calculate percent slope gradient (SLOP), combined curvature (CURV), profile curvature (PROF), plan curvature (PLAN), slope aspect in degrees azimuth (ASP) and topographic position indices (TPI) that show the relative topographic position compared to the predetermined neighborhood (Gallant and Wilson, 2000) using ArcGIS 10.2 (ESRI, Redlands, CA). Compound topographic indices were also derived (CTI; Evans and Oakleaf, 2011), which are also known as steady state wetness index and generally show the zones of soil water saturation within a landscape (Moore et al., 1991). Leaf area index is a key

indicator of plant physiological activity related to the C cycle through processes such as photosynthesis and net primary production. In addition, LAI can indicate the canopy contribution to rainfall interception and litter production, which both contribute to soil C variation. The LAI data were collected and presented by Taugourdeau et al. (2014) where detailed methods are described. Briefly, a satellite image acquired in March 2010 (drier season, minimum cloud cover and low LAI) from Worldview2 (DigitalGlobe, Longmont, CO), with 2 m resolution multispectral and 0.5 m resolution panchromatic images was utilized. Normalized Difference Vegetation Index (NDVI) was first obtained and converted to LAI through multiple calibration steps using two hand-held LAI-2000 devices (LI-COR Environmental, Lincoln, NE) as well as an actual LAI measurement by collecting 60,000 leaves. The leaves were selected based on systematic sampling and they were measured non-destructively on live plants through allometric relationships (Taugourdeau et al., 2014).

We have utilized averaged values for the topographic variables and LAI at a 12 m by 12 m cell in order to account for the GPS accuracy of ± 4m used for the soil sampling.

7.4.5 Visible-near-infrared reflectance spectroscopy

The VNIRS analysis was conducted for the spectra of 350 nm to 2500 nm at 10 nm spectral resolutions using a Fieldspec Pro hyperspectral sensor (Analytical Spectral Devices, INC., Boulder, CO). Oven-dried samples were placed in a 4 cm diameter optical quality Petri dish, and spectral reflectance was collected through the glass bottom at a constant angle (55 degrees from horizontal) from a distance of 4 cm, in an enclosed box. The prediction models were built using Unscrambler 10.2 (CAMO software, Oslo, Norway). We first smoothed the reflectance data using first-order derivatives of the Savitsky-Golay transformation function,

which is suitable for VNIRS applications because it removes noise from the data while retaining the original spectral characteristics with minimum distortion (Ruffin and King, 1999). A calibration model for predicting SOC concentrations was constructed utilizing partial least squares regression (PLSR; Geladi and Kowalski, 1986) with the reference sample set (n = 72), after log transforming the measured SOC concentrations (Kusumo et al., 2010; Kinoshita et al., 2012). It relates two variables, X (spectral readings) and Y (measured SOC concentrations) by a linear multivariate model. Orthogonal and weighted linear combinations of the spectral readings are used for predicting each Y variable. Partial least squares regression is suited to handle data with strong collinearity in the X variables, which are usually more numerous than the observations (Y) that they predict. The selection of the number of factors to include was critical to avoid over- or under-fitting of the model. Over-fitting of calibration models reduces their ability to predict SOC concentrations of new unknown soil samples. In this study, additional factors were only added if they reduced the total residual Y-variance at cross validation by the percentage equivalent to the number of factors (Kinoshita et al., 2012). Also, the maximum number of factors was restricted to 15 to prevent over-fitting. Spectral outliers were identified in the score plot of the first and the second factors using the 95% confidence ellipse (Hotelling T^2; Mouazen et al., 2010). Sample scores lying outside the plot were considered as spectral outliers and we assessed the prediction improvement by removing the outliers. The validity of the PLSR method was determined by cross validation using the leave-one-out method.

Following the validation of the PLSR method, SOC concentrations of the VNIRS sample set (n = 520) were predicted using the log-transformed model and then back transformed using the unbiased formula for further analysis (Yang, 2012).

7.4.6 Correlations among measured variables

We assessed a scatter matrix of our dataset and found non-linear associations among the measured variables. Therefore, Spearman rank correlation coefficients were calculated between SOC concentrations and all the other measured variables to elucidate the dominant factors controlling the spatial variation of SOC concentrations for the reference sample set (n = 72). This is a more robust method compared to the Pearson correlation when the sample set comes from non-normal distributions or when a non-linear relationship is observed (Seibert et al., 2007).

7.4.7 Random Forest

Random Forest was used as a feature-space (non-spatial) prediction method to interpolate SOC in unsampled locations using the measured covariates of the reference sample set. It is an extension of randomized classification and regression trees (CART; Breiman et al., 1983) with bagging of a set number of trees (Breiman, 1996). This method has several advantages over other statistical models (Breiman, 2001; James et al., 2013) such as its ability to model high dimensional non-linear relationships in mixed categorical and continuous predictors (Grimm et al., 2008). It is also resistant to overfitting and internally calculates unbiased measure of estimation error (James et al., 2013), which eliminates the need to undertake a separate cross validation.

We have assessed two separate scenarios in predicting SOC concentration using the *randomForest* package (Liaw and Wiener, 2002) in the R statistical computing environment. The first scenario did not include any andic soil properties ($RF_{reduced}$), considering that they are time-consuming and costly to measure, while the second scenario included all available covariates

(RF_{full}). Random Forest can restrict the number of predictors to utilize at each node when a split in tree is considered (m_{try}) in an effort to create uncorrelated trees to maximize variance reduction (James et al., 2013). In order to determine an optimum number of m_{try} for each scenario, we applied the RF models 100 times for each possible m_{try} value and calculated the mean RMSE. We used the number of trees (n_{tree}) of 1000 and the nodesize of 5 for this analysis. We then chose the m_{try} with the minimum mean RMSE for each scenario.

7.4.8 Geostatistical methods

The *gstat* package (Pebesma, 2004) in the R statistical computing environment was used for all geostatistical modeling. Ordinary kriging and two co-kriging approaches were used to interpolate SOC concentrations at unsampled locations. Ordinary kriging relies on a semivariogram that represents the variance of the linear increments of a variable under the assumption of intrinsic stationarity, and variogram models are then used to assign weights to neighboring points for the kriging function (Webster and Oliver, 2007). Both geometric and zonal anisotropy were tested, which incorporate directional effects on the spatial structure of a soil property and was assessed for the semivariogram due to the concave shape of the watershed. Co-kriging relies on a cross-variogram and was fitted using the linear model of co-regionalization that fits a single spatial structure to both direct and cross-variogram by optimizing the partial sills and the nugget by least squares (Rossiter, 2012). Any negative eigenvalues were set to zero and the eigenvectors were recomputed to ensure positive semidefinite matrices of the partial sills and the nugget (Pebesma, 2004). This ensures a valid linear model of co-regionalization, and is a quick approximation to iterative methods of adjusting partial sills and nuggets. Co-kriging methods were attractive in our study because they allow for

the effective incorporation of point-measured covariates if both feature space- and spatial-correlations can be verified. Two different co-kriging approaches were tested. The first co-kriging method (CK(Al_p)) used the most correlated analytically measured covariate available at the same locations as the reference sample set (n = 72) while the second co-kriging method (CK(SOC_{VNIRS})) incorporated the VNIRS estimated SOC concentrations (SOC_{VNIRS}; n = 520). However, the same cokriging model was used in both cases and this model is built only from co-located observations.

7.4.9 Prediction performance

The accuracy of the PLSR model for SOC concentration prediction was evaluated using the coefficient of determination (R^2) of the measured and the model-predicted values in the leave-one-out cross validation. Root mean square error of prediction (RMSE) was also determined, which is a measurement of accuracy calculated as the differences between model predicted values and measured values:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_{pred.i} - Y_{meas.i})^2}{n-1}} \quad (2)$$

The RMSE is the combination of bias and residual variance, the latter being the deviation from the 1:1 line of predicted values against measured values (Bellon-Maurel et al., 2010):

$$RMSE = \sqrt{Bias^2 + MSEc} \quad (3)$$

where MSEc is the mean square error of prediction corrected for bias.

Bias can be calculated as:

$$Bias = \bar{Y}_{pred.} - \bar{Y}_{meas.} \quad (4)$$

which is the difference between the mean of predicted values and the mean of measured values. Mean square error of prediction corrected for bias can be calculated as the quadratic sum of the differences between predicted values corrected for bias and measured values:

$$MSEc = \sum_{i=1}^{n} \frac{\left(Y_{pred.\mathrm{i}} - Bias - Y_{meas.\mathrm{i}}\right)^2}{n-1} \quad (5)$$

which describes the dispersion around the 1:1 line. Bias can be caused by systematic errors due to instrumental and methodological errors while the residual variance relates to random errors that can be improved by averaging the predicted values (Bellon-Maurel et al., 2010). The decomposition of RMSE is therefore important to understand the origin of errors.

Ratio of performance to interquartile distance (RPIQ) is a useful performance measure for comparing PLSR models from different datasets even when they are not normally distributed (Bellon-Maurel et al., 2010):

$$RPIQ = \frac{(Q3 - Q1)}{RMSE} \quad (6)$$

where Q1 and Q3 are the 1st and the 3rd quartiles respectively.

The prediction error of RF was calculated by predicting the SOC concentration of each out-of-bag (OOB) sample by its corresponding bootstrap training tree and subtracted from the measured value (Grimm et al., 2008). The root mean square error of prediction for the OOB samples was:

$$RMSE^{OOB} = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i^{OOB}\right)^2}{n}} \quad (7)$$

Kuhn and Johnson (2013) have shown that $RMSE^{OOB}$ results to present identical values compared to RMSE values predicted using cross validation across a range of m_{try} values.

The spatial prediction errors of SOC concentrations associated with the geostatistical methods were assessed using the leave-one-out cross validation at the 72 reference sample points.

For co-kriging we also performed the leave-one-out cross validation at the 72 reference sample points. For co-kriging with Al_p ($CK(Al_p)$), we assumed the covariate to be available at the validation locations and they were not taken out in each cross validation fold. For co-kriging with SOC_{VNIRS} ($CK(SOC_{VNIRS})$), we removed both the primary variable and the covariate information at the prediction location for each cross validation fold since we utilized the primary variable to build the PLSR model for the prediction of the covariate information. This ensured the independence of this cross validation procedure from the PLSR prediction model.

7.5 RESULTS AND DISCUSSION

7.5.1 Soil organic carbon and total N concentrations at the reference sample points

The range of measured SOC in the reference set (n=72) at 0-to-5 cm depth was from 48.1 to 172 g kg^{-1} with a mean of 94.8 g kg^{-1}, comparable values to the study conducted at a similar elevation in the Turrialba catena (Buurman et al., 2007). Soil organic carbon and total N concentrations displayed high variation within the reference sample set and the coefficient of variation (CV) were 33 and 29 % respectively. High variation of SOC at plot or field scale is common across different environmental systems, though the magnitude of variation is dependent on edaphic and environmental factors (Cambardella et al., 1994; Conant et al., 2003). Chesworth (2008) noted that Andisols commonly have SOC concentrations of more than 60 g kg^{-1} in both A and B horizons, and SOC concentrations can reach 200 g kg^{-1} under humid conditions. In

addition, allophanic Andisols can contain up to 150 g kg^{-1} SOC, while non-allophanic Andisols can accumulate up to 230 g kg^{-1} SOC in A_p, A and buried A horizon (Nanzyo et al., 1993). The C:N ratio was relatively constant across the watershed (mean = 10.9; SD = 0.67) indicating relatively homogeneous quality of organic inputs and minimal effects of leguminous shade trees.

7.5.2 Relationships among edaphic, vegetation, and topographic properties

Soil pH ranged widely from 4 to 6 (Table 7.1) and we found a patchy spatial distribution (Figure 7.2a; inverse distance weighting), possibly due to sporadic calcium carbonate applications to adjust soil pH according to the farm's soil management strategy (0 to 2760 kg $CaCO_3$ ha^{-1} yr^{-1}; L.G. Ramírez, personal communication). The ranges of allophane and Al_p concentrations were large with the CV of 50% and 26 % respectively.

The andic soil properties showed a strong correlation to SOC concentrations (Table 7.2). The stronger correlation was found between Al_p and SOC (r = 0.85), suggesting the occurrence of metal-humus complexes. This was in line with a larger sample set analyzed by Nanzyo et al. (1993) where they also found a high correlation between the two soil properties ($r = 0.84$). In our study, the correlation between SOC and allophane was relatively strong (r = 0.52; Table 7.2).

Table 7.1 Descriptive statistics of edaphic, andic, topographic, and vegetation variables at the reference sample points (n =72) and visible-near-infrared reflectance spectroscopy predicted soil organic carbon (SOC) values (n = 520).

	n	Min	Max	Mean	Median	1st quart.	3rd quart.	SD	Skewness	Kurtosis
Edaphic properties										
SOC_{ref}† (g kg^{-1})	72	48.1	172	94.8	91.8	71.6	113	31.5	0.55	-0.48
SOC_{VNIRS} (g kg^{-1})	520	16.8	187	85.9	84.4	62.5	107	28.6	0.34	-0.39
Total N (g kg^{-1})	72	4.9	14.3	8.6	8.4	6.64	10.6	2.5	0.34	-0.77
C:N	72	9.6	13.4	10.9	10.8	10.5	11.3	0.67	0.91	1.22
pH	72	4.0	6.0	5.1	5.0	4.8	5.5	0.47	0.11	-0.95
Andic soil properties										
Al_o (g kg^{-1})	72	12.6	74.8	43.2	43.6	30.1	56.0	17.0	0.02	-1.11
Al_p (g kg^{-1})	72	9.6	37.3	23.1	23.0	19.3	26.1	6.1	0.16	-0.12
Fe_o (g kg^{-1})	72	10.5	30.7	22.9	23.6	20.4	26.1	4.2	-0.64	0.14
Si_o (g kg^{-1})	72	2.7	36.3	18.1	18.3	11.1	24.4	8.2	-0.12	-1.01
Al_p/Al_o	72	0.33	1.07	0.59	0.54	0.44	0.72	0.18	0.60	-0.69
Allophane (%)	72	1.3	20.1	10.3	5.7	5.9	14.6	5.2	1.53	3.34
Topographic variables										
ELEV (m)	72	1014	1282	1135	1125	1070	1203	71.4	0.18	-1.1
SLOPE (%)	72	0.71	82.5	23.7	21.0	11.5	33.0	14.9	0.79	0.19
CURV (-)	72	-6.7	8.5	0.0	0.01	-0.55	0.54	1.2	0.02	3.1
PROF (-)	72	-4.8	4.2	0.02	0.03	-0.33	0.37	0.83	-0.11	2.5
PLAN (-)	72	-4.3	3.7	0.02	0.04	-0.19	0.28	0.55	-0.75	6.8
TPI (-)	72	-3.0	2.9	0.0	0.01	-0.28	0.28	0.57	-0.03	2.1
CTI (-)	72	4.0	16.7	7.4	7.2	6.4	8.1	1.5	1.2	3.4
Vegetation variable										
LAI (-)	72	0	7.0	2.5	2.4	1.4	3.5	1.5	0.29	-0.49

† SOC_{ref}, soil organic carbon measured by dry combustion; SOC_{VNIRS}, soil organic carbon estimated by ASD Fieldspec Pro hyperspectral sensor using partial least squares regression; Al_o, aluminum extracted by ammonium oxalate; Al_p, aluminum extracted by sodium pyrophosphate; Fe_o, iron extracted by ammonium oxalate; Si_o, silica extracted by ammonium oxalate; ELEV, elevation; SLOP, slope gradient; CURV, combined curvature; PROF, profile curvature; PLAN, plan curvature; TPI, topographic position index; CTI, compound topographic index; LAI, leaf area index

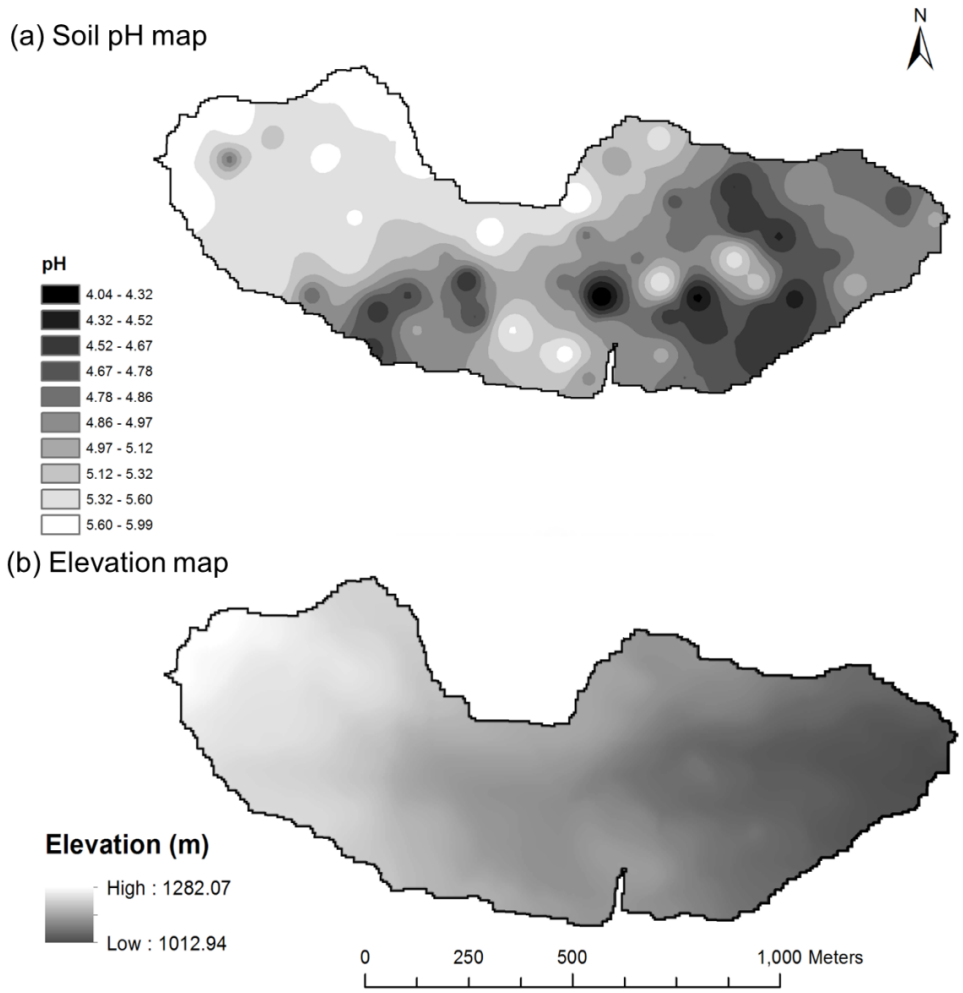

Figure 7.2 Range maps of a) soil pH (inverse distance weighting) and b) elevation of the experimental watershed.

Allophane is known to physically protect SOC against biodegradation in its mesoporic structures where the oxygen diffusion and soil organic carbon accessibility to microorganisms are low (Chevallier et al., 2008, 2010). Al hydroxide and Al-humus complexes are known to have stronger protective capacities against SOC degradation compared to allophane (Boudot, 1992) and their higher correlation to SOC indicates the importance of Al-humus complexes in our study site. This is in line with the study conducted by Powers and Schlesinger (2002) on Andisols in Costa Rica where they established the importance of Al-humus complexes when samples from a similar elevation range are compared.

There were also complex interactions among SOC, soil pH, allophane, Al_p, and elevation in our study site (Table 7.2). It has been shown that lower pH values release Al from allophane and promote the formation of Al-humus complexes (Huygens et al., 2005). However, the presence of both a short-term pH regulator like lime applications and a long-term regulator such as base cation leaching complicates the analysis using a simple correlation. There was a weak negative correlation between Al_p and CTI but we did not observe other statistically significant relationships between andic soil properties or SOC to available data of topography or vegetation at the sampling distance. The presence of shade tree showed no statistically significant effects on SOC assessed by Welch's two sample t-test (data not shown). The LAI data were obtained at a very high spatial resolution in March when the leaf area was at its yearly minimum after drier season and selective pruning of coffee and might not be representing the interactions between the vegetation cover and the SOC. We also used monthly MODIS LAI at a coarser resolution but we could not evidence any clear seasonal LAI effect either at that scale (data not shown). A recent study on coffee in the same region has shown that SOC poorly correlates with above-ground biomass stocks (Noponen et al., 2013). Overall, it appears that spatial variation of SOC was

Table 7.2 Spearman correlation coefficients for edaphic, andic, topographic, and vegetational variables at the reference sample points (n = 72).

	SOC	Total N	C:N	pH	Alp	Alp/Alo	Allophane	ELEV	SLOP	CURV	PROF	PLAN	ASP	TPI	CTI	LAI
SOC†	1.00															
Total N	**0.99**‡	1.00														
C:N	**0.57**	**0.45**	1.00													
pH	**0.28**	**0.28**	ns	1.00												
Alp	**0.85**	**0.83**	**0.53**	ns	1.00											
Alp/Alo	**-0.45**	**-0.46**	ns	**-0.31**	**-0.25**	1.00										
Allophane	**0.52**	**0.52**	ns	**0.35**	**0.34**	**-0.51**	1.00									
ELEV	0.20	ns	0.22	**0.55**	ns	ns	ns	1.00								
SLOP	ns	-0.23	ns	-0.23	ns	ns	-0.21	ns	1.00							
CURV	ns	ns	ns	ns	ns	ns	ns	ns	ns	1.00						
PROF	ns	ns	ns	ns	ns	ns	ns	ns	ns	**-0.84**	1.00					
PLAN	ns	ns	ns	ns	ns	ns	ns	ns	ns	**0.73**	**-0.35**	1.00				
ASP	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	1.00			
TPI	ns	ns	ns	ns	ns	ns	ns	ns	ns	**0.99**	**-0.82**	**0.75**	ns	1.00		
CTI	ns	ns	**-0.28**	ns	**-0.23**	ns	ns	ns	**-0.48**	**-0.35**	ns	**-0.58**	0.20	**-0.37**	1.00	
LAI	ns	ns	ns	**-0.30**	ns	ns	ns	-0.23	ns	ns	ns	ns	ns	ns	ns	1.00

† SOC, soil organic carbon measured by dry combustion; Al_p, aluminum extracted by sodium pyrophosphate; ELEV, elevation; SLOP, slope gradient; CURV, combined curvature; PROF, profile curvature; PLAN, plan curvature; ASP, slope aspect; TPI, topographic position index; CTI, compound topographic index; LAI, leaf area index

‡ Bold types are significant for $p < 0.05$ and the non-bold types are significant at $p < 0.1$.

largely controlled by andic soil properties in our study site, while vegetation and topographic variables were of secondary or low importance.

7.5.3 Visible-near-infrared reflectance spectroscopy analysis of soil organic carbon

The PLSR prediction model for the reference sample set (n = 72) was validated using the leave-one-out cross validation method and showed an RMSE of 12.3 g kg^{-1} (Table 7.3; Figure 7.3). A past study with an Andisol dataset showed an RMSE of 28.0 g kg^{-1} for soil total carbon estimation using PLSR with independent validation (McDowell et al., 2012). In our dataset, most of the RMSE associated with the model originated in the MSEc and very low bias existed (0.139 g kg^{-1}; Table 7.3). This indicates the presence of minimal systematic errors associated with the instrument, soil preparation, and the measurement method (Bellon-Maurel et al., 2010). The MSEc is caused by random errors and can be partly improved by increasing the number of observations. In our study, 100 spectral readings from two angles were taken and averaged per sample but there were no physical replicated samples at each location. The errors were particularly high (± 30 g kg^{-1}) for low SOC concentrations (< 52 g kg^{-1}; Figure 7.3). This is commonly observed when calibration sample sets show a form of Gaussian distribution (Bellon-Maurel et al., 2010), which leads to the underrepresentation of high or low SOC concentrations . There were two spectral outliers identified with the Hotelling T^2 values at the 95 % significance level but they were not removed from the analysis since it did not improve prediction.

Table 7.3 Results of a partial least squares regression (PLSR) model for predicting soil organic carbon with leave-one-out cross validation from reflectance spectra obtained with an ASD Fieldspec pro hyperspectral sensor (n=72).

	F†	R^2	Min	Max	Mean	1st quart.	3rd quart.	SD	MSE	RMSE	Bias^2	Bias	MSEc	RPIQ
SOC_{VNIRS}‡ (g kg^{-1})	6	0.85	35.3	187	95.1	74.2	114	30.4	151	12.3	0.019	0.139	151	3.21
SOC_{ref} (g kg^{-1})			48.1	172	94.8	71.6	113	31.5						

† F, number of factors used in PLSR; Min, minimum; Max, maximum; 1st quart, 1st quartile; 3rd quart, 3rd quartile; SD, standard deviation; MSE, mean square error of cross validation; RMSE, root mean square error of cross validation; MSEc, mean square error corrected for bias; RPIQ, ratio of performance to interquartile distance.

‡ SOC_{VNIRS}, soil organic carbon estimated by ASD Fieldspec Pro hyperspectral sensor using PLSR; SOC_{ref}, soil organic carbon measured by dry combustion.

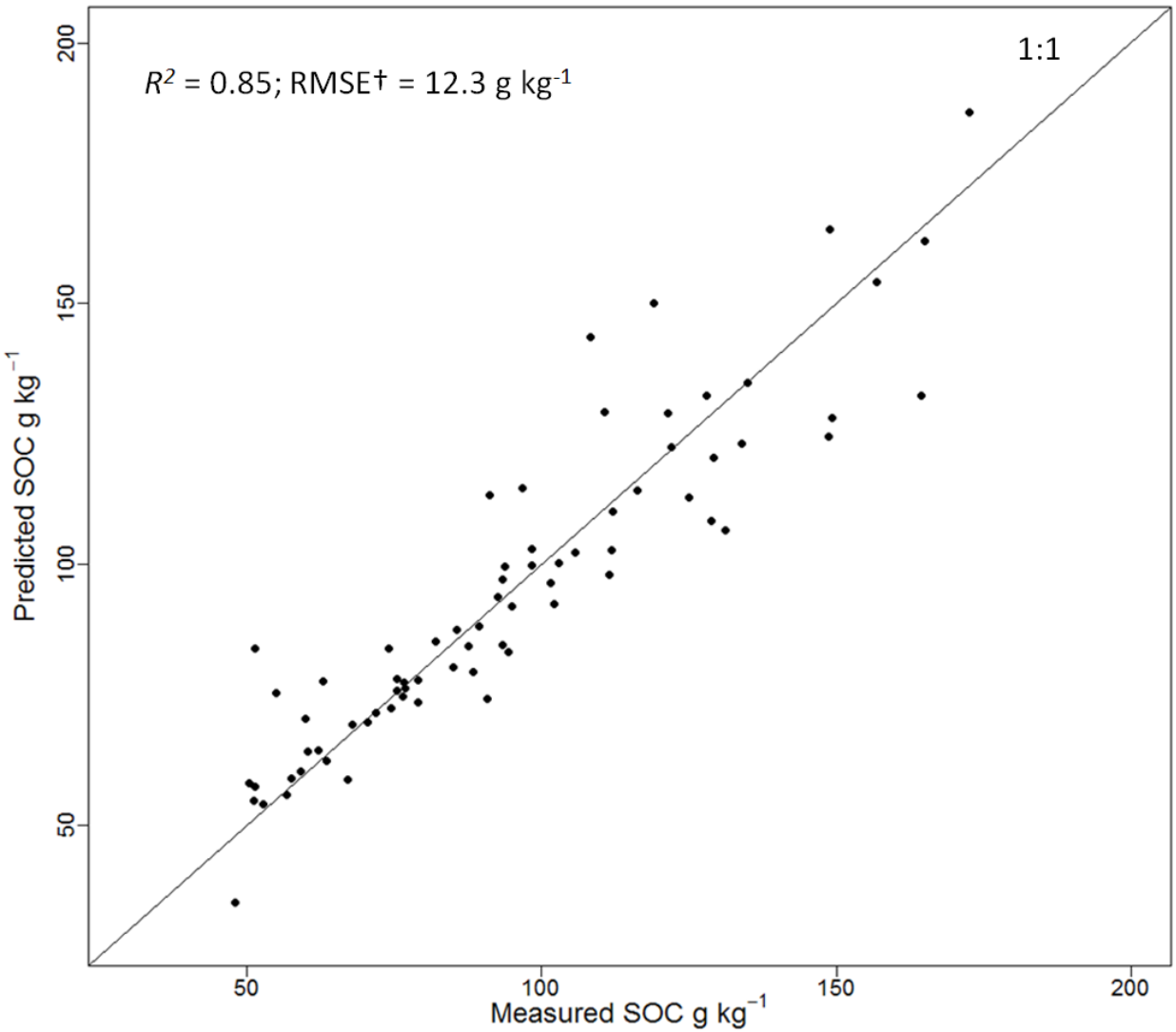

Figure 7.3 Scatter plot of soil organic carbon (SOC). The measured values were obtained by dry combustion and the predicted values by visible-near-infrared reflectance spectroscopy. † RMSE, root mean square error of cross validation.

7.5.4 Feature-space prediction of soil organic carbon by Random Forest

We tested the RF models with full and reduced datasets because the measurements of andic soil properties are generally costly and time-consuming. The 100 replicated model runs identified the m_{try} values of 3 and 11 as the optimal number of predictors for $RF_{reduced}$ and $RF_{full,}$ respectively. We obtained a large reduction in the $RMSE^{OOB}$ from 31.6 g kg^{-1} to 14.7 g kg^{-1} (Table 7.4) when we included the andic soil properties in the prediction. The range of predicted values using $RF_{reduced}$ showed significantly narrower ranges of SOC distribution and failed to predict high or low SOC values. The large reduction in the $RMSE^{OOB}$ was mostly attributed to Al_p, which was identified as the most important predictor assessed by the importance measures (Figure 7.4). In the RF_{full} model, most important predictors were related to the andic soil properties and ELEV and pH showed slight contributions (Figure 7.4). In the $RF_{reduced}$ model, ELEV was the most important predictor followed by SLOP and PLAN (Figure 7.4), however, the predictability remained poor unless the andic soil properties were included.

7.5.5 Spatial prediction of soil organic carbon by geostatistical methods

We have tested three geostatistical approaches to predict SOC across the entire study area:

1: Ordinary kriging with SOC concentrations estimated by the dry combustion method (n=72). The omni-directional variogram using a spherical model was chosen since anisotropic models showed no improvements in model predictability. The nugget value was high of 449 and the nugget to sill ratio was 44 %, indicating the presence of a moderate spatial dependency

Table 7.4 Statistical results of measured and predicted soil organic carbon (SOC) levels. The prediction methods were Random Forest without the andic soil properties ($RF_{reduced}$) and Random Forest with all available covariates (RF_{full}).

Prediction methods	Measured SOC levels ($g\ kg^{-1}$)						Predicted SOC levels ($g\ kg^{-1}$)			Validation statistics ($g\ kg^{-1}$)
	N	Mean	±	SD‡	Min.	Max.	Mean ± SD‡	Min.	Max.	$RMSE^{OOB}$§
$RF_{reduced}$†	72	94.8	±	31.5	48.1	172.3	95.0 ± 9.46	75.6	119	31.6
RF_{full}							95.2 ± 24.6	57.1	148	14.7

† $RF_{reduced}$ = Random Forest without andic soil properties; RF_{full} = Random Forest with all available covariates.
‡ Mean ± standard deviation; Min. = minimum and Max. = maximum.
§ $RMSE^{OOB}$ = root mean square error of prediction for the out-of-bag samples.

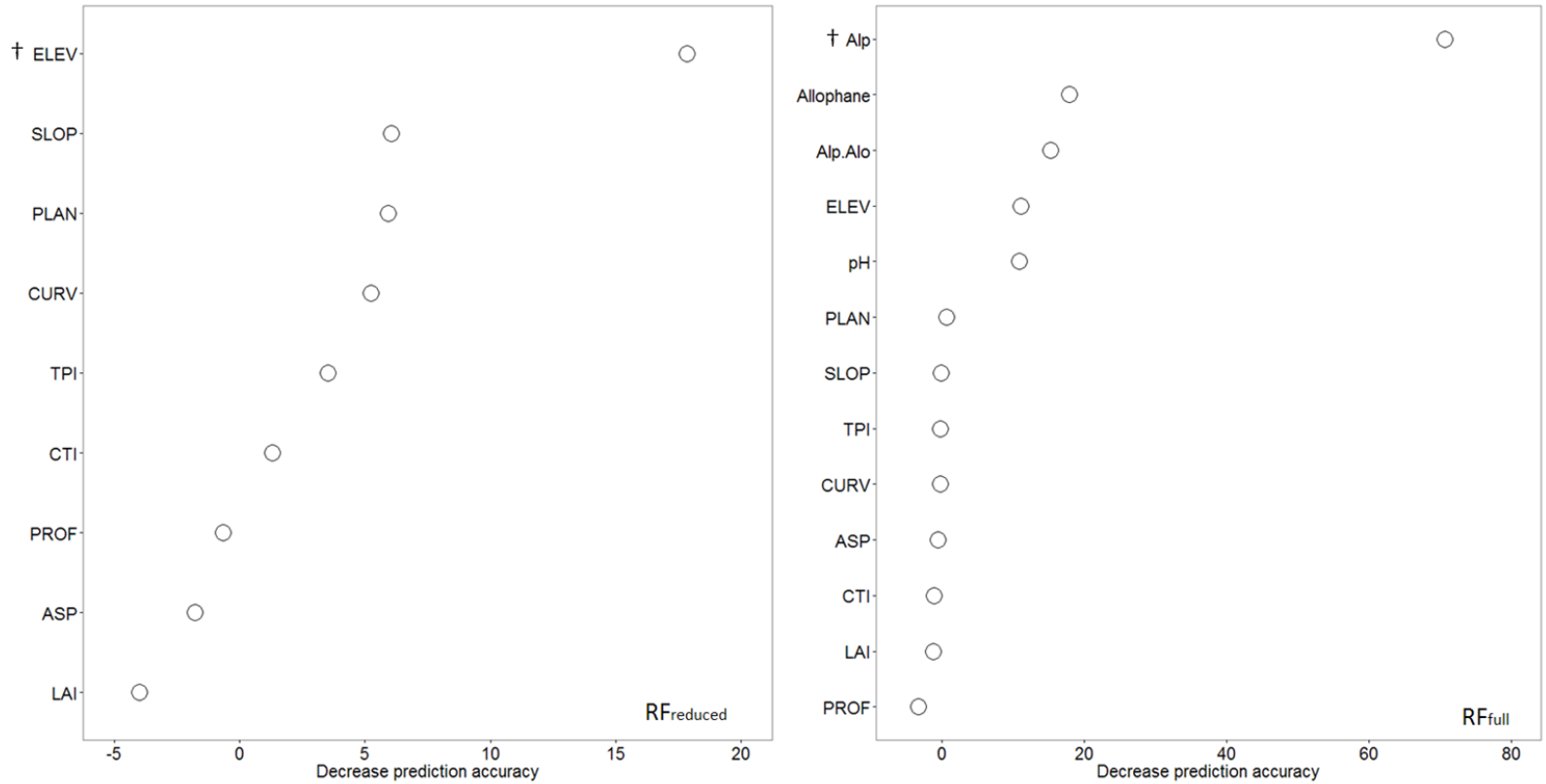

Figure 7.4 Variable importance of soil organic carbon predictions for a) Random Forest without the andic soil properties ($RF_{reduced}$) and b) Random Forest with all available covariates (RF_{full}). † ELEV, elevation; SLOP, slope gradient; PLAN, plan curvature; CURV, combined curvature; TPI, topographic position index; CTI, compound topographic index; PROF, profile curvature; ASP, slope aspect; LAI, leaf area index; Al_p, aluminum extracted by sodium pyrophosphate.

(Cambardella et al., 1994). High nugget values can originate from: 1) short-range variation in the SOC concentrations that cannot be explained by the spatial structure, 2) the sampling design, or 3) measurement errors associated with the dry combustion analysis of SOC_{ref} (RMSE = 2.42 g kg^{-1}; Goovaerts, 1997). In addition, overall SOC concentrations were high in our study site and thus inflated the absolute level of the nugget value. The area of high SOC concentrations was found towards the -N-W side of the watershed (Figure 7.5b) and the SOC concentrations varied from 63.1 g kg^{-1} to 141 g kg^{-1} across the study site (Figure 7.5b; Table 7.5). The cross validation of the OK method at 72 reference sample points yielded an RMSE of 28.0 g kg^{-1}, representing 29 % of the mean predicted SOC concentration. Spatially, validation points with high residual values (> ± 20 g kg^{-1}; Figure 7.5c) were observed across the watershed even in areas where the measured SOC concentrations were below the mean (Figure 7.5a). The residual values were not significantly correlated with any other measured covariates (data not shown).

2: Co-kriging with Al_p values, the most correlated covariate to SOC_{ref} in both feature space (*r = 0.85*; Table 7.2) and in spatial space at the 72 SOC_{ref} points. The spatial correlation was verified by the close proximity of the cross-variogram and the hulls of perfect correlation (Wackernagel, 2003) with the average distance of the two lines of 24.1 up to the range of 400 m (Figure 7.6b). This method improved the RMSE value to 16.0 g kg^{-1}, which was 17 % of the predicted mean SOC concentration. The range and the mean of the predicted SOC values were similar to the ordinary kriging approach (Table 7.5) and high residual values (> ± 20 g kg^{-1}) were primarily found where the SOC concentrations were high (> 125 g kg^{-1}; Figure 7.5a).

3: Co-kriging with SOC_{VNIRS} attempted to account for short-range SOC variation by reducing the sampling distances. The SOC_{VNIRS} values were correlated very well with SOC_{ref} in both feature space (R^2 = 0.85) and in spatial space shown by the almost identical line of the hulls of perfect

Table 7.5 Statistical results of geostatistical methods for interpolating soil organic carbon (SOC). The methods were ordinary kriging (OK), co-kriging with Al_p (CK(Al_p)), and co-kriging with VNIRS predicted SOC (CK(SOC_{VNIRS})).

Prediction methods	Measured SOC levels (g kg^{-1})				Predicted SOC levels (g kg^{-1})			Validation statistics (g kg^{-1})	
	N	Mean ± SD‡	Min.	Max.	Mean ± SD‡	Min.	Max.	RMSE§	RI (%)
OK†	72	94.8 ± 31.5	48.1	172.3	95.2 ± 17.3	63.1	141	28.0	
CK (Al_p)					95.2 ± 17.1	62.4	139	16.0	43
CK (SOC_{VNIRS})					87.2 ± 21.3	45.1	148	24.2	14

† OK = ordinary kriging; CK(Alp) = cokriging with sodium pyrophosphate extractable aluminum; CK (SOC_{VNIRS}) = cokriging with visible and near infrared reflectance spectroscopy predicted SOC.

‡ Mean ± standard deviation; Min. = minimum and Max. = maximum.

§ RMSE = root mean square error of cross validation; RI = relative improvement in root mean square error over OK [($RMSE_{OK}$ – $RMSE_{CK}$)*100/$RMSE_{OK}$].

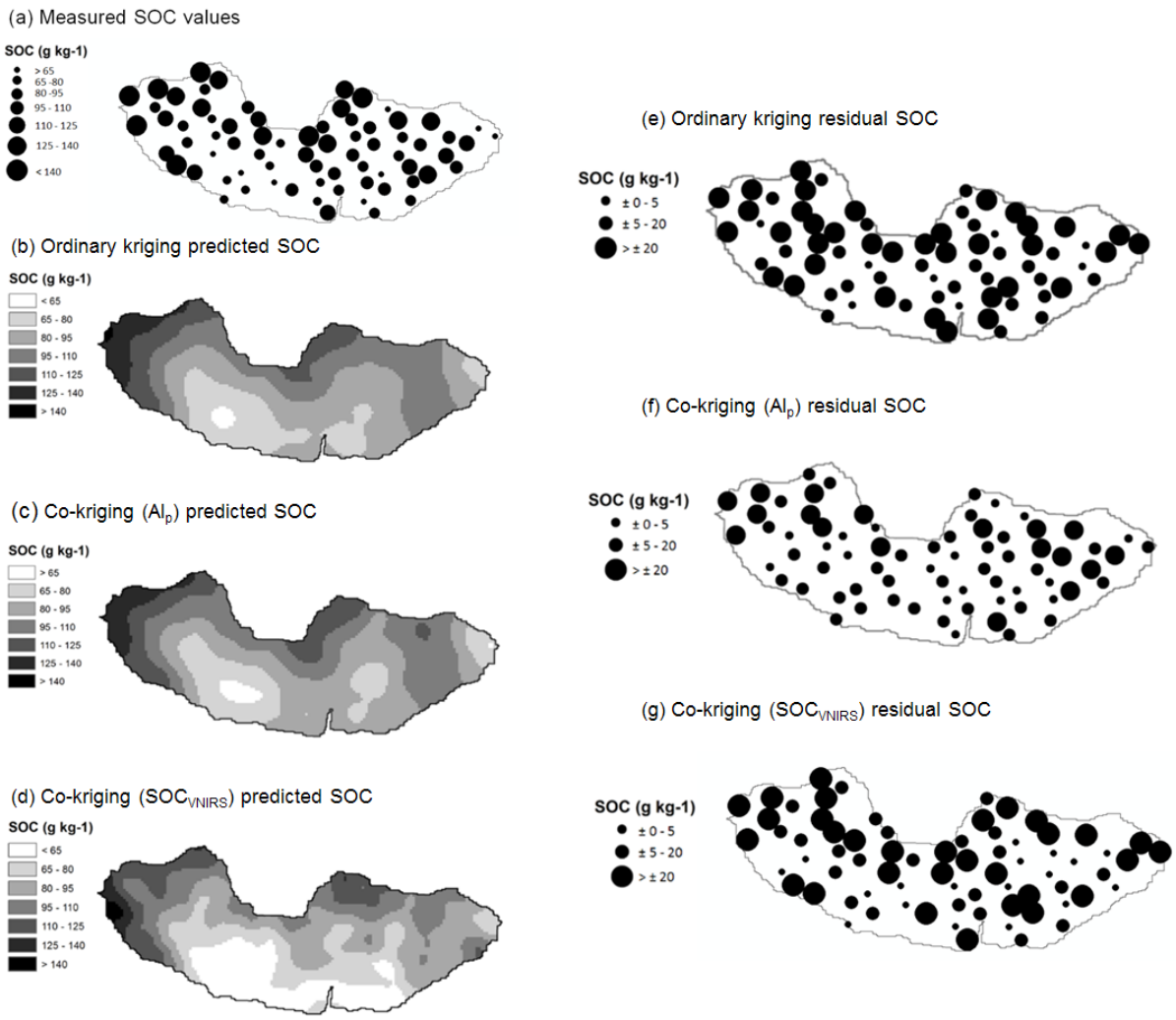

Figure 7.5 Maps of a) measured soil organic carbon (SOC) by dry combustion; b) c) and d) predicted soil organic carbon; and e), f) and g) prediction residuals. Figures b) and e) are associated with the ordinary kriging method, c) and f) with co-kriging with aluminum extracted by sodium pyrophosphate (Al_p), and d) and g) with co-kriging with visible-near-infrared reflectance spectroscopy (VNIRS) predicted SOC (SOC_{VNIRS}).

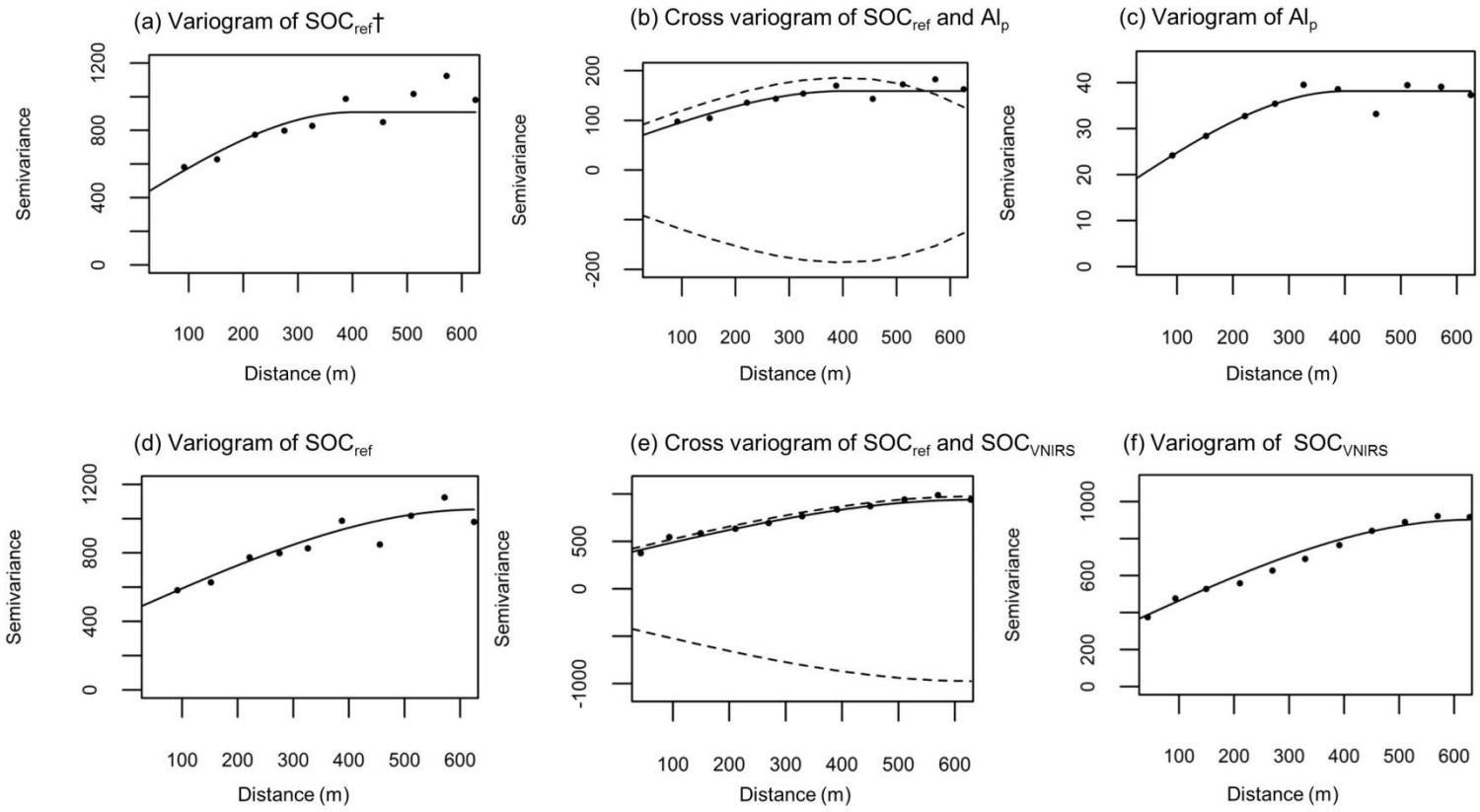

Figure 7.6 Direct- and cross- variograms used for co-kriging with aluminum extracted by sodium pyrophosphate (Al_p; a, b, and c), or visible-near-infrared reflectance spectroscopy (VNIRS) predicted SOC (SOC_{VNIRS}; d, e, and f). The dashed lines in the cross-variograms are the hulls of perfect correlation. †SOC_{ref}, soil organic carbon measured by dry combustion.

correlation and the cross-variogram with the average distance of the two lines of 24.2 up to the range of 631 m (Figure 7.6e). The higher sampling density decreased the nugget values shown in both the direct variogram for SOC_{VNIRS} and the cross-variogram between SOC_{ref} and SOC_{VNIRS} to 350 and 331, respectively (Figure 7.6e and f) compared to the direct-variogram of SOC_{ref} (449; Figure 7.6d). A higher number of sample points created more detailed boundaries between different SOC classes compared to ordinary kriging or co-kriging with Al_p (Figure 7.5d). This method reduced the RMSE value slightly to 24.2 g kg^{-1}, which was 28 % of the predicted mean SOC concentration. High residual values were found across the watershed (Figure 7.5g) similar to the ordinary kriging results (Figure 7.5e).

The improvement in the spatial estimation of SOC concentrations was high with the use of Al_p values with a relative improvement of 43 % compared to ordinary kriging (Table 7.5). Nevertheless, the cost of andic soil property analysis remains high even though our approach to use acid digestion for the clarification of the extracts will improve the efficiency of analysis.

The CK(SOC_{VNIRS}) did not reduce the RMSE significantly with a relative improvement of 14 % compared to ordinary kriging (Table 7.5) although it reduced the nugget value slightly by reducing sample distances. In this study, the calibration samples were selected systematically based on their spatial distributions, but this can be improved by selecting the reference sample set based on their spatial distributions as well as their spectral distributions to reduce the leverages (Bellon-Maurel et al., 2010). Furthermore, developing regional VNIRS-PLSR prediction models for SOC estimation can possibly replace all SOC predictions. Building regional models requires a large number of soil samples (Brown et al., 2006; Shepherd and Walsh, 2002). However, there are some recent approaches such as the inclusion of a spiking subset, i.e., a small number of local soil samples combined with national or regional spectral

library (Guerrero et al., 2010; Stenberg et al., 2010; Viscarra Rossel et al., 2009; Wetterlind and Stenberg, 2010). This has shown to reduce the prediction error with less sampling effort compared to a locally developed model especially when the spiking subset was selected according to its spectral distributions (Guerrero et al., 2014).

All three geostatistical approaches showed similar trends in the spatial distribution of SOC with higher SOC concentrations toward the western and northern part of the watershed (Figure 7.5b, c, and d). The western side was previously cultivated for cardamom and also the most recently converted area for coffee plantation (Figure 7.1a). A past study indicated increases in SOC concentrations under cardamom agroforestry with red alder depending on the amount of organic matter input from the shade trees (Sharma et al., 2009). The higher biomass production under this cultivation could have contributed in the accumulation of SOC in the western part of the watershed. Much of the rest of the study area was under housing and gardens until 1975 but the exact distribution of the houses remains unknown, and there is no available information whether the gardens were used for crop production or accumulating organic wastes. We are therefore not able to conclude whether the duration of coffee production has contributed to the spatial variation of SOC due to the absence of historical SOC records.

7.6 CONCLUSIONS

This study on spatial variability of topsoil soil organic carbon (SOC) in a coffee agroforestry micro-watershed on Andisols revealed i) high SOC values and large variation even at this small farm scale ii) the prevalence of andic soil properties in influencing spatial variation patterns, and iii) weak correlation with elevation but none with any other topographic or

vegetation covariates. Such covariates did not improve the prediction of SOC either by Random Forest (RF) or co-kriging models. We identified the presence of very consistent autocorrelation structures in both the measured SOC (SOC_{ref}; n=72) and the proxy measurement (SOC_{VNIRS}; n=520) but with high nugget effects. SOC_{VNIRS} showed the lowest root mean square error (RMSE = 12.3 g kg^{-1}) compared to RF and geostatistical methods and confirms its potential as a low-cost and rapid method for SOC estimation.

Large farm-scale variability of topsoil SOC in Andisols advocates for enhanced sampling strategies using spectral proxy measurement when assessing the impact of alternative management practices, or when designing chronosequences for SOC monitoring.

7.7 ACKNOWLEDGEMENTS

We are very grateful to Álvaro Barquero and his family, and Alexis Perez who helped in field campaigns. Dean Hively, Robert Schindelbeck, Sonam Sherpa, Michael Rutzke, Patricia Leandro, and Carlos Fernandez supported the soil analysis. David Rossiter provided us with invaluable suggestions in data analysis especially regarding the application of co-kriging. This work was supported by the European project CAFNET (EuropAid/121998/C/G), USDA-NIFA Special Grant on Computational Agriculture, the Cafetalera Aquiares farm (http://www.cafeaquiares.com), the ANR-Ecosfix project, the CIRAD-IRD SAFSE project, and the SOERE F-ORE-T network of observatories. We also acknowledge support from the Heiwa Nakajima Foundation, the Saltonstall family, the World Bank, and the Government of Japan. We thank reviewers for their useful suggestions.

CHAPTER 8: CONCLUSIONS

8.1 OVERALL SUMMARY

Spatio-temporally variable crop yields exist, which makes it difficult for growers to adjust their inputs while reductions in yields are economic and food security concerns. In this dissertation, we focused on transition and subsoil layers that are below the usually sampled top 15 cm of soil profiles, which have been identified as important sources of soil moisture and nutrients, especially under moisture-limited conditions. The deeper parts of the soil profile are not usually considered in crop management decisions nor assessed for the impacts of long-term soil and crop management. The current soil testing approach has also been heavily focused on plant available soil nutrient information, which is crucial but not exclusively when considering various yield constraining factors. The existing soil information such as the NRCS Soil Survey database can provide useful soil profile information, but may not have appropriate spatio-temporal resolution when considering soil properties that can be altered in a short time period by management. Therefore, a part of this dissertation was devoted to incorporating proximal sensing techniques for digital soil mapping to increase the efficiency of soil property assessment and also allowing us to determine the relationship between crop yields and three dimensional soil properties across space.

This dissertation contained six chapters and each contained one or more of the following specific goals: i) developing laboratory and in-situ soil property estimation and mapping approaches for soil biological, chemical, and physical properties, ii) developing approaches to

utilize yield monitor information to effectively assess within-field spatio-temporal yield variations, and iii) addressing the effects of crop and soil management on soil profile properties.

8.1.1 Rapid soil assessment

Proximal sensing techniques were incorporated as part of laboratory and in field assessment of soil biological, chemical, and physical properties to improve the spatial coverage of soil information and efficiency of measurement.

The in-situ measurement of soil profile properties using apparent electrical conductivity (ECa) was tested over nine fields in the Mid-Atlantic US, and showed promising results for mapping subsoil water retention parameters within the Coastal Plain physiographic province (Chapter 4). The cost of subsoil mapping is high and the ability of ECa to determine soil profile moisture retention parameters is highly beneficial when considering soil productivity under moisture-limited conditions.

Visible and near-infrared reflectance spectroscopy (VNIRS) was also tested in laboratory and field. In both cases, we confirmed significant effects of clay mineralogical differences in the predictability of soil properties including soil organic matter (SOM) and soil texture. In laboratory, we used k-means clustering on the spectra that grouped soil samples with low Fe-oxide spectral peaks and improved the predictability for sand, SOM, and permanganate oxidizable carbon (POXC), but the predictability was low for samples with higher Fe-oxide spectral peaks (Chapter 5). Interestingly, the correlations between clay or water retention at -1500 kPa and SOM or POXC were least in the sample set without Fe-oxide peaks. We also assessed the correlation matrix of the regression coefficients of partial least squares models, a new tool we proposed to assess whether the prediction of soil properties relies on similar spectral

absorption bands. Here, the spectral bands used to predict POXC was distinct from the ones for SOM and utilized 2123 nm, which is known to be related to organic matter, cellulose, glucan, pectin, and polysaccharides (Ben-Dor et al., 1997; Clark et al., 1990). Water stable aggregation (WSA) was much better predicted for sample sets with Fe-oxide peaks and the prediction relied on goethite and SOM related wavelengths in the visible range between 575 and 645 nm (Galvao and Vitorello, 1998; Stenberg et al., 2010), as well as protein and biotite/gibbsite absorption bands in the SWIR range of 2033, 2193, 2253, and 2383 nm (Ben-Dor et al., 1997; Clark, 1999; Clark et al., 1990; Post and Noble, 1993).

The relationship between clay mineralogy and soil organic carbon (SOC), and the predictability of SOC using VNIRS was clearer in our study undertaken within a coffee agroforestry farm in Costa Rica (Chapter 7). We built a SOC prediction model using VNIRS (R^2 = 0.85; root mean square error of prediction = 12.3 g kg^{-1}), and confirmed a strong positive correlation between SOC and Al extracted by sodium pyrophosphate, a proxy for Al-humus complex in volcanic ash soils.

In-situ measurements from 26 fields in the Mid-Atlantic US showed that a simplified version of VNIRS (660 and 940 nm) to be most correlated to Mehlich-3 extractable Al and Ca, not SOM unlike several past studies and products (Kweon et al., 2013; Kweon and Maxton, 2013). This suggests the incorporation of existing NRCS Soil Survey information to categorize the studied areas according to potential clay mineralogical differences when building a prediction model for SOM related properties.

8.1.2 Spatio-temporal assessment of yield data

The magnitude and characteristics of spatio-temporal variation of crop yields in the Mid-Atlantic US are expected to be different between physiographic provinces and also individual fields within each province. Grain harvester yield monitor information has two dimensional spatial structure as well as a time component when multiple years of yield data are collected. The spatio-temporal yield variations are affected by dynamic interactions of in-season weather conditions, soil properties, topography, pests and diseases, among others. In order to analyze these complex interactions, there was need to derive meaningful information from the multi-year yield data. We employed two statistical approaches, baseline functions and standardized principal component analysis (stdPCA). The former method fits a curvilinear model to the relationship between critical period precipitation amount and field-average yield, and shows whole field-scale expected yield loss under moisture-limited conditions. Baseline functions can be also used to forecast expected yield during the season, which we undertook using May and June precipitation for maize yield (R^2 = 0.38; Chapter 2) and can be used for in-season management (e.g. sidedress N application). Baseline function may be used at within-field scale if we accumulate larger datasets of yield for each field on a single crop type. This will allow us to assess the susceptibility in yield reductions in each within-field area under moisture-limited conditions. stdPCA had been used extensively in satellite remote sensing data at large geographical scales (i.e., continental scale) to explore multi-time step spatial data. In our study, we showed the ability of stdPCA to visualize characteristic yield patterns and assess the stability of the pattern within a range of critical period precipitation encountered. In addition, stdPCA can show that the characteristic yield patterns are generally only present under moisture-limited

conditions in the Coastal Plain province, while the patterns were more consistent across a range of critical period precipitation conditions in the Piedmont province.

8.1.3 Linking soil and crop yield information

Information of yield goals as well as the knowledge of yield constraints are necessary in order for growers to make decisions on site-specific management. Therefore, the relationships among soil-profile information, topography and derived characteristic yield patters were assessed for i) identifying potential yield constraints and ii) estimating within-field yield goals for fields with limited yield data availability (Chapter 4). The Random Forest machine learning approach was applied to predict the derived characteristic yield patterns (stdPCA score maps) using topographical variables and proximal sensor information. In the Delaware and Maryland Coastal Plain province, ECa, pH, and topographic wetness index were the most important parameters, while aspect, slope, and ECa were dominant factors in the Piedmont province. ECa was verified to be most correlated to subsoil moisture holding capacity, and thus we proposed the use of the Random Forest model to assess soil productivity in the Coastal Plain province. Since the spatial structure of grain yields is a result of complex interactions among many environmental and anthropogenic factors, accurate yield prediction without the incorporation of a wide array of parameters is challenging. However, approaches such as simulation models require a number of input variables, and moreover we did not find good agreement in yield prediction when we tried the Precision Nitrogen Management model at within-field scale (Melkonian et al., 2005). Instead, we have combined measured yield data and Random Forest prediction using statistical approaches to compare two spatial patterns using a moving window of 0.23 ha (0.5 acres), which is the minimal area with which growers can effectively undertake site-specific management. This

produced a map showing the locations where the Random Forest model failed to account for the yield variations, which helps to determine the locations where additional soil sampling may be advised to further determine the yield constraints.

8.1.4 Management induced change in soil profile properties

We also assessed the effects of long-term maize cropping systems (residue returned vs. residue removed) under no-till or plow-till management for soil-profile scale biological, chemical, and physical properties (Chapter 6). This work assessed what aspect of soil characteristics are more manageable by these practices and whether soil moisture holding capacity can be altered. We highlighted from this work i) little change in total quantity of available water, and ii) nutrient mining in the transition and subsoils when residue is removed under no-till. In order to improve the availability of soil moisture, the accessibility of larger soil volume by crop roots as well as the reduction in evaporation by surface cover appeared more manageable compared to the total quantity of available soil water by tillage and residue management. In addition, careful assessment of the effects of crop residue removal needs to be undertaken including soil assessment below the usual sampling depth of 15 cm, because a "hidden" effect on soil health may be present deeper in the soil profile, which can affect long-term sustainability of crop production.

8.2 CONCLUDING REMARKS

This dissertation helped advance our knowledge on two fronts; i) certain underlying conditions at which VNIRS prediction for SOM related properties are successful, and ii) the importance of including soil profile assessment when considering soil productivity or long-term sustainability of crop and soil management. The relationship between clay particles and SOM related properties needs to be assessed in more detail to explore the presence of the type of clay minerals. Then, we may be able to better stratify calibration sample sets to homogenize the relationship among certain clay minerals, SOM, and indicative spectral features. This will also help us improve the prediction of SOM in field using an on-the-go VNIRS sensors.

For soil profile assessment, we identified deep nutrient mining for those nutrients that were not supplied from fertilizers. This can be an important frontier for research because the uplifting of soil nutrients by plant root uptake and re-deposition by aboveground biomass may be masking the effects of crop production at the surface 15 cm of soil layer and causing “hidden” nutrient mining across a soil profile.

APPENDIX A. SUMMARY STATISTICS OF THE SHALLOW SAMPLE SET, CHAPTER 4

Area	Field ID	N	BD			Clay			Silt			Sand			-10 kPa			-33 kPa			-100 kPa			-1500 kPa			AWC			SOM		
			g cm^{-3}	SD	CV	%	SD	CV	%	SD	CV	%	SD	CV	m m^{-1}	SD	CV	m m^{-1}	SD	CV	m m^{-1}	SD	CV	m m^{-1}	SD	CV	m m^{-1}	SD	CV	%	SD	CV
1	1	15	1.22	0.11	9.10	8.90	2.28	25.62	50.80	9.04	17.80	40.30	10.40	25.81	0.26	0.03	12.07	0.18	0.03	19.27	0.14	0.03	20.35	0.06	0.02	26.49	0.20	0.03	13.66	1.43	0.44	30.98
	2	15	1.30	0.10	7.32	8.43	4.27	50.65	41.90	8.41	20.07	49.70	9.50	19.11	0.23	0.03	12.94	0.18	0.03	18.64	0.13	0.02	14.96	0.05	0.02	31.87	0.18	0.03	16.97	1.15	0.29	24.78
	3	6	1.35	0.04	2.77	6.31	1.03	16.32	43.50	7.49	17.22	50.20	8.32	16.57	0.24	0.03	11.23	0.13	0.02	14.26	0.11	0.01	12.43	0.04	0.00	13.06	0.20	0.02	11.41	0.55	0.22	39.45
	4	6	1.15	0.08	6.59	10.30	2.06	20.00	70.00	8.39	11.99	19.70	10.30	52.28	0.33	0.04	11.01	0.25	0.02	10.16	0.16	0.02	11.20	0.07	0.01	11.30	0.26	0.03	11.17	1.17	0.32	27.35
	5	6	1.19	0.12	9.92	9.69	1.77	18.27	71.10	9.51	13.38	19.20	9.15	47.66	0.33	0.04	10.96	0.23	0.03	11.47	0.17	0.02	12.97	0.07	0.01	11.36	0.26	0.03	12.70	1.52	0.29	19.14
	6	9	1.24	0.13	10.56	9.27	3.20	34.52	51.10	13.70	26.81	39.60	16.10	40.66	0.25	0.07	26.30	0.19	0.05	27.47	0.15	0.05	31.10	0.06	0.02	29.59	0.19	0.05	25.97	1.21	0.70	57.77
	7	9	1.17	0.10	8.55	8.96	3.25	36.27	50.00	14.40	28.80	41.10	17.00	41.36	0.26	0.06	22.32	0.22	0.05	24.86	0.15	0.05	32.21	0.06	0.02	37.17	0.20	0.04	19.39	1.27	0.32	25.20
	8	6	1.28	0.08	5.86	8.11	2.00	24.66	53.80	8.33	15.48	38.10	10.20	26.77	0.27	0.04	13.44	0.18	0.05	24.51	0.14	0.03	21.22	0.05	0.01	27.59	0.22	0.02	10.18	1.32	0.50	37.88
	9	3	1.13	0.08	6.70	6.62	1.42	21.45	44.30	16.80	37.92	49.10	17.90	36.46	0.23	0.06	28.01	0.18	0.05	27.70	0.13	0.04	27.60	0.06	0.02	30.46	0.17	0.05	28.34	1.47	0.25	17.14
	10	3	1.26	0.06	4.83	7.75	1.78	22.97	62.90	10.10	16.06	29.30	11.90	40.61	0.29	0.03	10.49	0.21	0.02	10.94	0.15	0.02	12.04	0.06	0.01	13.97	0.22	0.02	9.46	1.00	0.36	36.10
	11	7	1.22	0.17	13.69	6.43	3.10	48.21	48.20	15.90	32.99	45.30	18.90	41.72	0.25	0.07	26.43	0.19	0.06	30.93	0.13	0.03	27.32	0.05	0.02	29.43	0.20	0.05	25.81	1.10	0.34	31.09
	12	6	1.52	0.10	6.45	8.27	4.06	49.09	32.70	11.50	35.17	59.00	15.30	25.93	0.20	0.05	24.34	0.15	0.05	33.63	0.12	0.05	38.68	0.04	0.02	36.69	0.16	0.04	22.69	1.30	0.59	45.62
	13	6	1.46	0.05	3.47	2.88	1.09	37.85	17.40	3.72	21.38	79.80	4.61	5.78	0.14	0.02	12.22	0.08	0.01	16.57	0.07	0.01	18.89	0.02	0.00	19.82	0.11	0.01	11.25	0.85	0.27	32.24
2	14	12	1.14	0.11	9.65	17.60	4.05	23.01	56.40	3.63	6.44	26.00	5.13	19.73	0.31	0.02	6.63	0.27	0.02	8.98	0.19	0.01	6.99	0.10	0.01	6.68	0.21	0.02	11.91	1.50	0.57	37.73
	15	6	1.39	0.09	6.21	18.90	3.91	20.69	60.20	4.19	6.96	20.90	1.64	7.85	0.31	0.02	8.06	0.26	0.01	5.69	0.21	0.02	7.93	0.11	0.01	8.80	0.20	0.03	15.17	1.85	0.24	12.70
	16	12	1.18	0.14	11.95	14.30	4.65	32.52	47.60	9.68	20.34	38.00	13.50	35.53	0.32	0.05	16.17	0.24	0.05	20.41	0.20	0.05	22.76	0.11	0.03	28.99	0.21	0.02	10.66	1.74	0.45	25.98
	17	6	1.06	0.12	11.23	12.30	3.26	26.50	55.10	11.00	19.96	32.70	12.20	37.31	0.36	0.05	15.15	0.26	0.04	13.97	0.21	0.03	15.41	0.12	0.02	18.09	0.25	0.04	14.43	2.33	0.44	18.76
	18	6	1.15	0.15	12.70	12.90	3.86	29.92	54.60	13.00	23.81	32.60	16.50	50.61	0.34	0.06	18.05	0.24	0.05	21.19	0.19	0.04	21.96	0.10	0.02	23.63	0.24	0.04	16.02	1.87	0.32	17.11
	19	12	1.21	0.12	9.59	14.00	3.37	24.07	37.50	8.26	22.03	48.60	10.60	21.81	0.28	0.04	13.52	0.21	0.04	18.05	0.17	0.02	14.42	0.09	0.01	15.85	0.20	0.03	16.10	1.31	0.42	31.98
	20	6	0.99	0.05	5.40	10.60	1.58	14.91	35.30	6.89	19.52	54.10	8.32	15.38	0.29	0.02	6.77	0.16	0.02	12.42	0.14	0.02	14.44	0.07	0.01	13.82	0.22	0.01	6.35	0.62	0.13	21.56
	21	12	0.97	0.09	8.95	21.10	3.72	17.63	51.30	4.51	8.79	27.70	6.25	22.56	0.43	0.04	10.44	0.34	0.04	12.47	0.29	0.03	10.34	0.16	0.02	13.90	0.27	0.03	11.01	1.93	0.51	26.42
	22	12	1.17	0.08	6.84	17.80	1.95	10.96	40.20	4.19	10.42	42.00	5.08	12.10	0.33	0.03	9.23	0.25	0.04	14.05	0.19	0.02	8.29	0.11	0.01	11.87	0.22	0.03	14.22	1.13	0.49	43.72
	23	6	1.30	0.15	11.69	16.50	1.75	10.61	53.10	1.08	2.03	30.50	1.34	4.39	0.33	0.02	6.34	0.24	0.01	5.33	0.22	0.01	6.52	0.11	0.01	6.95	0.22	0.02	6.74	1.22	0.43	34.92
	24	6	1.12	0.05	4.78	16.50	6.63	40.18	62.60	6.84	10.93	20.90	11.20	53.59	0.41	0.03	8.41	0.28	0.03	12.23	0.23	0.03	11.29	0.12	0.03	24.55	0.29	0.02	5.63	2.33	0.76	32.66
	25	6	1.18	0.11	9.24	16.50	5.12	31.03	52.70	4.91	9.32	30.90	5.71	18.48	0.33	0.04	13.07	0.23	0.03	14.40	0.19	0.03	14.22	0.11	0.02	21.38	0.22	0.03	14.33	1.93	0.40	20.62
	26	6	1.18	0.16	13.56	17.40	1.89	10.86	51.60	4.16	8.06	31.00	5.10	16.45	0.38	0.02	5.47	0.30	0.03	9.90	0.26	0.02	7.26	0.12	0.01	6.17	0.25	0.02	5.91	1.87	0.45	23.85

Area	Field ID	N	pH			P			K			Mg			Ca			CEC			Al			Moisture		
				SD	CV	ppm	SD	CV	ppm	SD	CV	ppm	SD	CV	ppm	SD	CV	cmol kg^{-}	SD	CV	ppm	SD	CV	%	SD	CV
1	1	15	6.17	0.41	6.58	189.0	49.40	26.14	135	37.60	27.85	97	18.50	19.01	800	147	18.38	4.99	1.36	27.25	951	184	19.35	19.30	3.06	15.85
	2	15	6.33	0.36	5.66	37.3	24.10	64.61	122	44.50	36.48	156	45.70	29.29	607	144	23.72	3.99	1.12	28.07	729	154	21.12	18.30	2.13	11.64
	3	6	6.18	0.26	4.14	22.3	7.61	34.13	87	19.10	22.03	96	15.40	16.13	414	48	11.47	3.55	2.22	62.54	673	53	7.86	17.30	1.42	8.21
	4	6	6.52	0.10	1.51	65.3	41.10	62.94	165	32.40	19.64	121	15.70	12.98	1028	43	4.21	5.98	0.62	10.38	814	61	7.51	22.40	1.63	7.28
	5	6	6.09	0.36	5.94	51.3	28.70	55.95	136	23.90	17.57	81	24.30	30.15	718	181	25.21	4.13	0.96	23.29	914	161	17.61	24.10	2.66	11.04
	6	9	5.69	0.25	4.34	32.6	25.90	79.45	132	34.20	25.91	125	31.00	24.80	503	140	27.83	4.82	2.13	44.19	1036	209	20.17	19.40	4.29	22.11
	7	9	6.30	0.40	6.40	40.3	30.30	75.19	125	33.30	26.64	144	34.30	23.82	536	128	23.88	4.91	2.06	41.96	938	205	21.86	21.00	4.33	20.62
	8	6	6.33	0.16	2.58	34.2	14.70	42.98	139	27.90	20.07	134	9.65	7.20	612	32	5.16	3.88	0.39	9.97	911	181	19.87	20.70	1.70	8.21
	9	3	6.47	0.21	3.21	159.0	18.20	11.45	165	31.50	19.09	153	18.00	11.76	1185	313	26.41	6.73	1.80	26.75	974	174	17.86	19.50	2.64	13.54
	10	3	6.53	0.25	3.86	135.0	27.00	20.00	187	19.90	10.64	157	25.50	16.24	947	100	10.56	5.97	1.33	22.28	833	40	4.81	20.80	1.90	9.13
	11	7	6.33	0.45	7.11	106.0	37.00	34.91	148	45.60	30.81	104	33.60	32.31	840	244	29.05	5.16	1.25	24.22	764	232	30.37	17.50	2.62	14.97
	12	6	6.02	0.51	8.50	120.0	54.50	45.42	144	67.60	46.94	127	33.40	26.30	634	229	36.12	3.82	1.48	38.74	815	137	16.81	2.80	0.72	25.86
	13	6	6.37	0.33	5.13	90.8	24.90	27.42	112	27.80	24.82	104	21.30	20.48	492	129	26.22	2.93	0.78	26.76	620	78	12.61	2.45	0.52	21.27
2	14	12	6.05	0.66	10.98	76.6	37.30	48.69	134	57.50	42.91	126	34.50	27.38	962	320	33.26	7.36	1.94	26.36	929	211	22.71	12.30	10.20	82.93
	15	6	6.58	0.40	6.11	33.0	13.50	40.91	278	40.40	14.53	213	71.80	33.71	1253	202	16.12	8.30	1.06	12.77	895	110	12.29	7.13	0.80	11.25
	16	12	6.13	0.40	6.54	51.3	23.80	46.39	155	37.30	24.06	246	59.30	24.11	1292	311	24.07	9.02	3.15	34.92	794	127	15.99	24.70	3.78	15.30
	17	6	6.47	0.55	8.45	14.7	5.57	37.89	75	8.52	11.41	295	107.00	36.27	1475	367	24.88	10.60	2.25	21.23	692	87	12.57	28.00	5.08	18.14
	18	6	6.53	0.36	5.45	74.7	18.80	25.17	99	30.80	31.11	161	49.70	30.87	1663	531	31.93	9.07	2.94	32.41	788	39	4.95	24.40	4.08	16.72
	19	12	6.73	0.32	4.71	70.0	29.60	42.29	90	30.00	33.30	210	31.80	15.14	1140	114	10.00	6.79	1.01	14.87	699	87	12.39	11.20	8.95	79.91
	20	6	6.45	0.62	9.55	38.3	23.00	60.05	125	27.50	22.00	195	56.70	29.08	704	115	16.34	5.27	0.51	9.58	605	118	19.50	1.22	0.51	41.64
	21	12	6.56	0.37	5.69	50.2	21.60	43.03	132	43.50	32.95	265	65.10	24.57	1066	263	24.67	7.92	1.78	22.47	776	146	18.81	17.00	15.30	90.00
	22	12	5.78	0.36	6.18	21.4	6.10	28.50	97	26.40	27.33	126	27.20	21.59	530	96	18.19	4.32	0.87	20.23	722	107	14.82	12.70	11.30	88.98
	23	6	6.45	0.36	5.52	69.7	32.70	46.92	145	42.40	29.24	206	26.40	12.82	1004	180	17.93	7.05	1.07	15.18	819	59	7.14	2.80	1.02	36.43
	24	6	5.93	0.34	5.72	22.0	7.27	33.05	95	16.80	17.78	206	89.90	43.64	1451	514	35.42	11.90	1.35	11.34	734	99	13.54	27.50	3.67	13.35
	25	6	6.32	0.12	1.85	57.2	20.60	36.01	124	33.70	27.18	234	86.90	37.14	1180	288	24.41	7.22	1.99	27.56	738	75	10.20	23.00	2.93	12.74
	26	6	6.58	0.08	1.14	226.0	42.30	18.72	321	23.10	7.20	222	17.80	8.02	1296	127	9.80	9.03	0.64	7.10	924	93	10.11	2.77	0.37	13.43

APPENDIX B. SUMMARY STATISTICS OF THE DEEP SAMPLE SET, CHAPTER 4

Area	Field ID	n	Depth	BD			Clay			Silt			Sand			-10 kPa			-33 kPa			-100 kPa			-1500 kPa			AWC			WSA			SOM		
				g cm^{-3}	SD	CV	%	SD	CV	%	SD	CV	%	SD	CV	m m^{-1}	SD	CV	m m^{-1}	SD	CV	m m^{-1}	SD	CV	m m^{-1}	SD	CV	m m^{-1}	SD	CV	%	SD	CV	%	SD	CV
1	1	9	0-15	1.2	0.1	9.5	9.6	2.5	26.3	50.8	8.3	16.3	39.5	10.3	25.9	0.324	0.0	8.2	0.227	0.0	13.8	0.178	0.0	16.5	0.065	0.0	26.5	0.210	0.0	13.3	41.4	13.8	33.3	1.6	0.4	27.7
			15-30	1.36	0.11	8.09	12.71	3.62	28.48	49.49	9.01	18.21	37.80	11.23	29.71	0.361	0.02	5.65	0.263	0.03	13.08	0.195	0.03	14.72	0.067	0.02	26.85	0.199	0.03	12.81	23.50	17.00	72.34	1.04	0.34	32.69
			30-45	1.45	0.11	7.86	19.11	6.83	35.74	43.29	10.04	23.19	37.59	15.28	40.65	0.392	0.05	11.51	0.291	0.04	13.09	0.224	0.03	12.59	0.087	0.02	28.13	0.184	0.04	21.30	18.30	15.70	85.79	1.02	0.52	50.98
			45-60	1.52	0.05	3.14	20.01	5.23	26.14	38.56	10.75	27.88	41.41	14.78	35.69	0.403	0.04	10.00	0.300	0.04	14.43	0.227	0.03	13.48	0.089	0.02	22.78	0.175	0.03	19.66	11.10	8.32	74.95	0.68	0.44	64.71
			60-90	1.61	0.06	3.78	20.72	2.66	12.84	28.11	10.15	36.11	51.16	11.79	23.05	0.400	0.04	8.98	0.285	0.03	9.37	0.219	0.02	9.41	0.092	0.01	11.94	0.155	0.03	20.39	8.95	6.94	77.54	0.66	0.28	42.42
	2	9	0-15	1.29	0.10	8.06	9.07	5.02	55.35	42.40	6.74	15.90	48.50	7.62	15.71	0.222	0.03	12.12	0.195	0.02	10.56	0.130	0.02	14.08	0.058	0.02	30.53	0.164	0.02	12.93	20.20	6.58	32.57	1.24	0.22	17.66
			15-30	1.41	0.10	7.30	13.60	7.95	58.46	39.40	7.11	18.05	47.00	9.13	19.43	0.212	0.03	13.63	0.186	0.03	15.16	0.129	0.03	23.49	0.065	0.03	46.55	0.146	0.01	9.79	na	na	na	na	na	na
			30-45	1.49	0.07	4.87	18.30	6.86	37.49	38.60	8.20	21.24	43.10	11.20	25.99	0.224	0.04	19.91	0.196	0.04	19.49	0.144	0.03	23.19	0.077	0.03	40.08	0.146	0.02	14.79	na	na	na	na	na	na
			45-60	1.52	0.07	4.46	17.60	5.19	29.49	33.40	10.70	32.04	49.00	15.40	31.43	0.220	0.05	23.36	0.189	0.05	26.35	0.141	0.04	25.67	0.079	0.02	28.10	0.141	0.03	22.84	na	na	na	na	na	na
			60-90	1.65	0.06	3.57	13.70	4.17	30.44	20.00	9.73	48.65	66.30	13.80	20.81	0.186	0.04	23.55	0.157	0.05	28.92	0.115	0.04	30.96	0.070	0.01	20.14	0.116	0.03	27.76	na	na	na	na	na	na
	5	9	0-15	1.20	0.12	9.92	9.69	1.77	18.27	71.10	9.51	13.38	19.17	9.15	47.73	0.333	0.04	10.96	0.225	0.03	11.47	0.165	0.02	12.97	0.071	0.01	11.35	0.263	0.03	12.70	21.70	7.63	35.16	1.52	0.29	19.14
			15-30	1.33	0.11	8.12	16.90	5.23	30.95	66.90	9.15	13.68	16.23	7.97	49.11	0.307	0.04	12.44	0.233	0.02	10.47	0.173	0.02	11.45	0.088	0.02	20.11	0.220	0.04	18.73	na	na	na	na	na	na
			30-45	1.43	0.10	7.13	21.70	5.61	25.85	62.60	10.00	15.97	15.69	8.80	56.09	0.331	0.04	12.11	0.256	0.02	7.58	0.191	0.02	11.41	0.109	0.02	20.37	0.222	0.05	20.45	na	na	na	na	na	na
			45-60	1.50	0.16	10.93	20.80	4.55	21.88	55.80	13.80	24.73	23.39	14.70	62.85	0.315	0.05	15.08	0.241	0.03	12.41	0.183	0.03	17.27	0.105	0.02	18.57	0.209	0.04	21.00	na	na	na	na	na	na
			60-90	1.56	0.23	14.62	20.40	4.44	21.76	49.00	19.40	39.59	30.60	22.70	74.18	0.292	0.07	23.63	0.225	0.06	27.56	0.169	0.04	26.39	0.097	0.02	20.97	0.195	0.06	29.95	na	na	na	na	na	na
	6	9	0-15	1.24	0.14	11.29	9.27	3.20	34.52	51.10	13.70	26.81	39.60	16.10	40.66	0.254	0.07	26.30	0.194	0.05	27.47	0.146	0.05	31.10	0.063	0.02	29.59	0.191	0.05	25.97	31.20	14.30	45.83	1.21	0.67	55.37
			15-30	1.34	0.13	9.70	12.80	5.71	44.61	49.70	13.90	27.97	37.50	17.60	46.93	0.259	0.07	26.64	0.204	0.05	26.37	0.142	0.05	36.62	0.070	0.02	29.31	0.189	0.05	27.51	na	na	na	na	na	na
			30-45	1.40	0.11	7.86	16.60	8.05	48.49	44.50	15.60	35.06	38.90	21.20	54.50	0.276	0.08	28.59	0.213	0.07	31.03	0.155	0.05	32.19	0.082	0.03	33.25	0.195	0.06	30.87	na	na	na	na	na	na
			45-60	1.48	0.12	8.11	14.50	6.46	44.55	35.30	16.20	45.89	50.10	21.20	42.32	0.239	0.08	31.42	0.182	0.06	35.27	0.133	0.05	35.41	0.072	0.03	38.04	0.167	0.05	31.20	na	na	na	na	na	na
			60-90	1.62	0.12	7.41	10.60	5.17	48.77	18.60	11.70	62.90	70.80	15.90	22.46	0.180	0.06	35.50	0.124	0.05	42.98	0.090	0.04	40.53	0.054	0.02	43.18	0.126	0.04	35.08	na	na	na	na	na	na
2	14	6	0-15	1.18	0.14	11.69	18.10	5.29	29.23	55.30	4.24	7.67	26.60	6.91	25.98	0.322	0.03	7.80	0.257	0.03	11.25	0.194	0.01	6.24	0.100	0.01	6.19	0.222	0.03	12.93	31.80	4.77	15.00	1.90	0.53	27.84
			15-30	1.42	0.14	9.58	21.80	9.86	45.23	54.10	5.00	9.24	24.10	9.93	41.20	0.308	0.02	6.49	0.270	0.01	4.78	0.195	0.02	10.31	0.104	0.02	20.29	0.204	0.04	18.77	na	na	na	na	na	na
			30-45	1.59	0.14	8.87	30.30	15.30	50.50	46.00	9.75	21.20	23.70	17.00	71.73	0.310	0.05	14.58	0.264	0.04	16.10	0.200	0.03	16.35	0.123	0.04	32.60	0.187	0.03	14.55	na	na	na	na	na	na
			45-60	1.40	0.38	26.86	34.00	13.30	39.12	40.50	10.40	25.68	25.50	16.60	65.10	0.323	0.05	16.97	0.271	0.04	13.95	0.203	0.04	19.06	0.140	0.04	25.36	0.183	0.03	18.74	na	na	na	na	na	na
	16	6	0-15	1.29	0.05	3.69	15.70	4.62	29.43	45.60	8.25	18.09	38.70	12.20	31.52	0.330	0.04	11.39	0.256	0.04	16.33	0.222	0.04	18.60	0.112	0.03	27.95	0.219	0.01	3.53	12.70	8.89	70.00	1.93	0.47	24.46
			15-30	1.48	0.15	10.14	19.30	5.79	30.00	43.80	9.67	22.08	36.90	13.80	37.40	0.312	0.07	22.63	0.247	0.06	26.07	0.210	0.05	22.90	0.113	0.03	30.09	0.199	0.04	20.05	na	na	na	na	na	na
			30-45	1.71	0.17	9.65	24.40	6.15	25.20	39.10	11.80	30.18	36.50	17.20	47.12	0.325	0.08	23.60	0.257	0.06	25.14	0.214	0.05	24.21	0.131	0.04	26.87	0.194	0.05	24.95	na	na	na	na	na	na
			45-60	1.64	0.27	16.65	21.60	10.20	47.22	39.30	11.70	29.77	39.10	18.50	47.31	0.337	0.07	21.78	0.263	0.08	30.19	0.217	0.06	27.65	0.125	0.05	37.92	0.212	0.05	23.77	na	na	na	na	na	na
	19	6	0-15	1.31	0.05	3.73	14.10	4.91	34.82	39.20	8.00	20.41	46.70	12.40	26.55	0.306	0.02	7.84	0.228	0.03	13.99	0.179	0.02	13.91	0.086	0.02	19.70	0.220	0.01	4.59	19.10	7.53	39.42	1.52	0.40	26.45
			15-30	1.58	0.12	7.72	13.60	4.51	33.16	35.80	8.15	22.77	50.50	11.10	21.98	0.281	0.02	7.83	0.212	0.03	15.38	0.159	0.02	14.91	0.080	0.02	21.07	0.201	0.01	5.77	na	na	na	na	na	na
			30-45	1.59	0.12	7.23	16.90	5.42	32.07	32.90	8.75	26.60	50.20	12.60	25.10	0.281	0.03	9.15	0.224	0.05	21.12	0.167	0.03	19.28	0.095	0.02	26.22	0.186	0.01	6.72	na	na	na	na	na	na
			45-60	1.67	0.16	9.58	17.60	5.12	29.09	34.00	8.93	26.26	48.50	12.70	26.19	0.292	0.03	9.97	0.234	0.04	18.89	0.170	0.03	17.18	0.098	0.02	24.44	0.194	0.02	7.84	na	na	na	na	na	na
	21	6	0-15	1.01	0.11	10.69	20.60	4.67	22.67	50.60	6.34	12.53	28.80	8.47	29.41	0.431	0.06	13.39	0.355	0.04	10.42	0.304	0.03	11.05	0.159	0.02	14.78	0.272	0.04	15.18	26.00	12.50	48.08	2.32	0.44	18.79
			15-30	1.20	0.11	9.17	23.20	5.64	24.31	47.70	6.53	13.69	29.10	7.65	26.29	0.406	0.05	12.27	0.337	0.04	12.11	0.283	0.03	11.27	0.163	0.02	9.33	0.243	0.05	19.55	na	na	na	na	na	na
			30-45	1.44	0.23	15.76	30.40	8.89	29.24	39.40	7.21	18.30	30.30	14.00	46.20	0.407	0.07	17.49	0.335	0.06	18.54	0.277	0.06	21.08	0.167	0.04	23.35	0.241	0.04	17.55	na	na	na	na	na	na
			45-60	1.46	0.19	13.01	26.30	10.90	41.44	34.80	10.40	29.89	38.90	18.30	47.04	0.396	0.09	22.10	0.329	0.07	21.61	0.273	0.07	24.51	0.153	0.04	26.73	0.243	0.06	24.24	na	na	na	na	na	na
	22	6	0-15	1.23	0.06	4.59	17.60	2.55	14.49	38.40	2.57	6.69	44.00	3.76	8.55	0.336	0.03	9.64	0.241	0.04	14.73	0.188	0.02	11.49	0.099	0.01	11.47	0.237	0.02	10.08	19.90	7.93	39.85	1.53	0.15	9.87
			15-30	1.34	0.04	2.85	20.50	2.09	10.20	35.10	2.69	7.66	44.40	3.80	8.56	0.317	0.03	10.60	0.232	0.03	11.42	0.180	0.02	11.44	0.105	0.01	10.67	0.212	0.03	11.79	11.00	7.16	65.09	0.97	0.20	20.37
			30-45	1.38	0.08	5.98	20.20	3.39	16.78	33.90	4.88	14.40	46.00	4.90	10.65	0.328	0.03	7.87	0.240	0.03	12.17	0.186	0.02	10.38	0.107	0.02	14.21	0.221	0.04	16.24	10.60	7.58	71.51	0.85	0.37	43.88
			45-60	1.38	0.08	5.56	21.70	7.56	34.84	34.20	6.97	20.38	44.10	9.61	21.79	0.345	0.03	7.74	0.272	0.04	12.90	0.202	0.01	5.79	0.112	0.02	21.96	0.233	0.04	17.90	11.70	9.55	81.62	0.67	0.50	75.11

Area	Field ID	n	Depth	pH			P			K			Mg			Ca			CEC			Al		
					SD	CV	ppm	SD	CV	ppm	SD	CV	ppm	SD	CV	ppm	SD	CV	cmol kg^{-1}	SD	CV	ppm	SD	CV
1	1	9	0-15	6.0	0.4	6.7	208.0	48.0	23.1	154.0	34.4	22.3	103.0	19.6	19.0	829.0	135.0	16.3	5.1	1.6	30.3	1048.0	156.0	14.9
			15-30	5.97	0.14	2.35	86.60	44.80	51.73	89.20	21.10	23.65	91.00	19.90	21.87	667.00	117.00	17.54	3.88	1.48	38.14	1102.00	165.00	14.97
			30-45	6.09	0.24	3.94	29.40	12.40	42.18	92.00	24.50	26.63	137.00	85.90	62.70	670.00	175.00	26.12	4.38	2.25	51.37	1061.00	171.00	16.12
			45-60	6.00	0.33	5.50	10.90	9.64	88.44	85.10	14.70	17.27	173.00	107.00	61.85	609.00	108.00	17.73	4.13	1.34	32.45	1041.00	175.00	16.81
			60-90	5.80	0.39	6.72	6.11	3.18	52.05	76.10	14.90	19.58	187.00	105.00	56.15	489.00	68.40	13.99	3.63	1.03	28.37	923.00	118.00	12.78
	2	9	0-15	6.38	0.42	6.60	34.30	12.60	36.73	143.00	43.10	30.14	179.00	41.60	23.24	668.00	130.00	19.46	4.40	1.12	25.45	710.00	143.00	20.14
			15-30	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na
			30-45	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na
			45-60	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na
			60-90	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na
	5	9	0-15	6.09	0.36	5.94	51.30	28.70	55.95	136.00	23.90	17.57	80.60	24.30	30.15	718.00	181.00	25.21	4.13	0.96	23.29	914.00	161.00	17.61
			15-30	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na
			30-45	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na
			45-60	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na
			60-90	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na
	6	9	0-15	5.69	0.25	4.34	32.60	25.90	79.45	132.00	34.20	25.91	125.00	31.00	24.80	503.00	140.00	27.83	4.82	2.13	44.19	1036.00	209.00	20.17
			15-30	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na
			30-45	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na
			45-60	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na
			60-90	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na
2	14	6	0-15	6.02	0.75	12.41	89.30	40.70	45.58	177.00	37.80	21.36	137.00	38.30	27.96	1023.00	350.00	34.21	8.55	1.42	16.61	1019.00	258.00	25.32
			15-30	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na
			30-45	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na
			45-60	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na
	16	6	0-15	6.05	0.52	8.66	59.50	26.90	45.21	172.00	44.50	25.87	269.00	61.40	22.83	1205.00	259.00	21.49	9.70	3.46	35.67	774.00	164.00	21.19
			15-30	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na
			30-45	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na
			45-60	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na
	19	6	0-15	6.73	0.23	3.48	77.20	37.30	48.32	103.00	37.40	36.31	226.00	27.20	12.04	1203.00	86.90	7.22	7.42	0.62	8.41	742.00	77.30	10.42
			15-30	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na
			30-45	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na
			45-60	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na
	21	6	0-15	6.53	0.52	7.90	53.20	25.80	48.50	156.00	47.50	30.45	280.00	81.60	29.14	1138.00	332.00	29.17	8.92	1.72	19.28	835.00	187.00	22.40
			15-30	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na
			30-45	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na
			45-60	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na
	22	6	0-15	5.73	0.35	6.11	21.50	7.74	36.00	109.00	25.50	23.39	121.00	23.10	19.09	519.00	56.90	10.96	4.67	1.13	24.20	759.00	126.00	16.60
			15-30	6.25	0.36	5.79	8.17	4.36	53.37	61.70	12.00	19.45	135.00	16.80	12.44	552.00	23.80	4.31	3.50	0.54	15.43	795.00	95.80	12.05
			30-45	6.48	0.26	4.07	4.33	1.51	34.87	48.80	6.31	12.93	101.00	20.10	19.90	499.00	62.00	12.42	3.17	0.64	20.22	700.00	67.20	9.60
			45-60	6.35	0.36	5.70	4.00	2.45	61.25	45.70	5.35	11.71	80.30	18.30	22.79	493.00	132.00	26.77	2.98	0.81	27.21	675.00	145.00	21.48

APPENDIX C. SUMMARY STATISTICS OF THE PROXIMAL SENSING INFORMATION AND ELEVATION, CHAPTER 4

		Area	Soil order	RED			IR			ECsh			ECdp			pH			Elevation		
																			m		
Area	Field ID	ha		Mean	Range	SD	Mean	Range	SD	Mean	Range	SD	Mean	Range	SD	Mean	Range	SD	Mean	Range	SD
1	1	17.2	Ultisols	277	265-293	4.68	563	526-597	8.98	3.92	1.64-10.1	0.93	4.22	1.82-13.3	1.14	6.61	5.59-7.17	0.23	18.6	17.2-19.8	0.52
	2	12	Ultisols	288	265-310	6.28	566	500-626	9.76	4.78	0.70-12.0	1.86	4.91	-0.98-16.4	1.79	6.97	5.90-7.70	0.25	19.0	17.6-20.2	0.38
	3	12.9	Ultisols	288	274-307	4.53	568	524-609	9.27	3.84	1.76-9.27	0.94	3.93	2.05-9.56	0.94	6.61	5.84-7.39	0.28	19.6	18.1-20.8	0.55
	4	38.6	Ultisols	271	238-287	4.87	555	429-596	12.7	7.09	1.59-138	3.43	6.86	-3.94-85.15	2.63	7.19	6.24-8.04	0.22	22.8	17.7-24.8	1.14
	5	7.10	Ultisols	306	289-326	7.65	538	483-580	23.3	5.39	2.31-10.6	1.74	6.25	2.62-12.2	1.95	6.48	5.61-7.45	0.26	25.2	22.9-27.5	1.31
	6	16.8	Ultisols	303	280-321	6.79	551	511-581	10.3	4.50	0.95-8.66	1.12	4.51	1.02-18.39	1.12	5.97	5.24-7.02	0.25	21.3	18.5-23.0	0.88
	7	27.7	Ultisols	297	277-313	5.69	553	495-581	8.68	4.08	1.40-7.55	0.95	4.19	1.56-15.3	0.96	6.13	5.31-7.22	0.24	23.2	20.4-25.5	1.14
	8	23.7	Ultisols	299	270-320	7.33	562	494-610	9.69	4.99	1.88-15.2	1.37	5.27	2.14-18.2	1.74	6.37	5.47-7.17	0.24	20.6	18.6-22.4	0.77
	9	14.2	Ultisols	274	253-285	3.73	547	456-591	9.63	3.65	1.78-7.47	1.07	4.11	2.20-8.21	1.14	6.05	5.07-6.60	0.16	20.2	18.0-21.4	0.68
	10	14.1	Ultisols	272	248-357	8.09	557	448-714	17.1	5.76	1.62-10.0	1.24	6.31	1.97-22.1	1.60	6.11	5.31-6.68	0.18	18.8	15.6-20.7	1.16
	11	47.2	Ultisols	248	217-291	8.31	562	444-681	24.4	5.25	0.44-32.0	4.31	5.91	-0.36-29.9	4.37	6.54	5.50-7.39	0.26	25.4	23.5-28.0	0.86
	12	23.9	Ultisols	287	278-302	2.81	564	530-598	8.72	5.94	1.84-15.6	1.57	7.04	1.96-16.6	1.69	6.43	4.77-7.38	0.44	20.4	19.0-21.6	0.57
	13	16.4	Ultisols	283	274-291	2.59	569	536-616	11.6	3.81	0.99-13.4	1.52	4.44	1.25-13.7	1.56	7.19	6.18-7.92	0.29	19.6	17.2-21.8	0.91
2	14	21.8	Alfisols/Inceptisols	296	282-375	6.82	587	536-652	19.5	9.31	1.98-25.1	3.61	10.3	2.06-27.3	4.08	na	na	na	110	106-115	2.03
	15	20.2	Alfisols	287	280-295	1.71	549	517-575	6.03	13.0	3.67-163	9.44	14.1	4.34-35.5	5.54	na	na	na	180	177-183	1.43
	16	12.8	Ultisols/Alfisols	277	260-298	5.96	567	529-616	13.7	7.17	2.54-15.2	2.23	9.87	3.08-123	6.73	7.07	5.99-7.74	0.23	166	154-173	4.41
	17	5.37	Ultisols/Alfisols	239	219-252	4.46	493	431-519	11.4	7.97	3.65-14.2	2.24	9.34	3.36-18.8	2.93	7.21	6.11-7.59	0.19	200	192-205	3.39
	18	13.1	Ultisols/Alfisols	243	222-257	4.00	496	421-552	12.6	9.20	1.89-22.9	3.40	11.4	2.32-26.3	4.04	6.66	5.85-7.30	0.21	164	154-170	3.39
	19	11.8	Alfisols/Inceptisols	287	280-293	1.61	549	524-577	5.72	8.89	3.61-14.6	1.51	9.52	3.91-15.4	1.54	na	na	na	137	129-143	2.76
	20	7.63	Ultisols/Alfisols	291	285-298	2.31	571	456-601	11.5	6.03	2.54-11.5	1.73	6.21	2.72-23.4	1.88	na	na	na	105	98.6-107	1.69
	21	17.3	Ultisols	317	286-407	11.8	668	532-947	46.6	8.57	3.37-19.7	2.34	7.64	3.83-15.6	1.81	na	na	na	166	157-173	3.24
	22	14.6	Ultisols	303	291-316	4.93	619	577-653	12.8	5.11	2.47-9.74	1.14	4.67	1.90-8.99	1.06	na	na	na	160	148-166	3.58
	23	7.73	Ultisols/Inceptisols	294	289-303	1.91	550	527-583	6.46	6.15	2.16-12.1	1.42	5.98	1.83-12.8	1.41	na	na	na	276	269-283	3.30
	24	6.07	Ultisols/Alfisols	232	223-241	2.68	489	450-524	9.66	10.1	3.07-26.4	3.61	13.2	4.90-63.1	5.14	6.50	5.83-7.31	0.28	150	144-157	3.21
	25	11.9	Ultisols/Alfisols	267	247-289	5.22	553	495-601	16.2	8.56	2.75-27.0	2.89	10.6	3.37-35.5	3.96	7.23	6.54-8.04	0.21	158	153-162	1.99
	26	11.0	Ultisols/Inceptisols	290	284-301	2.13	553	532-593	7.49	6.74	3.95-13.2	1.33	5.84	3.17-10.1	1.14	na	na	na	250	238-257	4.85

REFERENCES

Abuzar, M., Rampant, P., Fisher, P., 2004. Measuring spatial variability of crops and soils at sub-paddock scale using remote sensing technologies, in: Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. 2004 IEEE International. Presented at the Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. 2004 IEEE International, pp. 1633–1636 vol.3. doi:10.1109/IGARSS.2004.1370642

Adamchuk, V.I., Hummel, J.W., Morgan, M.T., Upadhyaya, S.K., 2004. On-the-go soil sensors for precision agriculture. Comput. Electron. Agric. 44, 71–91. doi:10.1016/j.compag.2004.03.002

Adamchuk, V.I., Morgan, M.T., Ess, D.R., 1999. An automated sampling system for measuring soil pH. Trans. ASAE 42, 885–891.

Adams, W.A., 1973. The effect of organic matter on the bulk and true densities of some uncultivated podzolic soils. J. Soil Sci. 24, 10–17. doi:10.1111/j.1365-2389.1973.tb00737.x

Alakukku, L., 2000. Response of annual crops to subsoil compaction in a field experiment on clay soil lasting 17 years, in: Horn, R., Van den Akker, J.J.., Arvidsson, J. (Eds.), Subsoil Compaction. Distribution, Processes and Consequences, Advances in Geoecology. pp. 205–208.

Allan Jones, C., 1983. Effect of soil texture on critical bulk densities for root growth. Soil Sci. Soc. Am. J. 47, 1208–1211. doi:10.2136/sssaj1983.03615995004700060029x

Allen, B.L., Fanning, D.S., 1983. Composition and Soil Genesis, in: L.P. Wilding, N.E.S. and G.F.H. (Ed.), Developments in Soil Science, Pedogenesis and Soil Taxonomy I. Concepts and Interactions. Elsevier, pp. 141–192.

Amézketa, E., 1999. Soil aggregate stability: a review. J. Sustain. Agric. 14, 83–151. doi:10.1300/J064v14n02_08

Andrews, S.S., Karlen, D.L., Cambardella, C.A., 2004. The soil management assessment framework: A quantitative soil quality evaluation method. Soil Sci. Soc. Am. J. 68, 1945–1962.

Angers, D.A., Bolinder, M.A., Carter, M.R., Gregorich, E.G., Drury, C.F., Liang, B.C., Voroney, R.P., Simard, R.R., Donald, R.G., Beyaert, R.P., Martel, J., 1997. Impact of tillage practices on organic carbon and nitrogen storage in cool, humid soils of eastern Canada. Soil Tillage Res. 41, 191–201. doi:10.1016/S0167-1987(96)01100-2

Angers, D.A., Pesant, A., Vigneux, J., 1992. Early cropping-induced changes in Soil aggregation, organic matter, and microbial biomass. Soil Sci. Soc. Am. J. 56, 115–119. doi:10.2136/sssaj1992.03615995005600010018x

Archer, J.R., Smith, P.D., 1972. The relation between bulk density, available water capacity, and air capacity of soils. J. Soil Sci. 23, 475–480. doi:10.1111/j.1365-2389.1972.tb01678.x

Arnholt, M.W., 2001. Evaluating Adoption and Uses of Precision Farming Technologies (MS thesis). Ohio State University, Columbus, OH.

Arslan, S., Colvin, T.S., 2002. Grain yield mapping: Yield sensing, yield reconstruction, and errors. Precis. Agric. 3, 135–154. doi:10.1023/A:1013819502827

Aziz, I., Mahmood, T., Islam, K.R., 2013. Effect of long term no-till and conventional tillage practices on soil quality. Soil Tillage Res. 131, 28–35. doi:10.1016/j.still.2013.03.002

Baker, J.M., Ochsner, T.E., Venterea, R.T., Griffis, T.J., 2007. Tillage and soil carbon sequestration—What do we really know? Agric. Ecosyst. Environ. 118, 1–5. doi:10.1016/j.agee.2006.05.014

Barthes, B.G., Brunet, D., Ferrer, H., Chotte, J.L., Feller, C., 2006. Determination of total carbon and nitrogen content in a range of tropical soils using near infrared spectroscopy: influence of replication and sample grinding and drying. J. Infrared Spectrosc. 14, 341–348.

Barthès, B.G., Kouakoua, E., Larré-Larrouy, M.-C., Razafimbelo, T.M., de Luca, E.F., Azontonde, A., Neves, C.S.V.J., de Freitas, P.L., Feller, C.L., 2008. Texture and sesquioxide effects on water-stable aggregates and organic matter in some tropical soils. Geoderma 143, 14–25. doi:10.1016/j.geoderma.2007.10.003

Bartoń, K., 2015. MuMIn: multi-model inference. R package, version 1.15.1.

Basso, B., Cammarano, D., Chen, D., Cafiero, G., Amato, M., Bitella, G., Rossi, R., Basso, F., 2009. Landscape position and precipitation effects on spatial variability of wheat yield and grain protein in southern Italy. J. Agron. Crop Sci. 195, 301–312. doi:10.1111/j.1439-037X.2008.00351.x

Batjes, N.H., 1996. Total carbon and nitrogen in the soils of the world. Eur. J. Soil Sci. 47, 151–163.

Baveye, P., 2002. Comment on "Modeling soil variation: past, present and future" by G.B.M. Heuvelink and R. Webster. Geoderma 109, 289–293. doi:10.1016/S0016-7061(02)00173-8

Baveye, P.C., Laba, M., 2015. Moving away from the geostatistical lamppost: Why, where, and how does the spatial heterogeneity of soils matter? Ecol. Model., Complexity of Soils and Hydrology in Ecosystems 298, 24–38. doi:10.1016/j.ecolmodel.2014.03.018

Beatty, M., Corey, R.B., 1962. Subsoil fertility of Wisconsin soils, in: Scott, W.E. (Ed.), Wisconsin Academy Review. Wisconsin Academy of Science, Madison, WI, pp. 22–23.

Beegle, D.B., Sharpley, A.N., 1999. Approaches to managing phosphorus to protect the environment, in: Linenfelter, D. (Ed.), Penn State Agronomic Field Diagnostic Clinic. Pennsylvania State Univ. College of Agric. Sci. Coop. Ext., University Park, PA, pp. 42–46.

Beer, J., Muschler, R., Kass, D., Somarriba, E., 1998. Shade management in coffee and cacao plantations. Agrofor. Syst. 38, 139–164.

Belcher, B.N., DeGaetano, A.T., 2005. A method to infer time of observation at US Cooperative Observer Network stations using model analyses. Int. J. Climatol. 25, 1237–1251. doi:10.1002/joc.1183

Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.-M., McBratney, A., 2010. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. TrAC Trends Anal. Chem. 29, 1073–1081. doi:10.1016/j.trac.2010.05.006

Bellon-Maurel, V., McBratney, A., 2011. Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils – Critical review and research perspectives. Soil Biol. Biochem. 43, 1398–1410. doi:10.1016/j.soilbio.2011.02.019

Ben-Dor, E., Inbar, Y., Chen, Y., 1997. The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400–2500 nm) during a controlled

decomposition process. Remote Sens. Environ. 61, 1–15. doi:10.1016/S0034-4257(96)00120-4
Benegas, L., Ilstedt, U., Roupsard, O., Jones, J.R., Malmer, A., 2014. The effects of trees on infiltrability and preferential flow in a coffee plantation and in pastures: two contrasting agrosylvopastoral systems in Central America. Agric. Ecosyst. Environ. 183, 185–196.
Bennett, O.L., Mathias, E.L., Henderlong, P.R., 1972. Effects of north- and south-facing slopes on yield of Kentucky bluegrass (Poa pratensis L.) with variable rate and time of nitrogen application. Agron. J. 64, 630. doi:10.2134/agronj1972.00021962006400050025x
Beven, K., Germann, P., 1982. Macropores and water flow in soils. Water Resour. Res. 18, 1311–1325. doi:10.1029/WR018i005p01311
Bianchini, A.A., Mallarino, A.P., 2002. Soil-sampling alternatives and variable-rate liming for a soybean–corn rotation. Agron. J. 94, 1355. doi:10.2134/agronj2002.1355
Bilgili, A.V., Akbas, F., van Es, H.M., 2011. Combined use of hyperspectral VNIR reflectance spectroscopy and kriging to predict soil variables spatially. Precis. Agric. 12, 395–420.
Bishop, J.L., Pieters, C.M., Edwards, J.O., 1994. Infrared spectroscopic analyses on the nature of water in montmorillonite. Clays Clay Miner. 42, 702–716.
Blackmore, S., 2000. The interpretation of trends from multiple yield maps. Comput. Electron. Agric. 26, 37–51. doi:10.1016/S0168-1699(99)00075-7
Blackmore, S., Godwin, R.J., Fountas, S., 2003. The analysis of spatial and temporal trends in yield map data over six years. Biosyst. Eng. 84, 455–466. doi:10.1016/S1537-5110(03)00038-2
Blevins, R.L., Thomas, G.W., Cornelius, P.L., 1977. Influence of no-tillage and nitrogen fertilization on certain soil properties after 5 years of continuous corn. Agron. J. 69, 383–386. doi:10.2134/agronj1977.00021962006900030013x
Boehm, M.M., Anderson, D.W., 1997. A landscape-scale study of soil quality in three prairie farming systems. Soil Sci. Soc. Am. J. 61, 1147–1159. doi:10.2136/sssaj1997.03615995006100040022x
Bonilla, C.A., Norman, J.M., Molling, C.C., 2007. Water erosion estimation in topographically complex landscapes: Model description and first verifications. Soil Sci. Soc. Am. J. 71, 1524. doi:10.2136/sssaj2006.0302
Bonilla, C.A., Norman, J.M., Molling, C.C., Karthikeyan, K.G., Miller, P.S., 2008. Testing a grid-based soil erosion model across topographically complex landscapes. Soil Sci. Soc. Am. J. 72, 1745. doi:10.2136/sssaj2007.0310
Borin, M., Bigon, E., 2002. Abatement of NO3–N concentration in agricultural waters by narrow buffer strips. Environ. Pollut. 117, 165–168. doi:10.1016/S0269-7491(01)00142-7
Bou Kheir, R., Greve, M.H., Bøcher, P.K., Greve, M.B., Larsen, R., McCloy, K., 2010. Predictive mapping of soil organic carbon in wet cultivated lands using classification-tree based models: The case study of Denmark. J. Environ. Manage. 91, 1150–1160. doi:10.1016/j.jenvman.2010.01.001
Boudot, J.-P., 1992. Relative efficiency of complexed aluminum noncrystalline Al hydroxide, allophane and imogolite in retarding the biodegradation of citric acid. Geoderma 52, 29–39. doi:10.1016/0016-7061(92)90073-G
Boyer, D.G., Wright, R.J., Winant, W.M., Perry, H.D., 1990. Soil water relations on a hilltop cornfield in central Appalachia. Soil Sci. 149, 383–392.
Brady, N.C., Weil, R.R., 2008. The nature and properties of soils, Rev. 14th. ed. Prentice Hall, Upper Saddle River, N.J.

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. doi:10.1023/A:1010933404324
Breiman, L., 1996. Bagging Predictors. Mach. Learn. 24, 123–140. doi:10.1023/A:1018054314350
Breiman, L., Friedman, J.H., Olshen, R., Stone, C., 1983. Classification and Regression Trees. CRC Press, Boca Raton.
Brock, A., Brouder, S.M., Blumhoff, G., Hofmann, B.S., 2005. Defining yield-based management zones for corn–soybean rotations. Agron. J. 97, 1115. doi:10.2134/agronj2004.0220
Brown, D.J., Bricklemyer, R.S., Miller, P.R., 2005. Validation requirements for diffuse reflectance soil characterization models with a case study of VNIR soil C prediction in Montana. Geoderma 129, 251–267. doi:10.1016/j.geoderma.2005.01.001
Brown, D.J., Shepherd, K.D., Walsh, M.G., Dewayne Mays, M., Reinsch, T.G., 2006. Global soil characterization with VNIR diffuse reflectance spectroscopy. Geoderma 132, 273–290. doi:10.1016/j.geoderma.2005.04.025
Brunet, D., Barthes, B.G., Chotte, J.L., Feller, C., 2007. Determination of carbon and nitrogen contents in Alfisols, Oxisols and Ultisols from Africa and Brazil using NIRS analysis: Effects of sample grinding and set heterogeneity. Geoderma 139, 106–117. doi:10.1016/J.Geoderma.2007.01.007
Buurman, P., Peterse, F., Martin, G.A., 2007. Soil organic matter chemistry in allophanic soils: a pyrolysis-GC/MS study of a Costa Rican Andosol catena. Eur. J. Soil Sci. 58, 1330–1347. doi:10.1111/J.1365-2389.2007.00925.X
Calviño, P.A., Andrade, F.H., Sadras, V.O., 2003. Maize yield as affected by water availability, soil depth, and crop management. Agron J 95, 275–281.
Calviño, P.A., Sadras, V.O., 1999. Interannual variation in soybean yield: interaction among rainfall, soil depth and crop management. Field Crops Res. 63, 237–246. doi:10.1016/S0378-4290(99)00040-4
Cambardella, C.A., Moorman, T.B., Karlen, D.L., Novak, J.M., Turco, R.F., Konopka, A.E., 1994. Field-scale variability of soil properties in central Iowa soils. Soil Sci. Soc. Am. J. 58, 1501. doi:10.2136/sssaj1994.03615995005800050033x
Cambule, A.H., Rossiter, D.G., Stoorvogel, J.J., Smaling, E.M.A., 2014. Soil organic carbon stocks in the Limpopo National Park, Mozambique: Amount, spatial distribution and uncertainty. Geoderma 213, 46–56. doi:10.1016/j.geoderma.2013.07.015
Cañasveras, J.C., Barrón, V., del Campillo, M.C., Torrent, J., Gómez, J.A., 2010. Estimation of aggregate stability indices in Mediterranean soils by diffuse reflectance spectroscopy. Geoderma, Diffuse reflectance spectroscopy in soil science and land resource assessment 158, 78–84. doi:10.1016/j.geoderma.2009.09.004
Carter, M.R., 2002. Soil quality for sustainable land management: Organic matter and aggregation interactions that maintain soil functions. Agron. J. 94, 38–47.
Carter, M.R., Gregorich, E.G., 2010. Carbon and nitrogen storage by deep-rooted tall fescue (Lolium arundinaceum) in the surface and subsurface soil of a fine sandy loam in eastern Canada. Agric. Ecosyst. Environ. 136, 125–132. doi:10.1016/j.agee.2009.12.005
Casteel, S., 2011. Soybean Physiology: How Well Do You Know Soybeans?
CENIGA, 1998. Hojas Topográficas Escala 1:25000. Proy. TERRA.
Chang, C.W., Laird, D.A., Mausbach, M.J., Hurburgh, C.R., 2001. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. Soil Sci. Soc. Am. J. 65, 480–490.

Chen, X.-P., Cui, Z.-L., Vitousek, P.M., Cassman, K.G., Matson, P.A., Bai, J.-S., Meng, Q.-F., Hou, P., Yue, S.-C., Römheld, V., Zhang, F.-S., 2011. Integrated soil–crop system management for food security. Proc. Natl. Acad. Sci. 108, 6399–6404. doi:10.1073/pnas.1101419108

Chesworth, W., 2008. Encyclopedia of Soil Science. Springer Netherlands, Dordrecht.

Chevallier, T., Woignier, T., Toucet, J., Blanchart, E., 2010. Organic carbon stabilization in the fractal pore structure of Andosols. Geoderma 159, 182–188.

Chevallier, T., Woignier, T., Toucet, J., Blanchart, E., Dieudonné, P., 2008. Fractal structure in natural gels: effect on carbon sequestration in volcanic soils. J. Sol-Gel Sci. Technol. 48, 231–238. doi:10.1007/s10971-008-1795-z

Ciani, A., Goss, K.-U., Schwarzenbach, R.P., 2005. Light penetration in soil and particulate minerals. Eur. J. Soil Sci. 56, 561–574. doi:10.1111/j.1365-2389.2005.00688.x

Clark, R.N., 1999. Spectroscopy of rocks and minerals and principles of spectroscopy, in: Rencz, A.N. (Ed.), Remote Sensing for the Earth Sciences: Mannual of Remote Sensing. John Wiley & Sons, Chichester, pp. 3–58.

Clark, R.N., King, T.V.V., Klejwa, M., Swayze, G.A., Vergo, N., 1990. High spectral resolution reflectance spectroscopy of minerals. J. Geophys. Res. Solid Earth 95, 12653–12680. doi:10.1029/JB095iB08p12653

Conant, R.T., Smith, G.R., Paustian, K., 2003. Spatial variability of soil carbon in forested in cultivated sites: Implications for change detection. J. Environ. Qual. 32, 278–286.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for automated geoscientific analyses (SAGA) v. 2.1.4. Geosci. Model Dev. 8, 1991–2007.

Cornell University Cooperative Extension, 2012. Cornell Guide for Integrated Field Crop Management. Cornell Univ. Coop. Ext., Ithaca, NY.

Corwin, D.L., Lesch, S.M., 2005. Apparent soil electrical conductivity measurements in agriculture. Comput. Electron. Agric. 46, 11–43. doi:10.1016/j.compag.2004.10.005

Corwin, D.L., Lesch, S.M., 2003. Application of soil electrical conductivity to precision agriculture: Theory, principles, and guidelines. Agron. J. 95, 455–471.

Cox, M.S., Gerard, P.D., 2007. Soil management zone determination by yield stability analysis and classification. Agron. J. 99, 1357. doi:10.2134/agronj2007.0041

Cozzolino, D., Morón, A., 2003. The potential of near-infrared reflectance spectroscopy to analyse soil chemical and physical characteristics. J. Agric. Sci. 140.

Culman, S.W., Snapp, S.S., Freeman, M.A., Schipanski, M.E., Beniston, J., Lal, R., Drinkwater, L.E., Franzluebbers, A.J., Glover, J.D., Grandy, A.S., Lee, J., Six, J., Maul, J.E., Mirksy, S.B., Spargo, J.T., Wander, M.M., 2012. Permanganate oxidizable carbon reflects a processed soil fraction that is sensitive to management. Soil Sci. Soc. Am. J. 76, 494–504. doi:10.2136/sssaj2011.0286

Culman, S.W., Snapp, S.S., Green, J.M., Gentry, L.E., 2013. Short- and long-term labile soil carbon and nitrogen dynamics reflect management and predict corn agronomic performance. Agron J 105, 493–502. doi:10.2134/agronj2012.0382

da Silva, A.P., Kay, B.D., Perfect, E., 1994. Characterization of the least limiting water range of soils. Soil Sci. Soc. Am. J. 58, 1775–1781. doi:10.2136/sssaj1994.03615995005800060028x

Dalal, R.C., Henry, R.J., 1986. Simultaneous determination of moisture, organic carbon, and total nitrogen by near-infrared reflectance spectrophotometry. Soil Sci. Soc. Am. J. 50, 120–123.
Dane, J.H., Hopmans, J.W., 2002. Pressure plate extractor, in: Dane, J.H., Topp, G.C. (Eds.), Methods of Soil Analysis: Part 4 Physical Methods, SSSA Book Series. Soil Science Society of America, Inc., Madison, WI, pp. 688–690.
Davis, J.C., 1986. Statistics and Data Analysis in Geology, 2nd ed. Wiley, New York.
Dick, W.A., Thavamani, B., Conley, S., Blaisdell, R., Sengupta, A., 2013. Prediction of β-glucosidase and β-glucosaminidase activities, soil organic C, and amino sugar N in a diverse population of soils using near infrared reflectance spectroscopy. Soil Biol. Biochem. 56, 99–104. doi:10.1016/j.soilbio.2012.04.003
Duiker, S.W., Rhoton, F.E., Torrent, J., Smeck, N.E., Lal, R., 2003. Iron (hydr)oxide crystallinity effects on soil aggregation. Soil Sci. Soc. Am. J. 67, 606–611. doi:10.2136/sssaj2003.6060
Eastman, J.R., Filk, M., 1993. Long sequence time series evaluation using standardized principal components. Photogramm. Eng. Remote Sens. 59, 991–996.
Egli, D.B., Bruening, W.P., 2000. Potential of early-maturing soybean cultivars in late plantings. Agron. J. 92, 532. doi:10.2134/agronj2000.923532x
Erickson, B., Widmar, D.A., 2015. 2015 Precision Agricultural Services Dealership Survey Results. Purdue University, West Lafayette, IN.
Erskine, R.H., Green, T.R., Ramirez, J.A., MacDonald, L.H., 2007. Digital elevation accuracy and grid cell size: effects on estimated terrain attributes. Soil Sci. Soc. Am. J. 71, 1371. doi:10.2136/sssaj2005.0142
Evans, J.S., Oakleaf, J., 2011. ArcGIS - Geomorphometry and Gradient Metrics Toolbox [WWW Document]. URL http://conserveonline.org/workspaces/emt/documents/all.html (accessed 1.1.04).
Ewing, R.P., Wagger, M.G., Denton, H.P., 1991. Tillage and cover crop management effects on soil-water and corn yield. Soil Sci. Soc. Am. J. 55, 1081–1085.
Fenneman, N.M., 1938. Physiography of the Eastern United States. McGraw-Hill Book Company, Inc, New York.
Fischer, D., Uksa, M., Tischler, W., Kautz, T., Köpke, U., Schloter, M., 2013. Abundance of ammonia oxidizing microbes and denitrifiers in different soil horizons of an agricultural soil in relation to the cultivated crops. Biol. Fertil. Soils 1–4. doi:10.1007/s00374-013-0812-8
Florin, M.J., McBratney, A.B., Whelan, B.M., 2009. Quantification and comparison of wheat yield variation across space and time. Eur. J. Agron. 30, 212–219. doi:10.1016/j.eja.2008.10.003
Fontaine, S., Barot, S., Barré, P., Bdioui, N., Mary, B., Rumpel, C., 2007. Stability of organic carbon in deep soil layers controlled by fresh carbon supply. Nature 450, 277–280. doi:10.1038/nature06275
Foss, J.E., Fanning, D.S., Miller, F.P., Wagner, D.P., 1978. Loess deposits of the Eastern Shore of Maryland. Soil Sci. Soc. Am. J. 42, 329. doi:10.2136/sssaj1978.03615995004200020026x
Fourty, T., Baret, F., Jacquemoud, S., Schmuck, G., Verdebout, J., 1996. Leaf optical properties with explicit description of its biochemical composition: Direct and inverse problems. Remote Sens. Environ. 56, 104–117. doi:10.1016/0034-4257(95)00234-0

Franzluebbers, A.J., Hons, F.M., 1996. Soil-profile distribution of primary and secondary plant-available nutrients under conventional and no tillage. Soil Tillage Res. 39, 229–239.
Franzmeier, D.P., Pedersen, E.J., Longwell, T.J., Byrne, J.G., Losche, C.K., 1969. Properties of some soils in the Cumberland Plateau as related to slope aspect and position. Soil Sci. Soc. Am. J. 33, 755. doi:10.2136/sssaj1969.03615995003300050037x
French, R., Schultz, J., 1984a. Water use efficiency of wheat in a Mediterranean-type environment. I. The relation between yield, water use and climate. Aust. J. Agric. Res. 35, 743–764.
French, R., Schultz, J., 1984b. Water use efficiency of wheat in a Mediterranean-type environment. II. some limitations to efficiency. Aust. J. Agric. Res. 35, 765–775.
Gaiser, T., Perkons, U., Küpper, P.M., Puschmann, D.U., Peth, S., Kautz, T., Pfeifer, J., Ewert, F., Horn, R., Köpke, U., 2012. Evidence of improved water uptake from subsoil by spring wheat following lucerne in a temperate humid climate. Field Crops Res. 126, 56–62. doi:10.1016/j.fcr.2011.09.019
Gallant, J.C., Wilson, J.P., 2000. Primary topographic attributes, in: Wilson, J.P., Gallant, J.C. (Eds.), Terrain Analysis: Principles and Applications. Wiley, New York, pp. 51–85.
Galvao, L.S., Vitorello, I., 1998. Role of organic matter in obliterating the effects of iron on spectral reflectance and colour of Brazilian tropical soils. Int. J. Remote Sens. 19, 1969–1979. doi:10.1080/014311698215090
Gandah, M., Stein, A., Brouwer, J., Bouma, J., 2000. Dynamics of spatial variability of millet growth and yields at three sites in Niger, West Africa and implications for precision agriculture research. Agric. Syst. 63, 123–140. doi:10.1016/S0308-521X(99)00076-1
Garz, J., Schliephake, W., Merbach, W., 2000. Changes in the subsoil of long-term trials in Halle (Saale), Germany, caused by mineral fertilization. J. Plant Nutr. Soil Sci. 163, 663–668. doi:10.1002/1522-2624
Geladi, P., Kowalski, B.R., 1986. Partial least-squares regression: a tutorial. Anal. Chim. Acta 185, 1–17. doi:10.1016/0003-2670(86)80028-9
Gessler, P.E., Chadwick, O.A., Charman, F., Althouse, L., Holmes, K., 2000. Modeling soil-landscape and ecosystem properties using terrain attributes. Soil Sci. Soc. Am. J. 64, 2046–2056.
Gilbert, C.L., Morgan, C.W., 2010. Food price volatility. Philos. Trans. R. Soc. B 365, 3023–3034. doi:10.1098/rstb.2010.0139
Gilbert, R.O., 1987. Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold Co, New York.
Gobrecht, A., Roger, J.-M., Bellon-Maurel, V., 2014. Major issues of diffuse reflectance NIR spectroscopy in the specific context of soil carbon content estimation: A review. Adv. Agron. 123, 145–175.
Gómez-Delgado, F., 2010. Hydrological, Ecophysiological and Sediment Processes in a Coffee Agroforestry Basin: Combining Experimental and Modelling Methods to Assess Hydrological Environmental Services. (Ph.D. Dissertation). Supagro, Montpellier.
Gómez-Delgado, F., Roupsard, O., Moussa, R., le Maire, G., Taugourdeau, S., Bonnefond, J.M., Pérez, A., van Oijen, M., Vaast, P., Rapidel, B., Voltz, M., Imbach, P., Harmand, J.M., 2011. Modelling the hydrological behaviour of a coffee agroforestry basin in Costa Rica. Hydrol. Earth Syst. Sci. 15, 369–392.

Good, D., Irwin, S., 2012. The Historic Pattern of U.S Soybean Yields, Any Implications for 2012? [WWW Document]. URL http://farmdocdaily.illinois.edu/2012/02/the-historic-pattern-of-us-soy.html (accessed 4.11.15).
Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York.
Graham, C.J., 2012. An Examination of Various Segments of the Nitrogen Cycle in Diverse Agro-Ecosystems (MS thesis). Cornell University, Ithaca.
Gransee, A., Merbach, W., 2000. Phosphorus dynamics in a long-term P fertilization trial on Luvic Phaeozem at Halle. J. Plant Nutr. Soil Sci. 163, 353–357. doi:10.1002/1522-2624
Grassini, P., Eskridge, K.M., Cassman, K.G., 2013. Distinguishing between yield advances and yield plateaus in historical crop production trends. Nat. Commun. 4. doi:10.1038/ncomms3918
Grassini, P., Torrion, J.A., Yang, H.S., Rees, J., Andersen, D., Cassman, K.G., Specht, J.E., 2015. Soybean yield gaps and water productivity in the western U.S. Corn Belt. Field Crops Res. 179, 150–163. doi:10.1016/j.fcr.2015.04.015
Grime, J.P., 1998. Annual Report, 1996-1998. Unit of Comparative Plant Ecology. The University of Sheffield, Sheffield.
Grimes, D.W., Miller, R.J., Wiley, P.L., 1975. Cotton and corn root development in two field soils of different strength characteristics. Agron. J. 67, 519–523.
Grimm, R., Behrens, T., Märker, M., Elsenbeer, H., 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island — Digital soil mapping using Random Forests analysis. Geoderma 146, 102–113. doi:10.1016/j.geoderma.2008.05.008
Guber, A.K., Rawls, W.J., Shein, E.V., Pachepsky, Y.A., 2003. Effect of soil aggregate size distribution on water retention. Soil Sci. 168, 223–233. doi:10.1097/01.ss.0000064887.94869.d3
Guerrero, C., Stenberg, B., Wetterlind, J., Viscarra Rossel, R.A., Maestre, F.T., Mouazen, A.M., Zornoza, R., Ruiz-Sinoga, J.D., Kuang, B., 2014. Assessment of soil organic carbon at local scale with spiked NIR calibrations: effects of selection and extra-weighting on the spiking subset. Eur. J. Soil Sci. 65, 248–263. doi:10.1111/ejss.12129
Guerrero, C., Zornoza, R., Gómez, I., Mataix-Beneyto, J., 2010. Spiking of NIR regional models using samples from target sites: Effect of model size on prediction accuracy. Geoderma 158, 66–77. doi:10.1016/j.geoderma.2009.12.021
Gugino, B.K., Idowu, O.J., Schindelbeck, R.R., van Es, H.M., Wolfe, D.W., Moebius-Clune, B.N., Thies, J.E., Abawi, G.S., 2009. Cornell Soil Health Assessment Training Manual, Edition 2.0. Cornell University, Geneva, NY.
Hack, J.T., 1980. Rock control and tectonism - their importance in shaping the Appalachian Highlands. US Geol. Surv. Prof. Pap. 1126B, 16.
Haney, R.L., Haney, E.B., 2010. Simple and rapid laboratory method for rewetting dry soil for incubations. Commun. Soil Sci. Plant Anal. 41, 1493–1501. doi:10.1080/00103624.2010.482171
Hassink, J., 1997. The capacity of soils to preserve organic C and N by their association with clay and silt particles. Plant Soil 191, 77–87. doi:10.1023/A:1004213929699
Havlin, J., 2005. Soil Fertility and Fertilizers : An Introduction to Nutrient Management, 7th ed. Pearson Prentice Hall, Upper Saddle River.
Heming, S.D., 2004. Potassium balances for arable soils in southern England 1986–1999. Soil Use Manag. 20, 410–417. doi:10.1111/j.1475-2743.2004.tb00390.x

Hengl, T., Heuvelink, G.B.M., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., Mendes de Jesus, J., Tamene, L., Tondoh, J.E., 2015. Mapping soil properties of Africa at 250 m resolution: Random Forests significantly improve current predictions. PLoS ONE 10. doi:10.1371/journal.pone.0125814

Hengl, T., Heuvelink, G.B.M., Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma 120, 75–93.

Heuvelink, G.B.M., Webster, R., 2002. Reply to Comment on "Modelling soil variation: past, present, and future", by Philippe Baveye. Geoderma 109, 295–297. doi:10.1016/S0016-7061(02)00174-X

Hijmans, R.J., Phillips, S., Elith, J.L. and J., 2013. dismo: Species distribution modeling.

Hillel, D., 1980. Fundamentals of Soil Physics, 1st ed. Academic Press, New York.

Hillel, D., Rosenzweig, C., 2010. Handbook of Climate Change and Agroecosystems. Imperial College Press, London.

Hochman, Z., Gobbett, D., Holzworth, D., McClelland, T., van Rees, H., Marinoni, O., Garcia, J.N., Horan, H., 2012. Quantifying yield gaps in rainfed cropping systems: A case study of wheat in Australia. Field Crops Res. 136, 85–96. doi:10.1016/j.fcr.2012.07.008

Hong, N., Scharf, P.C., Davis, J.G., Kitchen, N.R., Sudduth, K.A., 2007. Economically optimal nitrogen rate reduces soil residual nitrate. J. Environ. Qual. 36, 354. doi:10.2134/jeq2006.0173

Horwath Burnham, J., Sletten, R.S., 2010. Spatial distribution of soil organic carbon in northwest Greenland and underestimates of high Arctic carbon stores. Glob. Biogeochem. Cycles 24, GB3012. doi:10.1029/2009GB003660

Hudson, B.D., 1994. Soil organic matter and available water capacity. J. Soil Water Conserv. 49, 189–194.

Hunt, G.R., 1977. Spectral signatures of particulate minerals in the visible and near infrared. Geophysics 42, 501–513. doi:10.1190/1.1440721

Huygens, D., Boeckx, P., Van Cleemput, O., Oyarzun, C., Godoy, R., 2005. Aggregate and soil organic carbon dynamics in South Chilean Andisols. Biogeosciences 2, 159–174.

ICRAF and ISRIC - World Soil Information, 2010. ICRAF-ISRIC Soil VNIR Spectral Library. ICRAF, Nairobi.

Idowu, O.J., van Es, H.M., Abawi, G.S., Wolfe, D.W., Ball, J.I., Gugino, B.K., Moebius, B.N., Schindelbeck, R.R., Bilgili, A.V., 2008. Farmer-oriented assessment of soil quality using field, laboratory, and VNIR spectroscopy methods. Plant Soil 307, 243–253.

Idowu, O.J., van Es, H.M., Abawi, G.S., Wolfe, D.W., Schindelbeck, R.R., Moebius-Clune, B.N., Gugino, B.K., 2009. Use of an integrative soil health test for evaluation of soil management impacts. Renew. Agric. Food Syst. 24, 214–224. doi:10.1017/S1742170509990068

ISSS-ISRIC-FAO, 1998. World Reference Base for Soil Resources, World Soil Resources Reports. FAO, Rome.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning. Springer, New York.

Jiang, P., Thelen, K.D., 2004. Effect of soil and topographic properties on crop yield in a north-central corn–soybean cropping system. Agron. J. 96, 252–258. doi:10.2134/agronj2004.0252

John, B., Yamashita, T., Ludwig, B., Flessa, H., 2005. Storage of organic carbon in aggregate and density fractions of silty soils under different types of land use. Geoderma 128, 63–79. doi:10.1016/j.geoderma.2004.12.013

Johnson, C.K., Doran, J.W., Duke, H.R., Wienhold, B.J., Eskridge, K.M., Shanahan, J.F., 2001. Field-scale electrical conductivity mapping for delineating soil condition. Soil Sci. Soc. Am. J. 65, 1829. doi:10.2136/sssaj2001.1829

Jokela, B., Magdoff, F., Bartlett, R., Bosworth, S., Ross, D., 2004. Nutrient Recommendations for Field Crop in Vermont. The University of Vermont Extension, Burlington.

Jozefaciuk, G., Czachor, H., 2014. Impact of organic matter, iron oxides, alumina, silica and drying on mechanical and water stability of artificial soil aggregates. Assessment of new method to study water stability. Geoderma 221–222, 1–10. doi:10.1016/j.geoderma.2014.01.020

Jung, W.K., Kitchen, N.R., Sudduth, K.A., Kremer, R.J., Motavalli, P.P., 2005. Relationship of apparent soil electrical conductivity to claypan soil properties. Soil Sci. Soc. Am. J. 69, 883. doi:10.2136/sssaj2004.0202

Kaiser, H.F., 1960. The application of electronic computer to factor analysis. Educ. Psychol. Meas. 20, 141–151.

Kane, M.V., Steele, C.C., Grabau, L.J., 1997. Early-maturing soybean cropping system: I. yield responses to planting date. Agron. J. 89, 454. doi:10.2134/agronj1997.00021962008900030014x

Karlen, D.L., Andrews, S.S., Doran, J.W., 2001. Soil quality: current concepts and applications. Adv. Agron. 74, 1–40.

Karlen, D.L., Mausbach, M.J., Doran, J.W., Cline, R.G., Harris, R.F., Schuman, G.E., 1997. Soil quality: A concept, definition, and framework for evaluation. Soil Sci Soc Am J 61, 4–10.

Katsvairo, T.W., Cox, W.J., Van Es, H.M., Glos, M., 2003. Spatial yield response of two corn hybrids at two nitrogen levels. Agron. J. 95, 1012. doi:10.2134/agronj2003.1012

Kaul, M., Hill, R.L., Walthall, C., 2005. Artificial neural networks for corn and soybean yield prediction. Agric. Syst. 85, 1–18. doi:10.1016/j.agsy.2004.07.009

Kautz, T., Amelung, W., Ewert, F., Gaiser, T., Horn, R., Jahn, R., Javaux, M., Kemna, A., Kuzyakov, Y., Munch, J.-C., Pätzold, S., Peth, S., Scherer, H.W., Schloter, M., Schneider, H., Vanderborght, J., Vetterlein, D., Walter, A., Wiesenberg, G.L.B., Köpke, U., 2013. Nutrient acquisition from arable subsoils in temperate climates: a review. Soil Biol. Biochem. 57, 1003–1022. doi:10.1016/j.soilbio.2012.09.014

Kemper, W.D., Rosenau, R.C., 1986. Aggregate stability and size distribution, in: Klute, A. (Ed.), Methods of Soil Analysis - Part 1. Physical and Mineralogical Methods. Soil Science Society of America, Inc., Madison, pp. 425–442.

Kettler, T.A., Doran, J.W., Gilbert, T.L., 2001. Simplified method for soil particle-size determination to accompany soil-quality analyses. Soil Sci. Soc. Am. J. 65, 849–852.

Kinoshita, R., Moebius-Clune, B.N., van Es, H.M., Hively, W.D., Bilgili, A.V., 2012. Strategies for soil quality assessment using visible and near-infrared reflectance spectroscopy in a Western Kenya chronosequence. Soil Sci. Soc. Am. J. 76, 1776–1788. doi:10.2136/sssaj2011.0307

Kinoshita, R., Roupsard, O., Chevallier, T., Albrecht, A., Taugourdeau, S., Ahmed, Z., van Es, H.M., 2016. Large topsoil organic carbon variability is controlled by Andisol properties and effectively assessed by VNIR spectroscopy in a coffee agroforestry system of Costa Rica. Geoderma 262, 254–265. doi:10.1016/j.geoderma.2015.08.026

Kirkegaard, J.A., Lilley, J.M., Howe, G.N., Graham, J.M., 2007. Impact of subsoil water use on wheat yield. Aust. J. Agric. Res. 58, 303–315.
Kitchen, N.R., Drummond, S.T., Lund, E.D., Sudduth, K.A., Buchleiter, G.W., 2003. Soil electrical conductivity and topography related to yield for three contrasting soil-crop systems. Agron. J. 95, 483–495.
Knadel, M., Viscarra Rossel, R.A., Deng, F., Thomsen, A., Greve, M.H., 2013. Visible–near infrared spectra as a proxy for topsoil texture and glacial boundaries. Soil Sci. Soc. Am. J. 77, 568. doi:10.2136/sssaj2012.0093
Kooistra, L., Wehrens, R., Leuven, R.S.E.W., Buydens, L.M.C., 2001. Possibilities of visible-near-infrared spectroscopy for assessment of soil contamination in river floodplains. Anal. Chim. Acta 446, 97–105.
Kravchenko, A.N., Bullock, D.G., 2000. Correlation of corn and soybean grain yield with topography and soil properties. Agron. J. 92, 75. doi:10.2134/agronj2000.92175x
Kuang, B., Mahmood, H.S., Quraishi, M.Z., Hoogmoed, W.B., Mouazen, A.M., van Henten, J., 2012. Sensing soil properties in the laboratory, in situ, and on-line: a review. Adv. Agron. 114, 155–223.
Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling. Springer, New York.
Kunkel, M.L., Flores, A.N., Smith, T.J., McNamara, J.P., Benner, S.G., 2011. A simplified approach for estimating soil carbon and nitrogen stocks in semi-arid complex terrain. Geoderma 165, 1–11. doi:10.1016/j.geoderma.2011.06.011
Kusumo, B.H., Hedley, M.J., Tuohy, M.P., Hedley, C.B., Arnold, G.C., 2010. Predicting soil carbon and nitrogen concentrations and pasture root densities from proximally sensed soil spectral reference, in: Viscarra Rossel, R.A., McBratney, A.B., Minasny, B. (Eds.), Proximal Soil Sensing. Springer, Dordrecht, pp. 177–190.
Kweon, G., Lund, E., Maxton, C., 2013. Soil organic matter and cation-exchange capacity sensing with on-the-go electrical conductivity and optical sensors. Geoderma 199, 80–89. doi:10.1016/j.geoderma.2012.11.001
Kweon, G., Maxton, C., 2013. Soil organic matter sensing with an on-the-go optical sensor. Biosyst. Eng. 115, 66–81. doi:10.1016/j.biosystemseng.2013.02.004
Lal, R., 2006. Enhancing crop yields in the developing countries through restoration of the soil organic carbon pool in agricultural lands. Land Degrad. Dev. 17, 197–209. doi:10.1002/ldr.696
Lamsal, S., 2009. Visible near-infrared reflectance spectrocopy for geospatial mapping of soil organic matter. Soil Sci. 174, 35–44.
Lauzon, J.D., O'Halloran, I.P., Fallow, D.J., von Bertoldi, A.P., Aspinall, D., 2005. Spatial variability of soil test phosphorus, potassium, and pH of Ontario soils. Agron. J. 97, 524. doi:10.2134/agronj2005.0524
Lawes, R.A., Oliver, Y.M., Robertson, M.J., 2009. Capturing the in-field spatial–temporal dynamic of yield variation. Crop Pasture Sci. 60, 834–843.
Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G.W., Prosser, J.I., Schuster, S.C., Schleper, C., 2006. Archaea predominate among ammonia-oxidizing prokaryotes in soils. Nature 442, 806–809. doi:10.1038/nature04983
Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. R News 2, 18–22.
Lory, J.A., Scharf, P.C., 2003. Yield goal versus delta yield for predicting nitrogen fertilizer need in corn. Agron. J. 95, 994–999.

Ma, B.-L., Wu, T.-Y., Shang, J., 2013. On-farm comparison of variable rates of nitrogen with uniform application to maize on canopy reflectance, soil nitrate, and grain yield. J. Plant Nutr. Soil Sci. n/a–n/a. doi:10.1002/jpln.201200338

Machado-Machado, E.A., Neeti, N., Eastman, J.R., Chen, H., 2011. Implications of space-time orientation for Principal Components Analysis of Earth observation image time series. Earth Sci. Inform. 4, 117–124. doi:10.1007/s12145-011-0082-7

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations, in: LeCam, L.M., Neyman, J. (Eds.), Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability. Vol. 1: Statistics. Univ. of California Press, Berkeley, pp. 281–297.

Magdoff, F.R., van Es, H.M., 2009. Building Soils for Better Crops. SARE Outreach, College Park.

Markewich, H.W., Pavich, M.J., Buell, G.R., 1990. Proceedings of the 21st Annual Binghamton Symposium in Geomorphology Contrasting soils and landscapes of the Piedmont and Coastal Plain, eastern United States. Geomorphology 3, 417–447. doi:10.1016/0169-555X(90)90015-I

Mausback, M., Tugel, A., 1995. Decision document for establishing a Soil Quality Institute. White Paper. Natural Resources Conservation Service, Washington DC.

McDowell, M.L., Bruland, G.L., Deenik, J.L., Grunwald, S., Knox, N.M., 2012. Soil total carbon analysis in Hawaiian soils with visible, near-infrared and mid-infrared diffuse reflectance spectroscopy. Geoderma 189–190, 312–320. doi:10.1016/j.geoderma.2012.06.009

McIntosh, J.L., 1969. Bray and Morgan soil extractants modified for testing acid soils from different parent materials. Agron. J. 61, 259–265. doi:10.2134/agronj1969.00021962006100020025x

Melkonian, J., van Es, H., DeGaetano, A., Joseph, L., 2008. ADAPT-N: Adaptive nitrogen management for maize using high-resolution climate data and model simulations. Presented at the Proceedings of the 9th International Conference on Precision Agriculture.

Melkonian, J., van Es, H.M., Joseph, L., 2005. Precision Nitrogen Management model: simulation of nitrogen and water fluxes in the soil-crop-atmosphere continuum in maize (Zea mays L.) production systems., Version 1.0. Dept. of Crop and Soil Sciences, Research series No. R05-2. Cornell University, Ithaca.

Mengel, K., Scherer, H.W., 1981. Release of nonexchangeable (fixed) soil ammonium under field conditions during the growing season. Soil Sci. 131, 226–232.

Miller, M.P., Singer, M.J., Nielsen, D.R., 1988. Spatial variability of wheat yield and soil properties on complex hills. Soil Sci. Soc. Am. J. 52, 1133. doi:10.2136/sssaj1988.03615995005200040045x

Minasny, B., McBratney, A.B., Bellon-Maurel, V., Roger, J.-M., Gobrecht, A., Ferrand, L., Joalland, S., 2011. Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon. Geoderma 167–168, 118–124. doi:10.1016/j.geoderma.2011.09.008

Minasny, B., McBratney, A.B., Malone, B.P., Wheeler, I., 2013. Digital mapping of soil carbon. Adv. Agron. 118, 1–47.

Mitášová, H., Hofierka, J., 1993. Interpolation by regularized spline with tension: II. Application to terrain modeling and surface geometry analysis. Math. Geol. 25, 657–669. doi:10.1007/BF00893172

Mizota, C., Van Reewijk, L.P., 1989. Clay Mineralogy and Chemistry of Soils Formed in Volcanic Material in Diverse Climatic Regions, Soil Monograph. ISRIC, Wageningen.
Moebius-Clune, B.N., 2010. Applications of integrative soil quality assessment in research, extention, and eduaction (MS thesis). Cornell University, Ithaca.
Moebius-Clune, B.N., Carlson, M., van Es, H.M., Melkonian, J.J., DeGaetano, A., Joseph, L., 2014. Adapt-N Training Manual, 1.0. ed, Extension Series. Department of Crop and Soil Sciences, Cornell University, Ithaca, NY.
Moebius-Clune, B.N., van Es, H.M., Idowu, O.J., Schindelbeck, R.R., Kimetu, J.M., Ngoze, S., Lehmann, J., Kinyangi, J., 2011. Long-term soil quality degradation along a cultivation chronosequence in Western Kenya. Agric. Ecosyst. Environ. 141, 86–99.
Moebius-Clune, B.N., van Es, H.M., Idowu, O.J., Schindelbeck, R.R., Moebius-Clune, D.J., Wolfe, D.W., Abawi, G.S., Thies, J.E., Gugino, B.K., Lucey, R., 2008. Long-term effects of harvesting maize stover and tillage on soil quality. Soil Sci. Soc. Am. J. 72, 960–969. doi:10.2136/sssaj2007.0248
Moebius-Clune, D., Moebius-Clune, B.N., Schindelbeck, R.R., Thies, J.E., van Es, H.M., 2014. Cornell Soil Health Training Manual 2014 Supplement. Cornell University College of Agriculture and Life Sciences, Ithaca.
Molling, C.C., 2011. Precision Agricultural-Landscape Modeling System Version 5 Combined User's and Developer's Manual. University of Wisconsin Board of Regents, Madison.
Molling, C.C., Strikwerda, J.C., Norman, J.M., Rodgers, C.A., Wayne, R., Morgan, C.L.S., Diak, G.R., Mecikalski, J.R., 2005. Distributed runoff formulation designed for a Precision Agricultural Landscape Modeling System. J. Am. Water Resour. Assoc. 41, 1289–1313. doi:10.1111/j.1752-1688.2005.tb03801.x
Moore, I.D., Grayson, R.B., Ladson, A.R., 1991. Digital terrain modeling - a review of hydrological, geomorphological, and biological applications. Hydrol. Process. 5, 3–30.
Mora-Chinchilla, R., 2000. Geomorfología de la Cuenca del Río Turrialba. Universidad de Costa Rica, San José.
Moral, F.J., Terrón, J.M., Silva, J.R.M. da, 2010. Delineation of management zones using mobile measurements of soil apparent electrical conductivity and multivariate geostatistical techniques. Soil Tillage Res. 106, 335–343. doi:10.1016/j.still.2009.12.002
Morón, A., Cozzolino, D., 2002. Application of near infrared reflectance spectroscopy for the analysis of organic C, total N and pH in soils of Uruguay. J. Infrared Spectrosc. 10, 215–221.
Morris, D.R., Gilbert, R.A., Reicosky, D.C., Gesch, R.W., 2004. Oxidation potentials of soil organic matter in Histosols under different tillage methods. Soil Sci. Soc. Am. J. 68, 817–826. doi:10.2136/sssaj2004.8170
Mouazen, A.M., Kuang, B., De Baerdemaeker, J., Ramon, H., 2010. Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. Geoderma 158, 23–31. doi:10.1016/j.geoderma.2010.03.001
Mueller, T.G., Hartsock, N.J., Stombaugh, T.S., Shearer, S.A., Cornelius, P.L., Barnhisel, R.I., 2003. Soil electrical conductivity map variability in limestone soils overlain by loess. Agron. J. 95, 496–507.
Muschler, R.G., 2001. Shade improves coffee quality in a sub-optimal coffee-zone of Costa Rica. Agrofor. Syst. 85, 131–139.

Nakagawa, S., Schielzeth, H., 2013. A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods Ecol. Evol. 4, 133–142. doi:10.1111/j.2041-210x.2012.00261.x
Nanzyo, M., Dahlgren, R., Shoji, S., 1993. Chemical characteristics of volcanic ash soils, in: Shoji, S., Nanzyo, M., Dahlgren, R. (Eds.), Volcanic Ash Soils - Genesis, Properties and Utilization. Elsevier Science Publishers B.V., Amsterdam, pp. 145–188.
Neild, R.E., Newman, J.E., 1990. Growing Season Characteristics and Requirements in the Corn Belt. Purdue University, Cooperative Extension Service, West Lafayette.
Noponen, M.R.A., Healey, J.R., Soto, G., Haggar, J.P., 2013. Sink or source—the potential of coffee agroforestry systems to sequester atmospheric CO2 into soil organic carbon. Agric. Ecosyst. Environ. 175, 60–68. doi:10.1016/j.agee.2013.04.012
Norman, J.M., 2013. Fifty Years of Study of S-P-A Systems: Past Limitations and a Future Direction. Procedia Environ. Sci. 19, 15–25. doi:10.1016/j.proenv.2013.06.003
Ogden, C.B., vanEs, H.M., Schindelbeck, R.R., 1997. Miniature rain simulator for field measurement of soil infiltration. Soil Sci. Soc. Am. J. 61, 1041–1043.
Oinuma, K., Hayashi, H., 1965. Infrared study of mixed-layer clay minerals. Am. Mineral. 50, 1213.
Oliver, M.A., 2010. An Overview of Geostatistics and Precision Agriculture, in: Oliver, M.A. (Ed.), Geostatistical Applications for Precision Agriculture. Springer Netherlands, Amsterdam, pp. 1–34.
Oliver, Y.M., Robertson, M.J., Wong, M.T.F., 2010. Integrating farmer knowledge, precision agriculture tools, and crop simulation modelling to evaluate management options for poor-performing patches in cropping fields. Eur. J. Agron. 32, 40–50. doi:10.1016/j.eja.2009.05.002
Osborne, T.M., Wheeler, T.R., 2013. Evidence for a climate signal in trends of global crop yield variability over the past 50 years. Environ. Res. Lett. 8, 24001. doi:10.1088/1748-9326/8/2/024001
Page, N.R., 1974. Estimation of organic matter in Atlantic Coastal Plain soils with a color-difference meter. Agron. J. 66, 652. doi:10.2134/agronj1974.00021962006600050014x
Pansu, M., Gautheyrou, J., 2006. Handbook of Soil Analysis Mineralogical, Organic and Inorganic Methods. Springer, Berlin.
Parfitt, R.L., 1990. Allophane in New Zealand - a review. Aust. J. Soil Res. 28, 343–360.
Passioura, J.B., Angus, J.F., 2010. Improving productivity of crops in water-limited environments. Adv. Agron. 106, 37–75.
Payán, F., Jones, D.L., Beer, J., Harmand, J.M., 2009. Soil characteristics below Erythrina poeppigiana in organic and conventional Costa Rican coffee plantations. Agrofor. Syst. 76, 81–93. doi:10.1007/S10457-008-9201-Y
Pebesma, E.J., 2004. Multivariate geostatistics in S: the gstat package. Comput. Geosci. 30, 683–691.
Pedersen, P., 2009. Soybean Growth and Development, PM 1945. Iowa State University, Ames.
Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Koppen-Geiger climate classification. Hydrol. Earth Syst. Sci. 11, 1633–644.
Peigné, J., Vian, J.-F., Cannavacciuolo, M., Lefevre, V., Gautronneau, Y., Boizard, H., 2013. Assessment of soil structure in the transition layer between topsoil and subsoil using the profil cultural method. Soil Tillage Res. 127, 13–25. doi:10.1016/j.still.2012.05.014

Pinheiro-Dick, D., Schwertmann, U., 1996. Microaggregates from Oxisols and Inceptisols: Dispersion through selective dissolutions and physicochemical treatments. Geoderma 74, 49–63. doi:10.1016/S0016-7061(96)00047-X
Plant, R.E., 2001. Site-specific management: The application of information technology to crop production. Comput. Electron. Agric. 30, 9–29. doi:10.1016/S0168-1699(00)00152-6
Post, J.L., Noble, P.N., 1993. The near-infrared combination band frequencies of dioctahedral smectites, micas, and illites. Clays Clay Miner. 41, 639–644.
Potgieter, A.B., Hammer, G.L., Butler, D., 2002. Spatial and temporal patterns in Australian wheat yield and their relationship with ENSO. Aust. J. Agric. Res. 53, 77–89.
Powers, J.S., Schlesinger, W.H., 2002. Relationships among soil carbon distributions and biophysical factors at nested spatial scales in rain forests of northeastern Costa Rica. Geoderma 109, 165–190. doi:10.1016/S0016-7061(02)00147-7
Powlson, D.S., Gregory, P.J., Whalley, W.R., Quinton, J.N., Hopkins, D.W., Whitmore, A.P., Hirsch, P.R., Goulding, K.W.T., 2011. Soil management in relation to sustainable agriculture and ecosystem services. Food Policy 36, Supplement 1, S72–S87. doi:10.1016/j.foodpol.2010.11.025
Pringle, M.J., McBratney, A.B., Whelan, B.M., Taylor, J.A., 2003. A preliminary approach to assessing the opportunity for site-specific crop management in a field, using yield monitor data. Agric. Syst. 76, 273–292. doi:10.1016/S0308-521X(02)00005-7
Pronk, G.J., Heister, K., Kögel-Knabner, I., 2011. Iron oxides as major available interface component in loamy arable topsoils. Soil Sci. Soc. Am. J. 75, 2158. doi:10.2136/sssaj2010.0455
Proulx, S., 2001. Evaluation of the Performance of Soil Moisture Sensors in Laboratory-Scale Lysimeters (MS thesis). University of Manitoba, Winnipeg.
QGIS Development Team, 2015. QGIS Geographic Information System. Open Source Geospatial Foundation Project.
R Core Team, 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.
Ramsey, W.L., 1984. A Comparison of Biological, Physical, and Organic Matter Characteristics of Soil in No-Till and Conventional Till Continuous Corn (Zea mays) (MS thesis). Cornell University, Ithaca.
Razakamanarivo, R.H., Grinand, C., Razafindrakoto, M.A., Bernoux, M., Albrecht, A., 2011. Mapping organic carbon stocks in eucalyptus plantations of the central highlands of Madagascar: A multiple regression approach. Geoderma 162, 335–346. doi:10.1016/j.geoderma.2011.03.006
Reeves, J.B., Van Kessel, J.S., 1999. Investigations into near-infrared analysis as an alternative to traditional procedures in manure N and C mineralization studies. J. Infrared Spectrosc. 7, 195–212.
Richter, J., 1987. The Soil as a Reactor. Catena Verlag, Cremlingen.
Roger-Estrade, J., Richard, G., Caneill, J., Boizard, H., Coquet, Y., Defossez, P., Manichon, H., 2004. Morphological characterisation of soil structure in tilled fields: from a diagnosis method to the modelling of structural changes over time. Soil Tillage Res. 79, 33–49. doi:10.1016/j.still.2004.03.009
Rosolem, C.., Takahashi, M., 1998. Soil compaction and soybean root growth, in: 5th Symposium of the International Society of Root Research. Presented at the Root

Demographics and their Efficiencies in Sustainable Agriculture, Grassland and Forest Ecosystems, Clemson, SC, pp. 295–304.
Rossiter, D.G., 2012. Technical Note: Co-kriging with the gstat package of the R environment for statistical computing [WWW Document]. URL http://www.css.cornell.edu/faculty/dgr2/teach/R/R_ck.pdf (accessed 6.12.15).
Ruffin, C., King, R.L., 1999. The analysis of hyperspectral data using Savitzky-Golay filtering - Theoretical basis (part 1), in: Geoscience and Remote Sensing Symposium. Hamburg.
Rumpel, C., Kögel-Knabner, I., 2010. Deep soil organic matter—a key but poorly understood component of terrestrial C cycle. Plant Soil 338, 143–158. doi:10.1007/s11104-010-0391-5
Russo, D., Bresler, E., 1981. Effect of field variability in soil hydraulic properties on solutions of unsaturated water and salt flows. Soil Sci. Soc. Am. J. 45, 675. doi:10.2136/sssaj1981.03615995004500040001x
Sadras, V.O., Calviño, P.A., 2001. Quantification of grain yield response to soil depth in soybean, maize, sunflower, and wheat. Agron. J. 93, 577. doi:10.2134/agronj2001.933577x
Sankey, J.B., Brown, D.J., Bernard, M.L., Lawrence, R.L., 2008. Comparing local vs. global visible and near-infrared (VisNIR) diffuse reflectance spectroscopy (DRS) calibrations for the prediction of soil clay, organic C and inorganic C. Geoderma 148, 149–158. doi:10.1016/j.geoderma.2008.09.019
SAS Institute Inc., 2015. SAS software. SAS Institute Inc., Cary.
Sawyer, J., Nafziger, E., Randall, G., Bundy, L., Rehm, G., Joern, B., 2006. Concepts and Rationale for Regional Nitrogen Rate Guidelines for Corn, Extension Publication. Iowa State University, Ames.
Schafer, R.L., Young, S.C., Hendrick, J.G., Johnson, C.E., 1984. Control concepts for tillage systems. Soil Tillage Res. 4, 313–320. doi:10.1016/0167-1987(84)90031-X
Scharf, P.C., Kitchen, N.R., Sudduth, K.A., Davis, J.G., 2006. Spatially variable corn yield is a weak predictor of optimal nitrogen rate. Soil Sci. Soc. Am. J. 70, 2154. doi:10.2136/sssaj2005.0244
Scharf, P.C., Shannon, D.K., Palm, H.L., Sudduth, K.A., Drummond, S.T., Kitchen, N.R., Mueller, L.J., Hubbard, V.C., Oliveira, L.F., 2011. Sensor-based nitrogen applications out-performed producer-chosen rates for corn in on-farm demonstrations. Agron. J. 103, 1683. doi:10.2134/agronj2011.0164
Scheinost, A.C., Chavernas, A., Barron, V., Torrent, J., 1998. Use and limitations of second-derivative diffuse reflectance spectroscopy in the visible to near-infrared range to identify and quantify Fe oxide minerals in soils. Clays Clay Miner. 46, 528–536.
Schindelbeck, R.R., van Es, H.M., Abawi, G.S., Wolfe, D.W., Whitlow, T.L., Gugino, B.K., Idowu, O.J., Moebius-Clune, B.N., 2008. Comprehensive assessment of soil quality for landscape and urban management. Landsc. Urban Plan. 88, 73–80. doi:10.1016/j.landurbplan.2008.08.006
Schmidt, J.P., 2015. Nitrogen Fertilizer for Soybean? [WWW Document]. Crop Insights. URL (accessed 11.29.15).
Schnitkey, G., 2015. Expected corn and soybean returns and shifts in acres. Farm Econ. Facts Opin. 15.
Schrumpf, M., Kaiser, K., Schulze, E.-D., 2014. Soil organic carbon and total nitrogen gains in an old growth deciduous forest in Germany. PLoS ONE 9, e89364. doi:10.1371/journal.pone.0089364

Schueller, J.K., 1997. Technology for precision agriculture. Precis. Agric. 97, 33–44.
Schulp, C.J.E., Veldkamp, A., 2008. Long-term landscape – land use interactions as explaining factor for soil organic matter variability in Dutch agricultural landscapes. Geoderma 146, 457–465. doi:10.1016/j.geoderma.2008.06.016
Schwertmann, U., 1991. Solubility and dissolution of iron oxides. Plant Soil 130, 1–25. doi:10.1007/BF00011851
Seibert, J., Stendahl, J., Sørensen, R., 2007. Topographical influences on soil properties in boreal forests. Geoderma 141, 139–148. doi:10.1016/j.geoderma.2007.05.013
Shahandeh, H., Wright, A.L., Hons, F.M., Lascano, R.J., 2005. Spatial and temporal variation of soil nitrogen parameters related to soil texture and corn yield. Agron. J. 97, 772. doi:10.2134/agronj2004.0287
Sharma, G., Sharma, R., Sharma, E., 2009. Impact of stand age on soil C, N and P dynamics in a 40-year chronosequence of alder-cardamom agroforestry stands of the Sikkim Himalaya. Pedobiologia 52, 401–414. doi:10.1016/j.pedobi.2009.01.003
Shepherd, K.D., Walsh, M.G., 2002. Development of reflectance spectral libraries for characterization of soil properties. Soil Sci. Soc. Am. J. 66, 988–998.
Sherman, D.M., Waite, T.D., 1985. Electronic spectra of Fe3+ oxides and oxide hydroxides in the near IR to near UV. Am. Mineral. 70, 8.
Shoji, S., Nanzyo, M., Dahlgren, R., 1993. Productivity and utilization of volcanic ash soils, in: Shoji, S., Nanzyo, M., Dahlgren, R. (Eds.), Volcanic Ash Soils - Genesis, Properties and Utilization. Elsevier Science Publishers B.V., Amsterdam.
Simbahan, G.C., Dobermann, A., Goovaerts, P., Ping, J., Haddix, M.L., 2006. Fine-resolution mapping of soil organic carbon based on multivariate secondary data. Geoderma 132, 471–489. doi:10.1016/j.geoderma.2005.07.001
Simonson, R.W., 1982. Loess in soils of Delaware, Maryland, and northeastern Virginia. Soil Sci. 133, 167–178.
Six, J., Bossuyt, H., Degryze, S., Denef, K., 2004. A history of research on the link between (micro)aggregates, soil biota, and soil organic matter dynamics. Soil Tillage Res. 79, 7–31. doi:10.1016/j.still.2004.03.008
Soil Survey Staff, 2014. Kellogg Soil Survey Laboratory Methods Manual, Version 5.0. ed. USDA-NRCS, Washington D.C.
Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture, n.d. Official Soil Series Descriptions [WWW Document]. URL http://soils.usda.gov/technical/classification/osd/index.html (accessed 8.29.15).
Sojka, R.E., Upchurch, D.R., 1999. Reservations regarding the soil quality concept. Soil Sci. Soc. Am. J. 63, 1039. doi:10.2136/sssaj1999.6351039x
Sojka, R.E., Upchurch, D.R., Borlaug, N.E., 2003. Quality soil management or soil quality management : Performance versus semantics. Adv. Agron. 79, 1–68.
Somarriba, E., Beer, J., Alegre-Orihuela, J., Andrade, H., Cerda, R., DeClerck, F., Detlefsen, G., Escalante, M., Giraldo, L., Ibrahim, M., Krishnamurthy, L., Mena-Mosquera, V., Mora-Degado, J., Orozco, L., Scheelje, M., Campos, J., 2012. Mainstreaming agroforestry in latin america, in: Nair, P.K.R., Garrity, D. (Eds.), Agroforestry - The Future of Global Land Use, Advances in Agroforestry. Springer, Dordrecht, pp. 429–453.
Spectrum Analytic Inc, 2010. Soil Sampling Guide. Spectrum Analytic Inc, Washington Court House.

Spoor, G., Tijink, F.G.., Weisskopf, P., 2003. Subsoil compaction: risk, avoidance, identification and alleviation. Soil Tillage Res. 73, 175–182. doi:10.1016/S0167-1987(03)00109-0
Stadler, A., Rudolph, S., Kupisch, M., Langensiepen, M., van der Kruk, J., Ewert, F., 2015. Quantifying the effects of soil variability on crop growth using apparent soil electrical conductivity measurements. Eur. J. Agron. 64, 8–20. doi:10.1016/j.eja.2014.12.004
Stamps, W.T., Dailey, T.V., Gruenhagen, N.M., Linit, M.J., 2008. Soybean yield and resource conservation field borders. Agric. Ecosyst. Environ. 124, 142–146. doi:10.1016/j.agee.2007.08.004
Stanford, G., 1973. Rationale for optimum nitrogen fertilization in corn production. J. Environ. Qual. 2, 159. doi:10.2134/jeq1973.00472425000200020001x
Stenberg, B., Viscarra Rossel, R.A., 2010. Diffuse reflectance spectroscopy for high-resolution soil sensing, in: Viscarra Rossel, R.A. et al. (Ed.), Proximal Soil Sensing. Springer, Dordrecht.
Stenberg, B., Viscarra Rossel, R.A., Mouazen, A.M., Wetterlind, J., 2010. Visible and near infrared spectroscopy in soil science. Adv. Agron. 107, 163–215.
Stone, J.R., Gilliam, J.W., Cassel, D.K., Daniels, R.B., Nelson, L.A., Kleiss, H.J., 1985. Effect of erosion and landscape position on the productivity of Piedmont soils. Soil Sci. Soc. Am. J. 49, 987. doi:10.2136/sssaj1985.03615995004900040039x
Stoner, E.R., Baumgardner, M.F., 1981. Characteristic variations in reflectance of surface soils. Soil Sci. Soc. Am. J. 45, 1161–1165.
Sudduth, K.A., Drummond, S.T., 2007. Yield Editor: software for removing errors from crop yield maps. Agron. J. 99, 1471. doi:10.2134/agronj2006.0326
Takata, Y., Funakawa, S., Akshalov, K., Ishida, N., Kosaki, T., 2007. Spatial prediction of soil organic matter in northern Kazakhstan based on topographic and vegetation information. Soil Sci. Plant Nutr. 53, 289–299.
Taugourdeau, S., le Maire, G., Avelino, J., Jones, J.R., Ramirez, L.G., Jara Quesada, M., Charbonnier, F., Gómez-Delgado, F., Harmand, J.-M., Rapidel, B., Vaast, P., Roupsard, O., 2014. Leaf area index as an indicator of ecosystem services and management practices: An application for coffee agroforestry. Agric. Ecosyst. Environ. 192, 19–37. doi:10.1016/j.agee.2014.03.042
Terra, J.A., Shaw, J.N., Reeves, D.W., Raper, R.L., van Santen, E., Mask, P.L., 2004. Soil carbon relationships with terrain attributes, electrical conductivity, and a soil survey in a coastal plain landscape. Soil Sci. 169, 819–831.
Thompson, J.A., Kolka, R.K., 2005. Soil carbon storage estimation in a forest watershed using quantitative soil-landscape modeling. Soil Sci. Soc. Am. J. 69, 1086–1093.
Timlin, D.J., Pachepsky, Y., Snyder, V.A., Bryant, R.B., 2001. Water budget approach to quantify corn grain yields under variable rooting depths. Soil Sci. Soc. Am. J. 65, 1219–1226. doi:10.2136/sssaj2001.6541219x
Tisdall, J.M., Oades, J.M., 1982. Organic matter and water-stable aggregates in soils. J. Soil Sci. 33, 141–163. doi:10.1111/j.1365-2389.1982.tb01755.x
Tollenaar, M., Dwyer, L.M., 1999. Physiology of maize, in: Smith, D.L., Hammel, C. (Eds.), Crop Yield: Physiology and Processes. Springer-Verlag, Berlin, pp. 169–204.
Topp, G.C., Galganov, Y.T., Ball, B.C., Carter, M.R., 1993. Soil water desorption curves, in: Carter, M.R. (Ed.), Soil Sampling and Methods of Analysis. Canadian Society of Soil Science, Lewis Publishers, Boca Raton.

Tremblay, N., Bouroubi, Y.M., Bélec, C., Mullen, R.W., Kitchen, N.R., Thomason, W.E., Ebelhar, S., Mengel, D.B., Raun, W.R., Francis, D.D., Vories, E.D., Ortiz-Monasterio, I., 2012. Corn response to nitrogen is influenced by soil texture and weather. Agron. J. 104, 1658. doi:10.2134/agronj2012.0184

Troy, T.J., Kipgen, C., Pal, I., 2015. The impact of climate extremes and irrigation on US crop yields. Environ. Res. Lett. 10, 54013. doi:10.1088/1748-9326/10/5/054013

Tugel, A.J., Herrick, J.E., Brown, J.R., Mausbach, M.J., Puckett, W., Hipple, K., 2005. Soil change, soil survey, and natural resources decision making. Soil Sci. Soc. Am. J. 69, 738. doi:10.2136/sssaj2004.0163

Tyson, T.W., Curtis, L.M., 2008. 60 Acre Pivot Irrigation Cost Analysis, Biosystems Engineering Series Timely Information Agriculture, Natrual Resources & Forestry. Department of Biosystems Engineering, Auburn University, Auburn.

Udelhoven, T., Emmerling, C., Jarmer, T., 2003. Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial-least square regression: a feasibility study. Plant Soil 251, 319–329.

University of Illinois, 2015. farmdoc [WWW Document]. URL http://www.farmdoc.illinois.edu/manage/pricehistory/price_history.html (accessed 11.5.15).

USDA, 2009. 2007 Census of Agriculture. USDA, National Agricultural Statistics Service, Washington D.C.

USDA-ERS, 2015. Commodity Costs and Returns [WWW Document]. URL http://www.ers.usda.gov/data-products/commodity-costs-and-returns.aspx (accessed 11.29.15).

USDA-ERS, 2000. Farm Resource Regions, Agricultural Information Bullentin. USDA-ERS, Washington D.C.

USDA-NRCS, 2015. SSURGO Soil Map Coverage versus the U.S. General Soil Map Coverage [WWW Document]. USDA-NRCS. URL http://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/geo/?cid=nrcs142p2_053626#data (accessed 2.22.15).

USDA-NRCS, 2005. Global Soil Regions. USDA-NRCS, Soil Survey Division, World Soil Resources, Washington D.C.

Van den Akker, J.J.., Arvidsson, J., Horn, R., 2003. Introduction to the special issue on experiences with the impact and prevention of subsoil compaction in the European Union. Soil Tillage Res. 73, 1–8. doi:10.1016/S0167-1987(03)00094-1

Van Eerd, L.L., Congreves, K.A., Hayes, A., Verhallen, A., Hooker, D.C., 2014. Long-term tillage and crop rotation effects on soil quality, organic carbon, and total nitrogen. Can. J. Soil Sci. 94, 303–315. doi:10.4141/cjss2013-093

van Ittersum, M.K., Cassman, K.G., Grassini, P., Wolf, J., Tittonell, P., Hochman, Z., 2013. Yield gap analysis with local to global relevance—A review. Field Crops Res. 143, 4–17. doi:10.1016/j.fcr.2012.09.009

van Ittersum, M.K., Rabbinge, R., 1997. Concepts in production ecology for analysis and quantification of agricultural input-output combinations. Field Crops Res. 52, 197–208. doi:10.1016/S0378-4290(97)00037-3

Van Uffelen, C.G.R., Verhagen, J., Bouma, J., 1997. Comparison of simulated crop yield patterns for site-specific management. Agric. Syst. 54, 207–222.

Vanotti, M.B., Bundy, L.G., 1994. Corn nitrogen recommendations based on yield response data. Jounral Prod. Agric. 7, 249–256.
Varmuza, K., Filzmoser, P., 2009. Introduction to Multivariate Statistical Analysis in Chemometrics. CRC Press, Boca Raton.
Veris Technologies, 2012. Operating Instructions MSP3, Pub.#OM18-MSP3. Veris Technologies Inc., Salina.
Viscarra Rossel, R.A., Adamchuk, V.I., Sudduth, K.A., McKenzie, N.J., Lobsey, C., 2011. Proximal soil sensing: an effective approach for soil measurements in space and time. Adv. Agron. 113, 243–291.
Viscarra Rossel, R.A., Cattle, S.R., Ortega, A., Fouad, Y., 2009. In situ measurements of soil colour, mineral composition and clay content by vis–NIR spectroscopy. Geoderma 150, 253–266. doi:10.1016/j.geoderma.2009.01.025
Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131, 59–75.
Viscarra Rossel, R.A., Webster, R., 2011. Discrimination of Australian soil horizons and classes from their visible–near infrared spectra. Eur. J. Soil Sci. 62, 637–647. doi:10.1111/j.1365-2389.2011.01356.x
Vonesh, E.F., Chinchilli, V.M., Pu, K., 1996. Goodness-of-fit in generalized nonlinear mixed-effects models. Biometrics 52, 572–587. doi:10.2307/2532896
Wackernagel, H., 2003. Multivariate Geostatistics: An Introduction with Applications, 3rd ed. Springer, Berlin.
Walker, J.M., 2002. The bicinchonic acid (BCA) assay for protein quantitation, in: Walker, J.M. (Ed.), The Protein Protocols Handbook. Humana Press, Totowa.
Walvoort, D.J.J., Brus, D.J., de Gruijter, J.J., 2010. An R package for spatial coverage sampling and random sampling from compact geographical strata by k-means. Comput. Geosci. 36, 1261–1267. doi:10.1016/j.cageo.2010.04.005
Ware, E.C., 2005. Corrections to Radar-Estimated Precipitation Using Observed Rain Gauge Data (MS thesis). Cornell University, Ithaca.
Warrick, A.W., Myers, D.E., Nielsen, D.R., 1986. Geostatistical Methods Applied to Soil Science, in: Methods of Soil Analysis, Part 1. Physical and Mineralogical Methods, Agronomy Monograph. American Society of Agronomy : Soil Science Society of America, Madison, pp. 53–82.
Weaver, K.N., 1967. Generalized Geologic Map of Maryland. Maryland Geological Survey, Baltimore.
Webster, R., 1997. Soil resources and their assessment. Philos. Trans. R. Soc. Lond. B Biol. Sci. 352, 963–973. doi:10.1098/rstb.1997.0075
Webster, R., Oliver, M.A., 2007. Geostatistics for Environmental Scientists, 2nd ed. Wiley, Chichester.
Weil, R.R., Islam, K.R., Stine, M.A., Gruver, J.B., Samson-Liebig, S.E., 2003. Estimating active carbon for soil quality assessment: a simplified method for laboratory and field use. Am. J. Altern. Agric. 18, 3–17. doi:10.1079/AJAA200228
Weiss, A.D., 2001. Topographic position and landform analysis.
Wetterlind, J., Stenberg, B., 2010. Near-infrared spectroscopy for within-field soil characterization: small local calibrations compared with national libraries spiked with local samples. Eur. J. Soil Sci. 61, 823–843. doi:10.1111/j.1365-2389.2010.01283.x

Wetterlind, J., Stenberg, B., Söderström, M., 2010. Increased sample point density in farm soil mapping by local calibration of visible and near infrared prediction models. Geoderma 156, 152–160. doi:10.1016/j.geoderma.2010.02.012

Whelan, B.M., McBratney, A.B., 2003. Definition and interpretation of potential management zones in Australia, in: Solutions for a Better Environment. Presented at the 11th Australian Agronomy Conference, Australian Society of Agronomy, Geelong, Victoria.

Wilks, D.S., 2008. High-resolution spatial interpolation of weather generator parameters using local weighted regressions. Agric. For. Meteorol. 148, 111–120. doi:10.1016/j.agrformet.2007.09.005

Williams, J.D., Long, D.S., Wuest, S.B., 2011. Capture of plateau runoff by global positioning system–guided seed drill operation. J. Soil Water Conserv. 66, 355–361. doi:10.2489/jswc.66.6.355

Williams, P.C., Sobering, D.C., 1993. Comparison of commercial near infra-red transmittance and reflectance instruments for analysis of whole grains and seeds. J. Infrared Spectrosc. 1, 25–33.

Wright, S.F., Upadhyaya, A., 1996. Extraction of an abundant and unusual protein from soil and comparison with hyphal protein of arbuscular mycorrhizal fungi. Soil Sci. 161, 575–586.

Yang, H.S., Dobermann, A., Cassman, K.G., Walters, D.T., Grassini, P., 2013. Hybrid-Maize (ver.2013.4). A Simulation Model for Corn Growth and Yield. Nebraska Cooperative Extension, University of Nebraska-Lincoln, Lincoln.

Yang, J., 2012. Interpreting Coefficients in Regression with Log-Transformed Variables, StatNews. Cornell University, Cornell Statistical Consulting Unit, Ithaca.

Yang, R.-C., Crossa, J., Cornelius, P.L., Burgueño, J., 2009. Biplot analysis of genotype × environment interaction: Proceed with caution. Crop Sci. 49, 1564. doi:10.2135/cropsci2008.11.0665

Zhang, J., Barber, S.A., 1992. Maize root distribution between phosphorus-fertilized and unfertilized soil. Soil Sci Soc Am J 56, 819–822. doi:10.2136/sssaj1992.03615995005600030024x

Zhu, A.X., Hudson, B., Burt, J., Lubich, K., Simonson, D., 2001. Soil mapping using GIS, expert knowledge, and fuzzy logic. Soil Sci. Soc. Am. J. 65, 1463. doi:10.2136/sssaj2001.6551463x