Centre for Geo-Information

Thesis Report GIRS-2014-13

Key soil property identification and delineation of management zones in precision agriculture

Tewelde Gebremichael Hagos

April, 2014

WAGENINGEN UNIVERSITY
WAGENINGEN UR

Key soil property identification and delineation of management zones in precision agriculture

Author

Tewelde Gebremichael Hagos

Registration number: 860727298080

Supervisors:

Dr. Lammert Kooistra

Dr. Harm Bartholomeus

A thesis submitted in partial fulfilment of the degree of Master of Science

at Wageningen University and Research Centre,

The Netherlands.

April, 2014

Wageningen, The Netherlands

Thesis code number: GRS-80436

Thesis Report: GIRS-2014-13

Wageningen University and Research Centre

Laboratory of Geo-Information Science and Remote Sensing

Abstract

Two adjacent precision farming fields located in south of the Netherlands were examined to delineate homogenous management zones for site-specific management. Spatially sensing like electrical conductivity (ECa), pH value, colour aerial photograph and digital elevation model (DEM) data sources were being obtained for these fields due to the advancements of proximal and remote sensing technologies. However, little attention was given to the over-information due to such a large volume of data from the technological progresses. The objective of the study was to develop and test the method of identifying key soil parameters to delineate management zones needed for precision farm management. Three scenario were considered for this study: i) combined analysis of both fields, ii) separate analysis of each field and iii) combined analysis of both fields but using national available geo-data sources (DEM and aerial photograph). Principal component analysis was used to identify soil variables which explain most of the soil variation for each scenario. Three principal components were retained for both scenario I and II while two principal components for scenario III. The spatial coherence and spatial distribution maps of the identified ECa's, optical soil indices and elevation soil parameters were analysed using geo-statistical techniques. Unsupervised k-mean clustering algorithm was then performed to delineate potential management zones using the identified soil parameter for each scenario. Three optimal management zones per scenario were found most convenient based on the separable and overlapping nature of the classes. To assess the goodness of the defined management zones for each scenario, geo-referenced potato yield were examined and compared using ANOVA. In addition geo-located organic matter samples on field one were used for validation purpose. Significant mean differences of potato yield among and between the management zones for scenario I and II were found. Significant mean differences of organic matter were also found. In general the method can be operable for precision agriculture at field level using apparent electrical conductivity, colour aerial photographs and elevation sensing data sources.

Key words: Precision farming, k-mean clustering, management zones, principal component analysis.

Acknowledgement

First and foremost, I would like to thank Almighty God who made me healthy and let me finish the course.

I am grateful to my supervisors Dr. Lammert Kooistra and Dr. Harm Bartholomeus from the Laboratory of Geo-information Science and Remote Sensing of Wageningen University in the Netherlands for their guidance, constructive comments and encouragements during the whole MSc research period and of course throughout my stay in MGI programme. Without their support completion of the research would have not been possible.

I would like also extend my appreciation to Mr. Jacob van den Borne and his colleagues for providing me with complete data sources needed for the study. I could not finish the study without the use of van den Bornes's data.

Lastly, but not least I want to acknowledge my study advisor Willy ten Haaf, Dr. Sytze de Bruin, Dr. Jan Verbesselt and all Geo-information Science and remote sensing staff members of Wageningen University for their great contribution in developing my knowledge and skills for future career and education.

Table of content

List of tables

List of figures

Chapter 1. Introduction

This chapter explores the background of precision agriculture, the statement of the problem, the research objective and research questions and describes the organization of the thesis.

1.1. Background

In contrast to traditional farming, today's site-specific farming system is capable of producing high quality and amount for the ever growing world population (Bongiovanni and Lowenberg-Deboer, 2004; McBratney et al., 2005; Tilman et al., 2002). The general aim of site-specific farming management is to increase profitability of crop production and to reduce unwanted environmental impacts (Gebbers and Adamchuk, 2010; Godwin et al., 2003).

Adamchuk et al. (2011a) defined precision agriculture as the management strategy based on information technologies implemented to optimize agriculture production. Corwin and Lesch (2005); and Viscarra Rossel and McBratney (1998) also extensively defined precision agriculture as the application of information and communication technologies to within-field data gathering and management driven systems which provides spatial and temporal information on how, where and when to apply inputs to the agricultural farming system.

Progress in farm management concepts and precision agriculture; have begun to exploit within-field heterogeneity so that variable input rate applications, more precise management and yield monitoring could be applied. It is based on observing and responding to intra- and inter-field variations. Precision agriculture relies on new technologies like sensing technology, information technology and geospatial tools with due consideration of farmers ability to locate precise Ground position system (GPS) locations and management experiences for best management decisions (Auernhammer, 2001; Cox, 2002; Plant, 2001). These practices in return maximize profitability, improve sustainability and reduce input demands (Bongiovanni and Lowenberg-Deboer, 2004; Gebbers and Adamchuk, 2010; Whelan and McBratney, 2000).

Besides recently evolving management practices, precision agriculture needs detailed soil information in terms of mapping and characterization of soil variations to support decision making and assist farmers for site-specific management. Precision agriculture is all about applying the correct input in the correct amount at a correct place in needed time (Fleming et al., 2000). There exist soil physical, chemical and biological variability and variation within short distances both horizontally and vertically which comes from complex processes and interactions that takes place in the soil environment (Stenberg et al., 2010). Mzuku et al. (2005) study showed that soil physical properties exhibited significant spatial variability across management zones of production fields. In a related study of (Gaston et al., 2001), location and density of weeds was influenced by spatial variability of soil clay and soil organic matter (SOM).

Recent innovations in sensing technology, water management, proximal sensing, on-the-go soil, and yield monitoring and crop management complemented the spatially sparse soil data to investigate the field variations. More specifically proximal soil sensing (PSS) techniques which is used for the investigation purpose is defined as field based techniques that can be used to measure soil chemical, physical, biological and mineralogical properties from a distance of approximately less than 2 m above the soil surface (Rossel et al., 2011). With the availability those innovations, characterization of unavoidable spatial and temporal soil and crop inter- and intra-fields variations is examined by (Boydell and McBratney, 2002; Reyniers et al., 2006).

Nowadays geo-referenced spatially dense and coverage of extensive information sources are becoming available. Given the fact that soil and crop properties measurements can be acquired at high spatial and temporal resolution, data availability in respect is growing quickly (Vitharana et al., 2008). The adoption and implementation of such precision agriculture innovations are also advancing in developing countries (Batte and Arnholt, 2003; Maohua, 2001; McBratney et al., 2005) even though, proper decision-support systems of implementing, insufficient recognition of temporal variation and lack of integrated farm management remains major stumbling obstacles for adoption (McBratney et al., 2005; Tilman et al., 2002).

1.2. Statement of the problem

Precision agriculture practice is becoming evident especially due to the increment of available datasets from development of different proximal sensors. Soil properties like apparent electrical conductivity (ECa), soil pH, digital elevation model (DEM) and their derivative information used for soil characterization and management zone delineation apparently are becoming available through high spatial and temporal resolution mobile and on-the-go sensors (Thessler et al., 2011).

Consequently, large volume of soil and crop data is available starting recently in precision agriculture farming due to the advancement of remote sensing and proximal soil sensing technologies (Gibbons, 2000; Zhang et al., 2002b). However, little attention was given to the over-information due to such a large volume of data from the technological advancements. It is believed that the availability of ancillary information or information from different sensors with different accuracy might contribute to characterization of variability or inter-correlation of soil and crops at high spatial and temporal resolution. But this might also result or end-up in a sort of data redundancy. Consequently, selection of most important parameters which explain most of the existing variations remains essential component of precision farming. In line with this (Van Meirvenne et al., 2013) selected key parameters based from a list of soil properties, topography attribute, EMI measurement, and gamma ray measurements data sources for site-specific management. A previous study of (Vitharana et al., 2008) also conducted a related study based on ECa and top and sub-soils data sources.

Further development and testing of the methods is then critical in this regard as: proximal sensing technology, on-the-go soil and crop sensing and measurement, water management and focuses on management zones focuses are evolving through time. Improvement and testing of the method on broader scope of spatial coverage of fields, crops production system, and data sources is not intensively investigated yet to confirm broader implementation and application to define better performing management zones. Besides efficient use of inputs, sustainability of environmental is becoming centre of concerns (Gibbons, 2000). This gap motivates the researcher to develop and test the methodology introduced by (Van Meirvenne et al., 2013).

1.3. Research Objective and question

The main objective of the research is to develop and test the methodology of identify key sensing based soil variables to delineate management zones for a precision agriculture farm. To achieve this objective the following research questions have been developed.

RQ1. Which sensing based soil variables describes most of soil variations within fields?

RQ2. How to define management zones in precision agricultural farm using soil parameters?

RQ3. How to evaluate the potential performance of the method used to define management classes?

1.4. Organization of the thesis

The report includes seven chapters. The first chapter introduced background and problem statement of precision farming, proximal sensing and management zone delineation. It also presented the objectives and research questions of the study.

The second chapter presents literature review regarding field variations, sensing based technology, used sensors and definitions of management zones.

The third chapter introduced the study area, descriptions of input dataset and the methods used. This chapter also presented the description of validation methods.

The fourth chapter presented the results section. Conventional statistics of input data, principal component analysis, interpolation of selected soil variables, management zone delineation, management zone and crop yield map visual assessment, and management zone delineation are presented in this chapter.

The fifth chapter discusses the result of the paper. Own results and other finding results are discussed. Conclusion sections are included in the sixth chapter. Finally recommendation remarks are included in chapter seven.

Chapter 2. Literature review

This section covers literature reviews on field variation and sensing technology, definition and concepts of management zones, and clustering algorithms.

2.1. Field variation and sensing technology

Spatial heterogeneity in soil properties is part of an agricultural field due to the soil forming factors and soil forming process. Such a variability of chemical and physical soil properties obviously happens in agricultural farming and is even unavoidable in small fields of precision agricultural farms (McBratney and Pringle, 1999). Acquiring high temporal and spatial resolution data, in a real-time data with high quality sensors to detects within-field variations increases agricultural production (Pierce and Elliott, 2008). Due to this, the knowledge of spatial variability of soil properties within an agricultural field is a critical issue for successful site-specific crop management, especially in precision farming (Thessler et al., 2011).

Obtaining and analysing soil properties are important bottlenecks in the traditional farming system. Traditional management practices assume similar soil type and management practice over the entire field. However, intra-and inter- field spatial variability of soil properties and yield is unavoidable. Inherently variable water holding capacity, resistance to root growth, acidity, nutrient deficiency, texture, depth and other related soil properties are some of the causes for yield variations within a field. So homogenous management practices for such a field with ends-up with less eco-friendly and an-uneconomical practices (Patabendige et al., 2003).

The basic principle of precision agriculture lies in: sensing, recognizing, managing field variability and implement decision accordingly (Pierce et al., 1999). Finding sensed data which addresses within-field variability is a critical step in precision agriculture, as proper decision management is impossible without understanding such variability. Precision farming is based on observing and responding to intra-field variations. Advancement in recent remote sensing, VIS-NIR reflectance spectrum and proximal soil sensing methods has made it possible to rapidly acquire soil data and crops data (Figure 2). Large quantities of inexpensive high spatial and temporal resolution soil data and even in real time data can be acquired nowadays by proximal sensors as frequent as every second (Chang and Laird, 2002; Moral et al., 2010; Rossel et al., 2011; Viscarra Rossel and McBratney, 1998), besides identifying variation of soil properties due increasing of resolutions for crop and soil measurement (Figure 2).

Today quite a numbers of advanced soil sensors have been investigated to manage field variability. In a broad scope field variability can be achieved by either map-based or sensor-based approaches. Sensor-based approaches are relatively expensive and less available. This sensor-based approach describes soil properties and crop parameters using real time sensors in an on-the-go fashion to control variable rate application (Zhang et al., 2002a). In contrast map-based approach is implemented with the availability of GPS technologies, remote sensing, yield monitoring and soil sampling which help to produce site-maps for field management.

Focusing on proximal soil sensing, it can be performed when a sensor detects signal from the soil being in contact with or close to the soil (Rossel et al., 2011). Kuang et al. (2012) categorised proximal sensors into five main categories:

I. Reflectance based soil sensors
II. Conductivity, resistivity and permittivity based soil sensors
III. Radiation based soil sensors

IV. Strength based soil sensors
V. Electro-chemical soil sensors

Proximal sensors can also be furthered grouped in to different groups based on: the way they measure, the energy sources, the way they operate and method of inferring with soil (Figure 1). The mobile on-the-go operating sensors are becoming an interdisciplinary field of research (Adamchuk et al., 2011b).

Although different sensors out of the listed above are implemented in the mechanized precision farming of the study field, for this study EMI (EM38-MK2) sensor, Veris pH Manager sensing and unmanned aerial vehicle techniques are used.

EM38-MK2

The ability of soil to conduct electricity is usually quantified by electrical resistivity or electrical conductivity. The EM38-MK2 Ground Conductivity Meter is contact less and invasive methods of measuring soil electrical conductivity. ECa first was introduced for salinity evaluation (Rhoades, 1993), which nowadays is greatly used for field-variability characterization to define management zones for efficient and sustainable utilization of farm inputs (Fleming et al., 2000; Moral et al., 2010; Vitharana et al., 2008). In contrast to aerial and satellite remote sensing, on-the-go sensors that operate beneath the soil surface minimize weather and field disturbances (Frazier et al., 1997).

The EM38-MK2 sensor which measures the soil ECa consists of a transmitter and receiver coils on both ends of horizontal bar at a distance of 1 meter. The transmitter coil at or above the ground surface is energised with an alternating current creating a primary magnetic field in the soil. This magnetic field causes a current to flow in the soil and that inspires a second magnetic field on that with the receiver is measured. The technical details of the instrument are described by Davis et al. (1997) and McNeill (1980).

Apparently soil ECa has become one of the most frequently used measurements to characterize within field variability of precision farm as it is a function of a multiple soil properties (Al-Gaadi, 2012; Corwin and Lesch, 2003; Sudduth et al., 2001). Different authors investigated the use of ECa to characterize several soil properties with varying degree of accuracy and some of them are presented (Table 1). More details about the sensing techniques and accuracy assessed can be found at (Khakural et al., 1998; Sheets and Hendrickx, 1995; Sudduth et al., 2001; Sudduth et al., 2003; Sudduth et al., 2005).

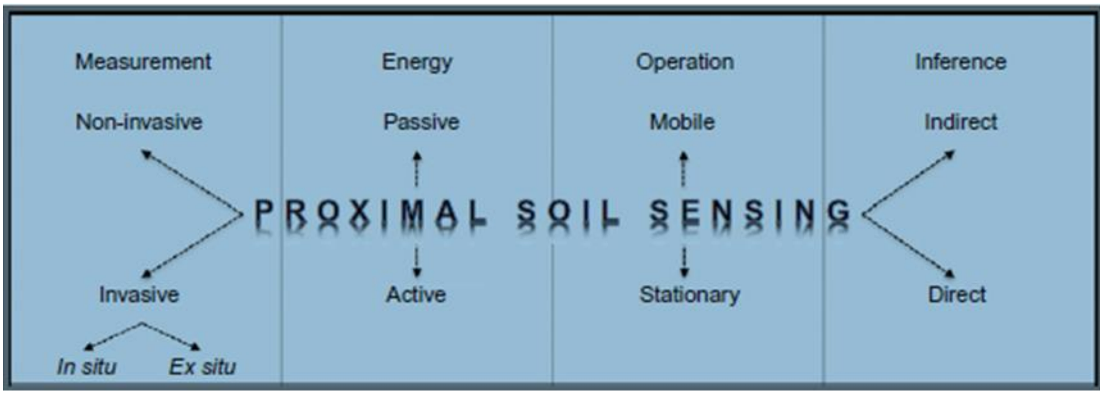


Figure 1. Different categories of proximal soil sensors based on how they operate taken from (Rossel et al., 2011).

Table 1. Soil properties assessed by EMI sensing techniques and with respective accuracy

Soil property	Plat form	R^2	Reference
soil compaction	On-line	0.66	(Al-Gaadi, 2012)
clay	On-line	0.693	(Williams and Hoey, 1987)
Soil Moisture content	Online	0.67- 75	(Christy, 2008; Mouazen et al., 2005)
salinity	On-line		(Farahani and Buchleiter, 2004)
Chemical properties (Ca^{++}, Mg^{++}, K^{+})	On-line	0.26 – 0.85	(Christy, 2008; Shaner et al., 2008)
SOM	On-line	0.44 – 0.83	(Christy, 2008; Muñoz and Kravchenko, 2011; Shonk, 1991)
Total nitrogen	On-line	0.86	(Christy, 2008; Kuang and Mouazen, 2013)
Soil pH	On-line	0.43	(Christy, 2008)

Veris pH

Veris pH manager which take soil samples within grid samples was used. The sensor is commercially built on mobile sensor platform for soil pH measurement (Veris Technologies). It collects geo-referenced soil samples so that soil pH map can be produced. Regarding the accuracy performance of the instrument: soil pH might vary from 5.4 to 8.0 over distances of about 150 m in most transects. In some sections soil pH varied about 2 pH units over a 12 m distance (Bianchini and Mallarino, 2002). On the other hand (Brouder et al., 2005) concluded soil data samples from smaller grids provide much information on the natural distribution pH or lime requirement. Lauzon et al. (2005) added a grid spacing of 30 meter or less is required to adequately asses the spatial distribution of soil pH, phosphorous and potassium properties.

Imaging cameras

Aerial and satellite imagery is a promising approach especially in contactless and inaccessible fields (Mulder et al., 2011). It is also an excellent means of analysing the landscape variability of soil based on reflectance (Frazier et al., 1997; Lamb and Brown, 2001). However, vegetation, weather, crop residue and other factors limits the proper functionality of this technique. Moreover colour aerial photographs nowadays are available or can be acquired at low cost to estimate spatial distribution of soil properties. Unmanned airborne vehicles (UAVs) are used for this purpose. A UAV remote sensing technique provides or has potential to provide high spatial and temporal resolutions, highly flexible image acquisition program though payload of the platform is the limiting factor. Different indices can be derived from the colour aerial photographs to assess spatial distribution of some soil parameters like organic matter and moisture content (Bartholomeus and Kooistra, 2012).

2.2. Definition and concepts of management zones

Designing management zones contributes a lot for better management of farm inputs, crop management, and environmental impacts. These activities are the core practices in precision agriculture and in variable input rate application. Though defined indifferent ways based on scale and goal; the most comprehensive one “management zones are sub-regions of a field that express a homogeneous combination of yield limiting factors for which a single crop input is appropriate to attain maximum efficiency of farm inputs” (Doerge, 1999). Consequently, a management zone within a field may be crop, nutrient and parameter specific.

Delineation of management zones for a specific input or parameter considers factors directly influencing the status and availability of that particular input or parameter (Zhang et al., 2002b). This implies that management zone with in a field may be different for different inputs used, nitrogen versus phosphorus fertilizer for example. So while delineating nitrogen management zone, factor mainly influencing the status of nitrogen situations are considered accordingly. Fleming et al. (2000) extends the definition of management zone to individual nutrient maps for variable rate input application or on-go fertilizer treatment. Referring to management of field variation, management zones can be defined as geographical areas that can be treated as homogenous so that input application and decision making can be treated separately for each zone.

Referring back to historical background of management zones, farmers experiences have been playing an important role (Crookston, 1996). Farmers might have qualitative information on their field variation based on their past production history. Conventional farmers' management practices are based on expertise of farmers and long-term knowledge of the respective fields in contrast to the recent developed data-driven management approach. Due to further advancement of GPS, geographic information systems (GIS), remote sensing technology, real-time yield and soil proximal sensors, delineation of management zone nowadays is data-driven approaches.

The concept of management zone in precision farming is the most important concept for variable input application rate and efficient management of fields. Depending on: natural variability within a field, size of the field (spatial coverage) and management factors, the numbers of appropriate management zones could vary for a certain field (Zhang et al., 2002b). The number of appropriate management zones within a field is influenced by tehnological equipiment used and practical applicability patterns of classes. So it is common practice to modify or remove excessive detailes among the mamagement zones as it prohobits practicability of operations in a real situation. On the other hand it is believed that considering more limiting variables can more advantageous to define better performing management

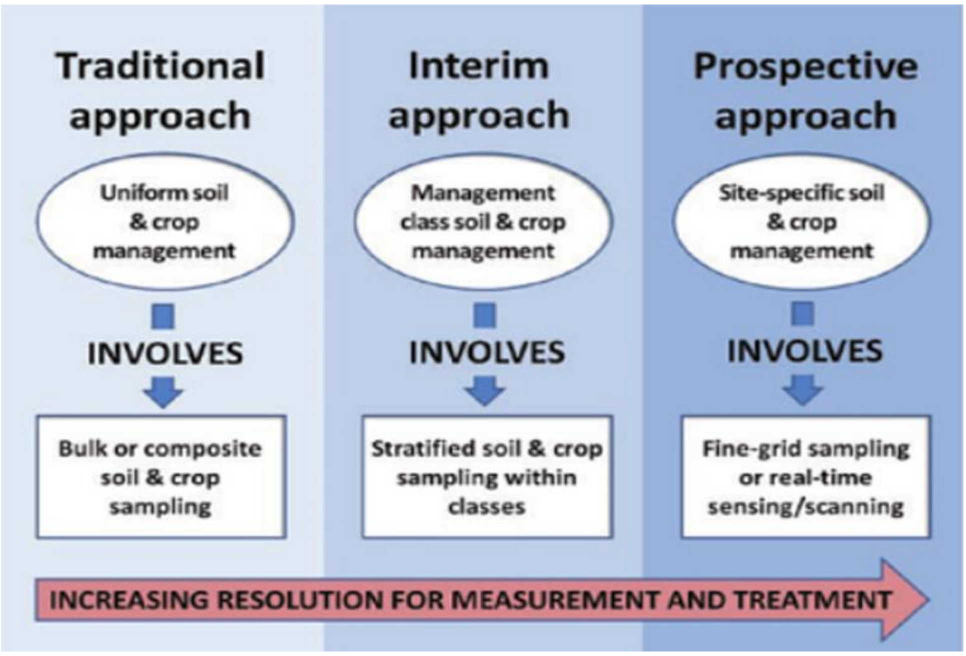


Figure 2. The evolution of site-specific crop management from a uniform to a totally site-specific approach (Whelan and Taylor, 2013)

zones. Density and spatial scale of data, cost of the data, relationship with the crop, stability and repeatable characterstics of data needs to be considered during management zoning strategey (Doerge, 2009).

The concept of management zone in precision farming again is critical for variable input application rate and efficient management of fields. Different authors indicated that management zone delineation approach combines different techniques to visualize, identify and locate spatial distributions within fields for input rate (Doerge, 1999). Soil properties, aerial photographs, topography factors and yield maps have all been suggested as logical basis to define homogenous zones in agricultural fields (Fleming et al., 2000; Schepers et al., 2004; Schepers et al., 2000).

Soil sensors are becoming more fast, accurate, wireless and efficient to provides on time and cost efficient quantitative results with time (Rossel et al., 2011). Remote and proximal soil sensing principles on broader scope are expected to fulfil the spatial and temporal resolution demands of soil characterization. Airplane or satellite remote sensing is becoming a promising approach to characterize soil field variations. However remote sensing suffers from spatial and temporal resolution inadequacy short comings (McBratney et al., 2003). Moreover, remote sensing is preferred under low atmospheric disturbance and less cloud effects.

Currently, there is no site-specific management prescription method and algorithm proven to be the most favourable for all variables involved in crop production. As a result authors developed different methods to define management zones in precision agriculture (Li et al., 2007; Van Meirvenne et al., 2013). Depending on; data sources, growers' expertise, soil and crop characteristics, computer literacy, location of field and methods or technique used, different delineation approaches are forwarded. Many researchers used one or multiple information sources and methods to delineate sub-field management zones. Following paragraphs presents some of the approaches used to delineate potential management zones implement by different studies depending on the determining factors listed in the above lines. They are presented in the following paragraphs.

Franzen and Nanna (2003) presented basic data anlysis approach for delineation of management zones in Des Moines, IA, United States. They used topography, yield, soil survery aerial photographs, satellite imagery and ECa attributes data sources for analysis. Results of individual and combination of these variables and their correlation coefficient to nitrate were compared. Subset of variables consisting topogarphy, satellite imagery and yield mapping results highest and most consisitent correlation while use of combination of all the variables together resulted worse the result. Three years later in 2006, subsequent study by smae authors confirmed the same result. (Schepers et al., 2004) conducted a study with a hypothesis of spatial and temporal variations of soil properties affects the spatial variability of yield. The authors used aerial-photos, elevation, yield, soil samples and apparent electrical conductivity (ECa) and soil attributes on irrigated field. Principal component analysis (PCA) used to aggregate the landscape attributes and finally four management zones were identified.

In another study (Jaynes et al., 2003) used multi-year sampled data to determine management zones on Iowa corn field, USA. Three step processing of partitioning, interpretation and profiling is followed by *K-mean cluster* analysis. The authors used cluster analysis to interpret temporal and spatial patterns in six years of corn yield measurements. Once the clusters are obtained, a conical multiple discriminate analysis is performed to reveal which field attribute contributed significantly towards classifiyning yield in to clusters. (Brock et al., 2005) also proposed unsupervised clustering with special focus on usage of *fuzzy c-mean* to develop management zones from geo-referenced multi-year yield data.

Management zones are compared for diffent crops on the same field to evaluate the effectivenss of the study strategy.

Management zone analyst software was developed for management zone delination though (Kitchen et al., 2005). The authors used ECa and elevation multiyear data sources to define management zones which later was compared with management zones defined using yiled map. King et al. (2005) also investigated the relationship between ECa and yield map for delination of management zones and fuzzy clustering was uesd to define the management zones. Finally they conclude that the combination use of yield and ECa data defined more approprite classes than individual use.

Song et al. (2009) introduced an integrated approach of soil and yield with remotely sensed (Quickbird) image in northern China. A fuzzy k-mean algorithm was used for delineation purpose. Long et al. (1994) also investigated the uses of aerial photographs remotely sensed images of growing crops to accurately delineate management units which later used to predict grain yield.

Ortega and Santibáñez (2007) compared and evaluated three zoning methods using multiple soil fertilities data sources for consecutive years. Xin-Zhong et al. (2009) also presented an approach to delineate management zone using soil properties sampled from tobacco field in central China. The authors assessed the spatial variability and spatial distribution and maps were constructed using geo-techniques. Principal component analysis and clustering algorithms were then performed to define homogenous management zones.

Li et al. (2007) presented an approach which integrates soil and landscape properties with remote sensing images to delinate mangement zones. Normalized fifference vegetation index (NDVI), EC, organic matter, total Nitrogen, recent yield and cation exchange capacity data sources are used for implementation. They performed principal component analysis and resulted two principal components explaining 87.7 % of the total variability. Fuzzy c-mean cluster was allowed for clustering and finally resulted three management zones. To assess the the accuracy of the defined management zones, 139 georeferenced soil and crop yield sample points accross each management zone was analyzed using statistical variance. Significant statistical differences between the chemical properties of soil samples and yield data in each defined management zone was found.

More recently, Vitharana et al. (2008) investigated an approach for selection of key soil parameters due to large dataset coming out of advancement on-go-soil, terrain modelling and yield mapping sensing. Top and sub-soil properties, EC and topographic attributes are collected from an agriculture field in Belgium. 110 top and sub-soil point samples at two depths, EMI (ECa), topographic attribiutes and three year consecutive yield data sources were used. Principal component analysis retained three principal components explaining 67 % of the total variations. A fuzzy k-mean classification is also applied to define four potential management class. Four years later (Van Meirvenne et al., 2013) tested and developed the method. Fortunatily the same key parameters as the previous study were selected. The authors also implementd stepwise multivariate regression analysis to evaluate the performance power of the key variables. This helps identiy how helpful the selected variableas were for specific crop on the field.

Cluster analysis is the search for clusters in the data, in such a way that classes belonging to the same cluster resemble each other, whereas classes in different clusters are dissimilar (Ortega and Santibáñez, 2007). Fields with similar soil properties, landscape and crop properties are divided in to potential management zones, beside quantifying variability pattern and reducing the empirical nature of defined management zones. Boydell and McBratney (2002); Fraisse et al. (2001); Stafford et al.

(1999) implemented fuzzy clustering on yield monitor data to define spatial potential management zones. On the other hand (Jaynes et al., 2003) implemented cluster analysis to interpret spatial and temporal patterns of multi-year corn yield.

The split of the field to homogenous subfields and consecutive merge can be performed using different clustering algorithms depending on the measure of similarity used and in their weighting criterion as reviewed by different authors in connection with management zones. The clustering algorithm methods could be classified as: fuzzy k-mean, fuzzy c-mean and hierarchical clustering. The simplest unsupervised *K-means* or *fuzzy K-mean* to solve clustering problems is most commonly used clustering method (Ortega et al., 2002). Hierarchical clustering which start with each point in a single cluster and subsequently merge clusters according to some criterion or constraint is also a common clustering approach.

Chapter 3. Materials and Methods

3.1. Study area and datasets

The study is conducted on two adjacent potato fields of about an area 22 hectare ($51^0$18'54" North and $5^0$10'12" East) at van den Borne Aardappelen precision farm in south of the Netherlands (Figure 3). ECa, pH value, colour aerial photographs and digital elevation model (DEM) data sources from both fields were used for this study. Three case scenarios were considered for this study depending on data usage and spatial scale or coverage.

i. Combined analysis of both adjacent fields
ii. Separate analysis of the adjacent fields (field one and field two, Figure 3)
iii. Combined analysis of the adjacent fields but, using easily available public elevation and RGB aerial photographs data sources.

Potato yield from 2013 growing season and partly located organic matter sample data sources were used to validate the defined management zones.

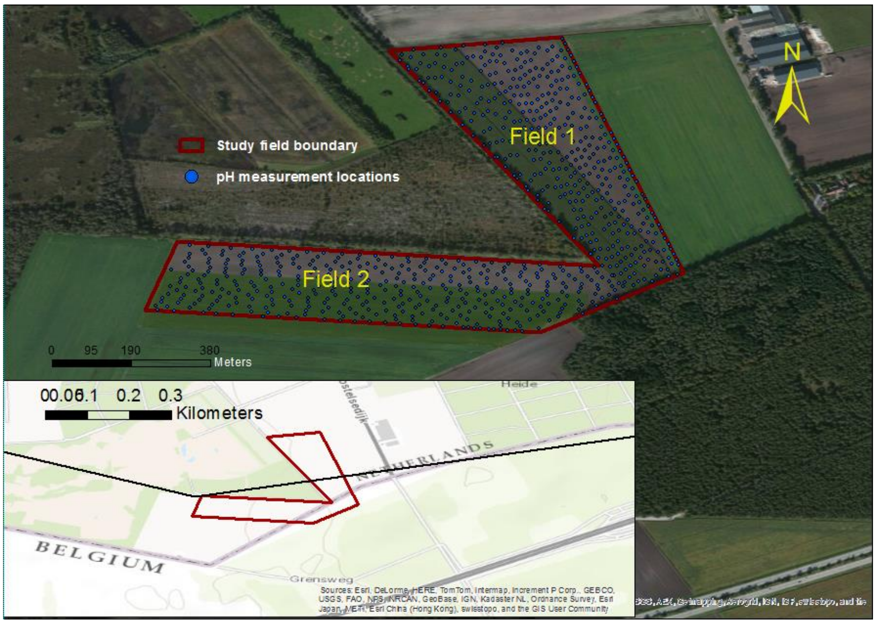


Figure 3. Study area and location of the experimental fields with pH value measurement locations

3.2. Datasets

3.2.1. Electromagnetic induction measurement

A physical and chemical characteristic of a soil determines the electrical conductivity (ECa) of soil. The ability of soil to conduct electricity is usually quantified by electrical resistivity or conductivity. Electrical conductivity measurements have been used to determine salinity using probes inserted to soil. However, recently soil ECa in millisiemens per metre (mS/m), which is used for multiple soil properties characterization, can be measured using the non-invasive, non-destructive EM38-MK2 (Geonics Ltd., Canada) sensor.

The EM38-MK2 sensor is composed of a transmitter coil and receiver coil installed on both ends of the bar. It consists of both horizontal and vertical coils at a 0.5 and 1.0 m distance where measurements are taken in both horizontal (ECa0.5H and ECa1.0H) and vertical (ECa0.5V and ECa1.0V) orientation at three effective depths. The depths ranges are: 1.5 m and 0.75 m for vertical dipole mode and 0.75 m and 0.38 m for horizontal dipole model (Figure 4). The sensor is mounted on wooden sled and pulled by an all-terrain vehicle which drove with a speed range of 1-13 m/s and an average speed of 4 m/s along about 5 m transect and more than meter spacing. The sensor is connected to GPS receiver and towed at ground level using an all-terrain vehicle so that geo-referenced data can recorded on- the-go fashion (Vitharana et al., 2008). ECa values could be affected by moisture content, porosity, salt content, temperature and clay content. As a result ECa measurement values are standardized to reference temperature of 25^{O}C for further analysis according to (Sheets and Hendrickx, 1995).

$ECa_{25}= ECa_{T}\ (0.4470 + 0.403e^{-T/26.815})$

Where ECa25 is the standard temperature at 25^{O}C and ECa_{T} is the ECa value at soil temperature in field. The theoretical depth of influence (DOI) is the depth below the sensor at which 70 % of the cumulative influence of the signal is obtained (Van Meirvenne et al., 2013). The ECa measurements for this study were done on 2013-03-25.

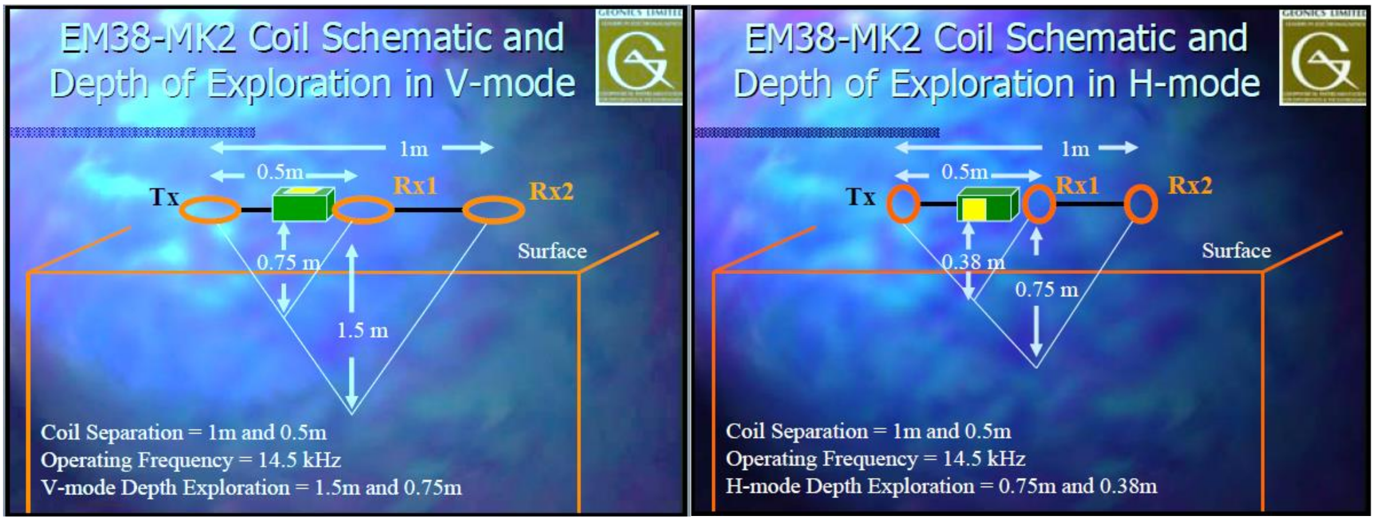


Figure 4. Coil schematic and depth of exploration in both a) vertical b) horizontal orientation (Catalano, 2011)

3.2.2. DEM generation and colour aerial photographs

Recent, very detailed and accurate, AHN2 elevation data were used for this study (AHN.NL). The data was collected by airborne laser scanning (LiDAR) on aircraft platform. The detail elevation data, one point in every half meter is sampled as part of the country's Current Height Netherlands (AHN) project and with accuracy of 5 cm was interpolated to build DEM. The height of the study fields ranges from 28 m to 32 m with maximum difference of 3.86 m. For the analysis, the AHN2 DEM value, where the other soil parameters and topographic measurement located, was extracted.

A colour aerial photograph of the study area taken on 2008 was obtained from the Wageningen University GeoDesk data portal. The RGB colour aerial photograph composed of three bands (red, green and blue) with spatial resolution of 0.25 x 0.25 meter was projected to overlay other datasets.

RGB aerial photograph was then cropped by the extent of study fields for further usage. Four indices, normalized difference Red and Green (NDRG), Red minus Blue (R-B), ratio index (R/B) and sum of all the visible light (SUMVIS), were derived from the colour aerial photo for spatial distribution investigation of soil properties (Bartholomeus and Kooistra, 2012). Then elevation and derived indices values were extracted at exact locations where the pH measurements located (

Figure 5).

3.2.3. Veris sensing technology and validation datasets

Geo-referenced pH values of the study fields were obtained from the Veris pH Manager Instrument (Veris technologies). The Veris pH Manager is built on the Mobile Sensor Platform (MSP) which also holds a GPS to recorded spatial locations. The pH measurements were done at sampling density of about 40 samples per ha. On the other hand denser ECa measurements were done at sampling density of about 1860 samples per hectare.

The pH measurements were located at different spatial locations than the ECa measurements besides the sampling density differences. To spatially link these two spatial datasets, nearest co-located ECa measurements at the pH measurement locations were selected using spatial join algorithm in commercial ArcGIS software (

Figure 5). In this way a dataset consisting of 10 variables at 849 coincident locations with pH measurement location were prepared. Namely pH values, four ECa measurements, elevation, and four aerial photograph indices variables in total were used for this study.

Geo-referenced potato yield (ton/ha) collected in 2013-10-6/7 growing season for both adjacent fields by yield combiner machine was used for validation purpose. The yield combiner in this case is composed of a GPS receiver, yield sensor and user terminal in the cabin so that spatial analysis could be done. In addition to potato yield, existing partly located soil organic matter samples for field one were available for validation.

3.3. Principal component analysis

Principal component analysis (PCA) is a dimension reduction method which uses correlated variables and identifies orthogonal linear recombination of the variables that summarize the principal sources of variability in the data as described by (Abdi and Williams, 2010; Demšar et al., 2013). The authors also explained PCA as most popular and dimension reduction method besides exploring the relationship of the variables. Measurements values of ECa (both vertical and horizontal), pH value, and extracted topographic attributes (elevation value and RGB colour soil indices), composing a total of 10 input variables were used as input for PCA. The data sources were subjected to PCA to identify key variables explaining most variation. So applying the PCA, set of variables were reduced to components prior to analysis. A correlation matrix was used to equal consideration of all the variables.

The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy (MSA) was evaluated to assure the viability of the data set for PCA analysis. Higher values conveniently between 0.5 and 1 value is suggested for good analysis as explained by (Van Meirvenne et al., 2013). Besides to this KMO measure, the Bartlett's test of sphericity which checks the significance of correlation between variables was evaluated (Jolliffe, 2005). The correlation matrix of the input variables was also considered.

To decide the selection of retained principal components (PC), explained variances percentage for each PC and plot of their eigen values (scree plot) were considered (Cattell, 1966; Van Meirvenne et al., 2013; Vitharana et al., 2008). To improve the analysis of the retained factors, a varimax rotation was applied for scenario I. Finally from each of the retained factors, a key variable with highest loading was identified (Van Meirvenne et al., 2013). All the multivariate analysis was done in SPSS (v. 20.0) software.

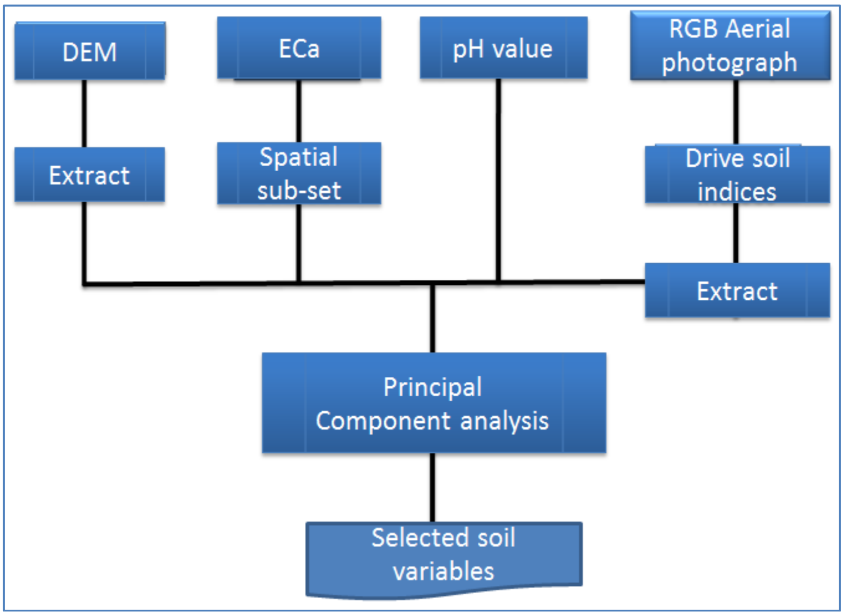


Figure 5. Methodological framework of principal component analysis

3.4. Geo-statistics and clustering algorithm

Soil variation is more continuous than discrete and therefore calls for a continuous approach for soil classification. To characterize the spatial distribution of the selected soil variables, semi-variance analyses were carried out on these selected soil variables using a statistical software package (R version-3.0.2). The spatial structure of each selected soil variable for each retained PCA was analysed geo-statistically. A linear variograms were computed, modelled and fitted using residual maximum likelihood (RMEL) method. Finally the fitted model was used in ordinary kriging technique to produce the continuous values of each selected soil parameters. Five meter grid resolution was used to interpolate the selected soil variables for each scenario to generate continuous soil variable maps (Ortega and Santibáñez, 2007).

With the objective of identifying clusters occurring in the dataset used and generating homogenous sub-fields, interpolated maps of the selected soil variables were classified in to potential management classes using k-means classification clustering algorithm (De Gruijter and McBratney, 1988; Whelan and McBratney, 2003; Fridgen et al., 2000; Li et al., 2007; Moral et al., 2010; Morari et al., 2009; Van Meirvenne et al., 2012; Vitharana et al., 2008). The grouping algorithm method produces a continuous grouping of objects by assigning partial class membership values, which is to be preferred for grouping properties in the soil continuum (Odeh et al., 1992). It assigns and determines individuals to geographical and taxonomical continuous classes. The algorithm search clusters in the data, in such a way that objects belonging to the same cluster resembles each other, whereas objects in different clusters are dissimilar (Ortega and Santibáñez, 2007)..

Figure 6 shows the potential management zonation method for scenario I in this regard. Three of the selected soil variables (NDRG, ECa-V1 and ECa-H1) were interpolated to 5 meter grid resolution to generate continuous property maps. *K*-mean clustering was then implemented on the three soil variables to create potential management zones. R statistical software package was used to perform the k-mean clustering and ArcGIS commercial software was used to produce the maps.

3.5. Crop yield of potential management zones

The potato yield data was pre-processed for spatial distribution yield investigation. Geo-referenced potato yield values were assigned to the defined management zones for each scenario as similarly implemented by (Fraisse et al., 2001; Van Meirvenne et al., 2013). Final analysis of variance (ANOVA) was performed on potato yield between and among the defined management zones to validate whether there is significant mean yield difference between and among the defined management zones. The Tukey HSD test ("Honest Significant Difference") was used to make comparisons between management zones for samples of different sizes (Molin and Castro, 2008). Geo-referenced organic matter samples partly located on field one collected on 2013-06-28 were also used for validation purposed.

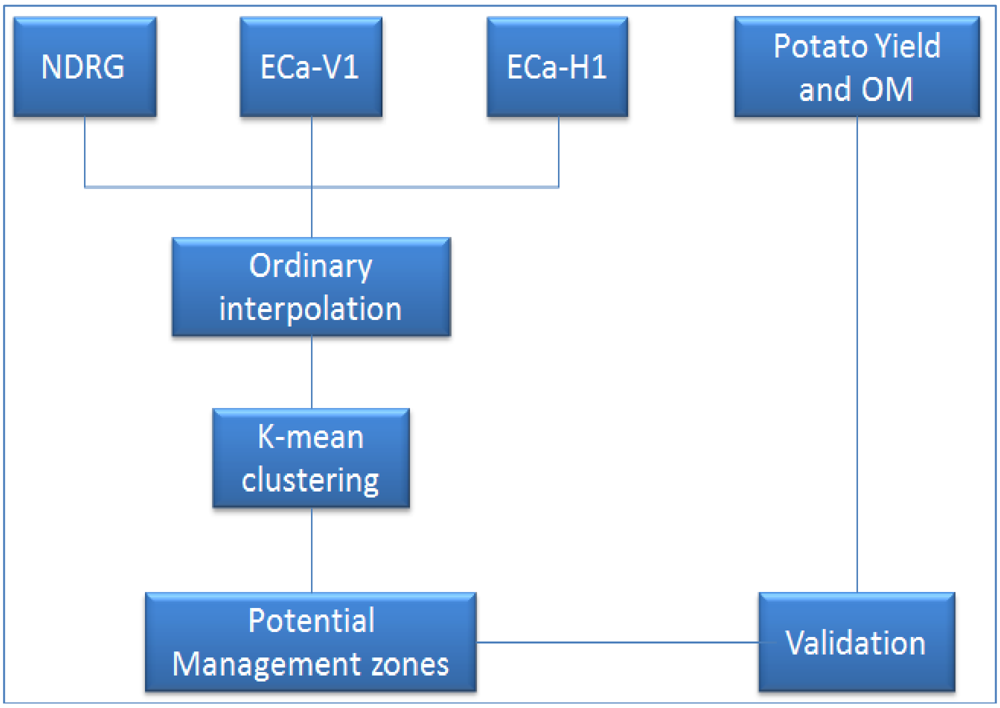


Figure 6. Potential management zonation framework for scenario I

Chapter 4. Result

The result section of the study as stated in the outline of the paper is organized according to the three scenarios. Results are presented and analysed per scenario simplicity. The basic principles and implementations of the methods used are presented per scenario of each topic.

4.1. Conventional statistics of soil properties and crop yield

Descriptive statistics: means, standard deviation (SD), coefficient of variation (CV), maximum values and minimum values, ECa's measurements, pH value, RGB aerial photograph indices, DEM and potato yield data sources are presented in Table 2.

Table 2. Descriptive statistics of the soil variables (N for number of observation and CV for coefficient of variation) for field one.

Soil parameters	N	Mean	Minimum	Maximum	Variance	CV(%)
EMI*(mSm^{-1})*						
ECa-V.5	19,098	21.75	7.9	31.83	1.88	6.31
ECa-H.5	19,098	16.07	-1.21	27.74	1.34	7.21
ECa-V1	19,098	8.80	-1.87	19.43	2.09	19.38
ECa-H1	19,098	-0.39	-4.65	14.44	1.29	-294
Veris pH Manager						
PH value	446	7.09	6.3	8.29	0.21	6.51
RGB aerial photographs						
R-B	446	-1.42	-17.00	48.00	48.06	-485.9
R/B	446	1.01	0.80	3.18	0.02	15.78
NDRG	446	0.00	-0.11	0.52	0.0024	-1301.0
SUMVIS	446	410.0	117.0	502.0	4140.2	15.70
DEM						
Elevation(m)	446	30.91	29.68	32.41	0.42	2.11
Yield (ton/ha)						
Potato yield	27,114	63.91	0.00	446.2	328.4	28.35
Organic matter (OS %)	22	3.47	2.5	4.5	0.27	15.3

The summary statistics for the sensing based soil properties are presented in Table 2 and Table 3 for field one and field two respectively. Of the EMI measurements, ECa-Vertical (both at 0.5 and 1 m) showed higher average value than ECa-Horizontal (both at 0.5 and 1meter). The variance of field one vertical and horizontal of ECa measurements at both 0.5 and 1 meter remains similar while some variation is observed for field two. Exceptionally higher absolute coefficient of variance value for ECa-H1 of field one is observed compared to the other ECa measurement. The low average value of ECa-H in both fields might be due to the lower conductivity of the top soil as the area is known for water stress. On the other hand the mean value of ECa-V.5 for both fields is higher than ECa-V1 (Table 2 and Table 3).

The soil pH of the area ranges from 6.3 to 8.29 and 6.26 to 8.77 for field one field two respectively. Field one has 7.09 mean pH value while 7.43 for field two. The soil in this regard might be roughly categorised as neutral soil. The coefficient of variance was also 6.51 and 6. 87 for field one and field two respectively. The descriptive statistics also shows high (above 328) variability of potato yield in both fields. An average mean potato yield of 63.91 and 61.67 was recorded for field one and field two respectively. Lower (~ 0) value of yield was recorded probably refers to harvesting or driving lines. In contrast few high values outliers of yield records were also found.

Table 3. Descriptive statistics of the soil variables (N for number of observation and CV for coefficient of variation) for field two.

Soil parameters	N	Mean	Minimum	Maximum	Variance	CV(%)
EMI*(mSm⁻¹)*						
ECa-V.5	21,875	20.53	11.35	35.64	1.88	6.67
ECa-H.5	21,875	17.33	6.78	32.93	4.45	12.17
ECa-V1	21,875	9.25	0.00	16.70	2.30	16.41
ECa-H1	21,875	2.00	-4.02	11.30	0.77	44.09
Veris pH Manager						
PH value	404	7.43	6.26	8.77	0.25	6.78
RGB aerial photographs						
R-B	404	-2.60	-14.00	13.00	19.75	-170.97
R/B	404	0.98	0.91	1.10	0.0008	2.95
NDRG	404	-0.004	-0.05	0.05	0.0002	-181.6
SUMVIS	404	466.2	350.0	537.0	1127	7.2
DEM						
Elevation(m)	404	30.09	28.55	31.94	0.84	3.05
Yield (ton/ha)						
Potato yield	26,717	61.67	0.0	559.9	344.9	30.11

4.2. Principal component analysis

4.2.1. Scenario I

The output of PCA, KMO and Bartlett's test of spherity was used to assess the suitability of factor analysis for the variables. Bartlett's test of spherity indicates significant correlation level between the variables having a value of 0.00 which is less than 0.05 satisfying the requirement. The KMO measure of sampling adequacy was found 0.60 indicating the PCA is appropriate for analysis for the variables. KMO should be between 0.5 and 1 for the appropriateness consideration of further step (Van Meirvenne et al., 2013). In addition, substantial correlations matrix was used for appropriateness check-up of the factor analysis for the variables. For the used soil variables, more than 20 correlations between the soil parameters were found satisfying the requirement (greater than 0.30 correlations Table 4). According to Johnson and Wichern (2002) if there are few correlations above 0.3, it is not appropriate to carry further analysis, clearly we do not have that problem.

Table 4.Correlation matrix of the soil variables used for principal component analysis in scenario I.

Correlation Matrix

	pH value	ECa-V.5	ECa-H.5	ECa-V1	ECa-H1	Elevation	NDRG	RB	RmB	SUMVIS
pH value	1,000	-,067	,182	,125	,391	-,162	-,037	-,044	,023	,071
ECa-V.5	-,067	1,000	,063	,694	-,242	,092	,266	,222	,333	-,315
ECa-H.5	,182	,063	1,000	,332	,500	-,169	-,097	-,089	-,096	,085
ECa-V1	,125	,694	,332	1,000	,218	-,244	,147	,115	,205	,022
ECa-H1	,391	-,242	,500	,218	1,000	-,323	-,149	-,134	-,147	,314
Elevation	-,162	,092	-,169	-,244	-,323	1,000	-,027	-,009	-,062	-,420
NDRG	-,037	,266	-,097	,147	-,149	-,027	1,000	,960	,880	-,525
RB	-,044	,222	-,089	,115	-,134	-,009	,960	1,000	,764	-,487
R-B	,023	,333	-,096	,205	-,147	-,062	,880	,764	1,000	-,441
SUMVIS	,071	-,315	,085	,022	,314	-,420	-,525	-,487	-,441	1,000

A first rule which considers principal components with eigen value greater than one needs to considered is followed. The scree plot which shows the plot of eigen value on ordinate versus number of components on abscissa was considered next to help for deciding the number of components (Appendix 2). Considering the default setting, four PC's appear to be important components with eigen value greater than 1. Four of these PC's explained 80.32 % of the total variation in the dataset. However, the natural break from the scree plot was not so clear to decide (Appendix 1). Therefore forced factor extraction PCA procedure was applied to optimize the interpretation. Then three PC's explaining 70% (PC1 33.9%, PC2 21.7% and PC3 14.35%) of the total variance were retained (Li et al., 2007). Table 5 presents the loadings and communalities of the three rotated PCs of sensing based soil parameters. Almost all of the variables resulted communality value greater than 0.6 except pH value which gives less than 0.51.

Table 5. Rotated factor loading of the first three PCs and the communalities of each variable of scenario I.

		Principal component loadings		
Soil parameters	Communality	PC1	PC2	PC3
EMI measurements (mSm^{-1})				
ECa-V.5	0.91	0.22	-0.24	**0.88**
ECa-H.5	0.62	-0.13	0.56	0.37
ECa-V1	0.90	0.08	0.29	**0.89***
ECa-H1	0.76	0.15	**0.83***	0.01
Veris pH Manager				
PH value	0.51	0.04	0.57	-0.02
RGB aerial photographs				
R-B	0.85	**0.90**	0.02	0.14
R/B	0.89	**0.94**	-0.01	0.24
NDRG	0.96	**0.98***	-0.01	0.60
SUMVIS	0.77	-0.58	0.44	-0.15
DEM				
Elevation(m)	0.84	-0.04	**-0.67**	-0.01

Figure 7 presents the rotated loadings matrix plot or factor pattern matrix of the retained PCs to show the correlation among the components and variables. The first PC was strongly associated with optical soil indices. Most of the indices contributed high (> 0.90) loadings values, with NDRG the highest loading (0.98) value. All the other soil variables contributed lower loadings for PC1. Combination of soil variables contributed for the PC2. ECa-H1, pH value and elevation contributed for PC2 loadings with highest loading value of 0.83 from ECa-H1. Elevation was in an inverse relationship with ECa-H1. The third component is mainly related to ECa-vertical (both at 0.5 and 1 meter) depths. Both ECa-V.5 and ECa-V1 showed comparable high loadings of 0.89 and 0.88 values respectively. NDRG has 0.60 loading for the third PC.

In this regard PCA determined which soil variables are most important in characterization most of the variability in data source used. As a result PCA identified, NDRG, ECa-H1 and ECa-V1 as foremost soil variables for delineation of potential management zones in the current study under scenario I, unlike to (Fraisse et al., 1999; Van Meirvenne et al., 2012; Vitharana et al., 2008) who identified ECa, soil pH and elevation as key soil parameters. Out of the total 10 soil variables used, three of them were retained by explaining about 70 % of the variations that exists. Looking back at the correlation matrix, there exist weak correlations among these selected soil variables indicating to an independent spatial distribution. Both ECa-H1and ECa-V1 this days can be obtained from an intensive on-the-go observations and of course NDRG can also be easily obtained by UAV sensors (Adamchuk et al., 2005). RGB colour aerial photographs are available geo-data at the national level in the country.

a) b)

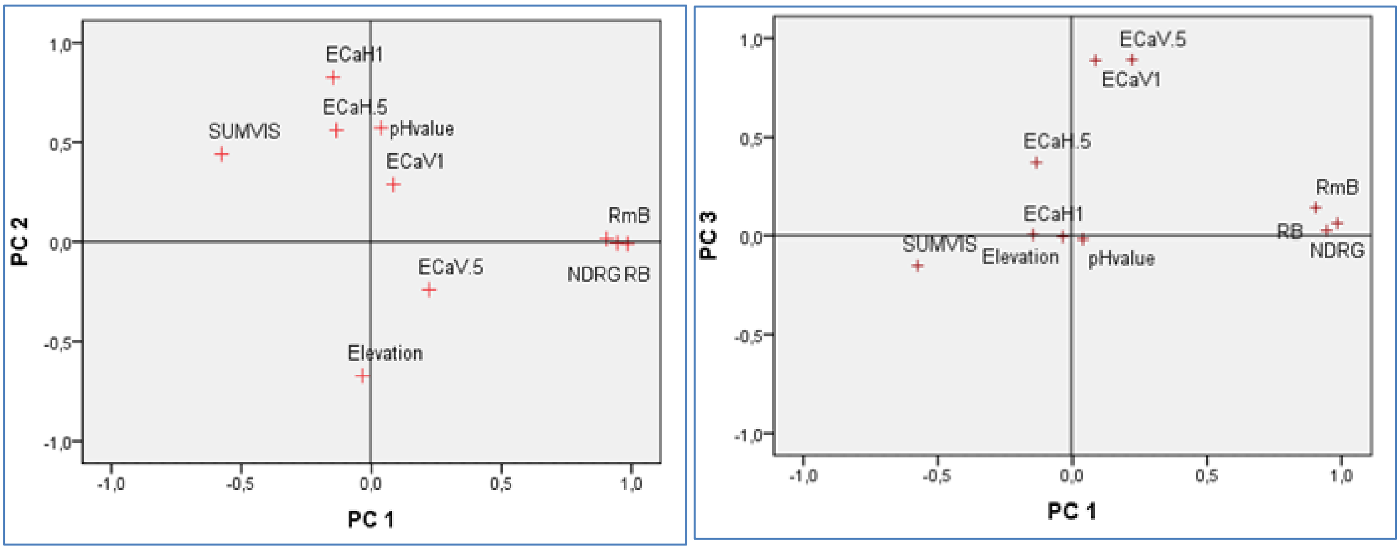


Figure 7. Rotated loading matrix plots of scenario I, a) first and second principal component and b) first and third Principal component.

4.2.2. Scenario II

In this section the adjacent fields are treated separately for analysis purpose and the output results are presented side by side. KMO and Bartlett's test of spherity was used to assess appropriateness of data for PCA for both fields and a significant correlation among soil variables was found. The Bartlett's test of spherity was found 0.00 for both fields. The KMO measure of sampling adequacy which tells the appropriateness of the data was found to be 0.62 and 0.66 for field one and field two respectively (Van Meirvenne et al., 2013; Vitharana et al., 2008). Correlations assessment was also done for appropriateness check-up of the factor analysis in similar manner as in scenario I. And more correlations with values greater than 0.30 among the soil variables with were found for both fields.

PCs with eigen value greater than one were retained for further analysis, the first rule. The scree plot was also considered as a complement to help the decision of PCs selection (Appendix 2). Using a default setting, three PCs appear to be with clear Eigen eigenvalue greater than one. The scree plots test of both fields in this case also showed a natural break at eigen value greater than one (Appendix 2). And it was decided to retain three PCs for further analysis. The three PCs accounts for 72.87% and 69.35% (Table 6) of the variability for field one and field two respectively (Li et al., 2007). The total variance explained by each PC and cumulative percentage of variations explained for both field one and field two are presented (Table 7).

Table 8 gives the principal communality loadings of the three principal components and the communalities of each soil variable for field one. Almost all of the soil variables resulted communality value greater than 0.55 except pH value with lower 0.37 value. The first PC was dominantly connected with the four optical soil indices. Out of these four, NDRG has the largest loading (0.96) and SUMVIS with absolute lowest loading (-0.79) value. The correlation matrix (Table 6) showed a high correlation matrix among NDRG, RmB, and R-M. Other soil variables showed by far low loadings contribution s to PC1in this regard.

ECa-V1 was the strongest contributor for the second PC with loadings value of (0.88). ECa-V.5 also was the second highest (0.86) contributor for the second PC. Elevation though with opposite sign contributed (-0.72) loading for the same PC.

The third PC was strongly associated with ECa-H1 and ECa-H.5 while ECa-H1 having the highest (0.74) loading. ECa-H1 and ECa-H.5 also showed strong association with positive sign. The pH value with 0.59 loading was a third contributor for this PC but still it remains with very low contribution for the other two PCs.

In summary, the result of PCA for field one, suggested that the overall spatial variability of the considered total soil variables was aggregated in to three principal components and NDRG, ECa-V1 and ECa-H1were identified as the most dominant soil variables for the three components respectively.

Table 6. Correlation matrix of the soil variables used for principal component in scenario II. The top is for field one and bottom for field two.

Correlation Matrix

		pH.value	ECa.V.5	ECa.H.5	ECa.V1	ECa.H1	Elevation	NDRG	RB	RmB	SUMVIS
Correlation	pH.value	1,000	,105	,087	,060	,232	-,083	-,004	-,017	,081	-,186
	ECa.V.5	,105	1,000	-,057	,823	-,086	-,416	,304	,253	,408	-,205
	ECa.H.5	,087	-,057	1,000	-,048	,268	-,255	-,064	-,074	,017	-,024
	ECa.V1	,060	,823	-,048	1,000	-,042	-,413	,211	,171	,305	-,076
	ECa.H1	,232	-,086	,268	-,042	1,000	,051	-,116	-,098	-,118	-,052
	Elevation	-,083	-,416	-,255	-,413	,051	1,000	-,198	-,136	-,348	-,130
	NDRG	-,004	,304	-,064	,211	-,116	-,198	1,000	,965	,898	-,627
	RB	-,017	,253	-,074	,171	-,098	-,136	,965	1,000	,800	-,569
	RmB	,081	,408	,017	,305	-,118	-,348	,898	,800	1,000	-,582
	SUMVIS	-,186	-,205	-,024	-,076	-,052	-,130	-,627	-,569	-,582	1,000

Correlation Matrix

		pH.value	ECa.V.5	ECa.H.5	ECa.V1	ECa.H1	Elevation	NDRG	RB	RmB	SUMVIS
Correlation	pH.value	1,000	,095	,084	,064	,046	,106	,044	,054	,054	-,098
	ECa.V.5	,095	1,000	,338	,848	,295	,136	,198	,193	,203	-,117
	ECa.H.5	,084	,338	1,000	,551	,727	,035	-,167	-,191	-,174	-,110
	ECa.V1	,064	,848	,551	1,000	,435	-,006	,094	,080	,092	-,062
	ECa.H1	,046	,295	,727	,435	1,000	,016	-,088	-,103	-,094	-,110
	Elevation	,106	,136	,035	-,006	,016	1,000	,128	,132	,136	-,532
	NDRG	,044	,198	-,167	,094	-,088	,128	1,000	,971	,980	-,022
	RB	,054	,193	-,191	,080	-,103	,132	,971	1,000	,992	,001
	RmB	,054	,203	-,174	,092	-,094	,136	,980	,992	1,000	-,019
	SUMVIS	-,098	-,117	-,110	-,062	-,110	-,532	-,022	,001	-,019	1,000

Table 7. Principal components and variations explained by each component for both fields of scenario II

	Field one		Field two	
Principal Components	Variance Explained (%)		Variance Explained (%)	
	individual	Cumulative	individual	Cumulative
1	31.24	31.24	36.59	36.59
2	26.56	57.8	18.21	54.8
3	15.07	72.87	14.54	69.35

Table 8. Unrotated factor loading of the first three principal component and the communalities of soil parameters for field one of scenario II

		Principal component loadings		
Soil parameters	Communality	PC1	PC2	PC3
EMI measurements (mSm^{-1})				
ECa-V.5	0.67	0.21	**0.86**	-0.05
ECa-H.5	0.47	-0.04	-0.06	**0.68**
ECa-V1	0.79	0.08	**0.88***	-0.05
ECa-H1	0.57	-0.05	-0.11	**0.74***
Veris pH Manager				
PH value	0.37	0.08	0.09	0.59
RGB aerial photographs				
R-B	0.87	**0.87**	-0.32	-0.00
R/B	0.88	**0.93**	0.10	-0.10
NDRG	0.95	**0.96***	0.16	-0.09
SUMVIS	0.66	**-0.79**	0.08	-0.17
DEM				
Elevation(m)	0.55	-0.04	-0.72	0.17

Table 9 gives the principal communality loadings of the three principal components and the communalities of each soil variable for field two. Most of the soil parameters showed a communality greater than 0.6, with exceptionally low (0.09) from pH value. R-B, R/B and NDRG showed most significant influence on first PC with highest (0.987) loading from R-B. For the second PC four of the ECa measurements showed large contributions rate relative to other soil variables. Highest loading (0.89) was contributed from ECa-V1. Elevation and SUMVIS on the other hand contributed large contribution for the third PC with highest loading (0.87) from elevation, although each of them correlated in opposite direction. The pH value variable appeared weakly related to the three PC appearing to be less informative compared to other soil variables.

In summary R-B, ECa-V1 and elevation were selected for the three PCs respectively by explaining most of the variability of data source for field two. Although the same datasets was used as field one, different soil variables were selected out of the PCA analysis.

Table 9. Unrotated factor loading of the first three Principal component and the communalities of soil parameters for field two of scenario II

	Principal component loadings			
Soil parameter	Communality	PC1	PC2	PC3
EMI measurements (mSm^{-1})				
ECa-V.5	0.67	0.27	**0.77**	0.12
ECa-H.5	0.71	-0.24	**0.80**	0.09
ECa-V1	0.82	0.14	**0.89***	0.00
ECa-H1	0.59	-0.16	**0.75**	0.08
Veris pH Manager				
PH value	0.09	0.05	0.10	0.28
RGB aerial photographs				
R-B	0.98	**0.987***	-0.00	0.06
R/B	0.97	**0.985**	-0.07	0.05
NDRG	0.96	**0.98**	0.00	0.06
SUMVIS	0.75	0.05	-0.05	**-0.86**
DEM				
Elevation(m)	0.75	0.09	-0.04	**0.87***

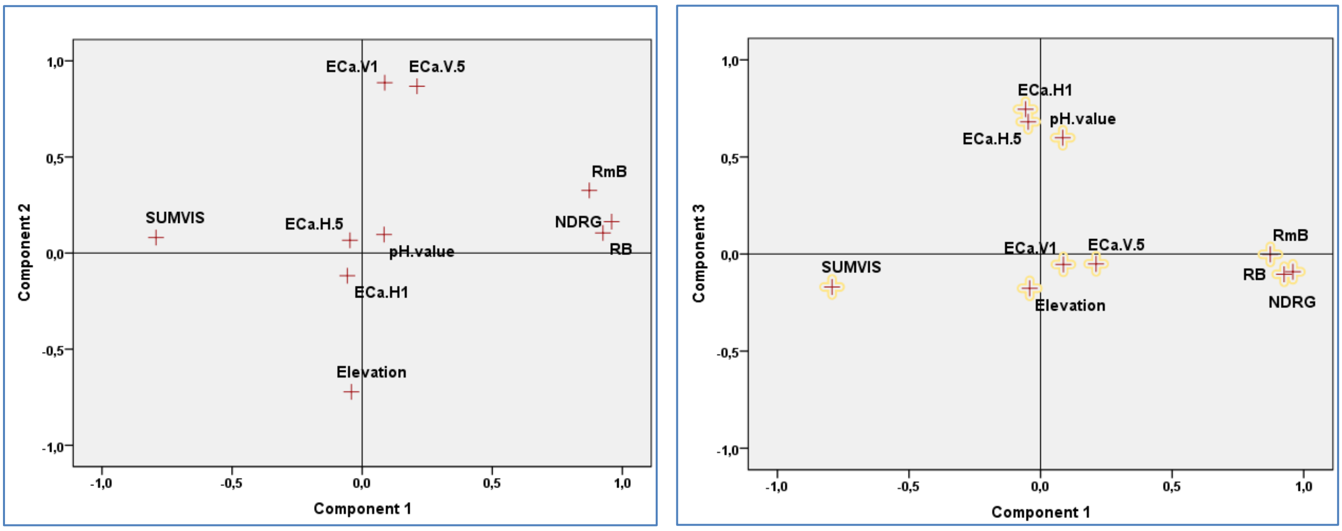


Figure 8. Field one loading matrix plots of scenario II, first and second principal component (left) and first and third principal component (right).

It seems reasonable to tentatively identify the first principal component as NDRG, R-B, and R/B, and all have high loadings on it (Figure 8). On the other hand elevation and SUMVIS seems identifiable on the second principal component though on opposite directions (Jolliffe, 2005).

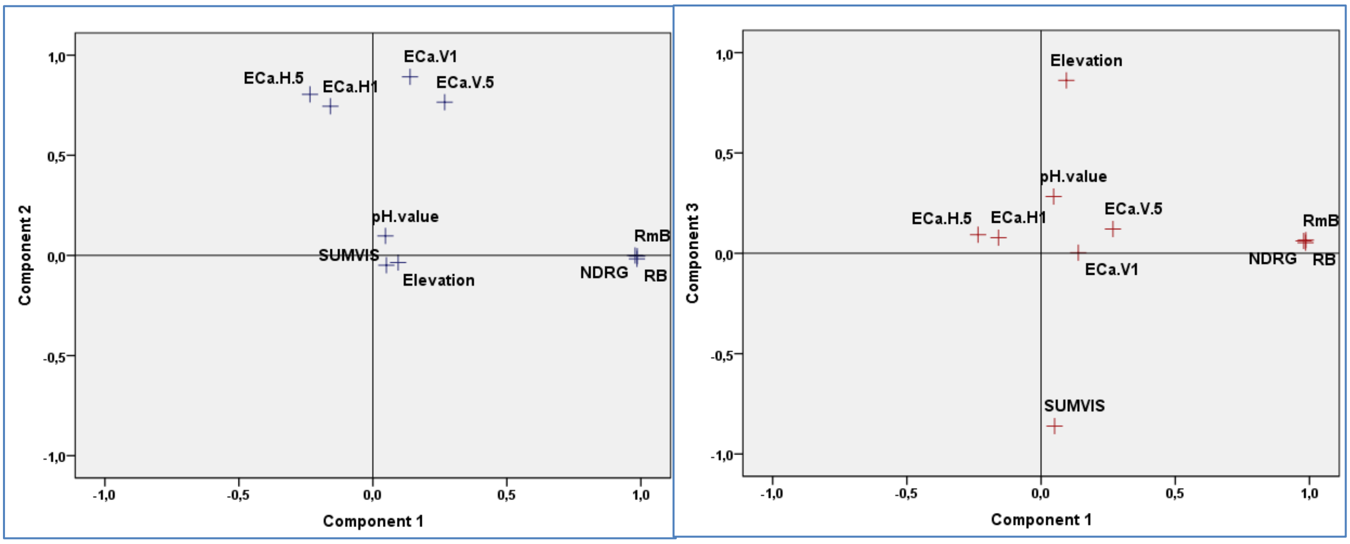


Figure 9. Field two loading matrix plots of scenario II, first and second principal components (left) and first and third principal component (right).

4.2.3. Scenario III

For this scenario available geo-data, elevation and colour aerial photographs data sources were used. Table 10 provides the correlations matrix between the soil variables which basically showed the appropriateness of the dataset for factor analysis. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy for the overall dataset and the KMO measure for each individual variable were found 0.56. Bartlett's test of sphericity for data suitability reduction was also found significant with value of 0.000.

Table 10. Correlation matrix of soil variables of scenario III.

variables	Elevation	NDRG	R-B	RmB	SUMVIS
Elevation	1.00	-0.03	-0.01	-0.06	-0.42
NDRG	-0.03	1.00	0.96	0.88	-0.52
RB	-0.01	0.96	1.00	0.76	-0.48
RmB	-0.06	0.88	0.76	1.00	-0.44
SUMVIS	-0.42	-0.52	-0.48	-0.44	1.00

Two meaningful initial principal components were retained for further factor analysis based eigen value greater one. The natural break of the scree plot test criterion also coincides with eigen value greater than one. The total variance explained by each PC and cumulative percentage of explained variations is presented in Table 11.

Table 11. Principal Components and variations explained by each component of scenario III.

Principal Components	Variance explained (%)	
	individual	Cumulative
1	61.6	61.6
2	25.4	87.1

Table 12 gives the principal communality loadings of the two principal components and the communalities of each soil variables. All soil variables in this case showed high communality value. Communality apparently refers to the total influence on a single observed variable from all the components associated with and it ranges from zero to one (Jolliffe, 2005).

The first PC was dominantly connected to R-B, R/B, and NDRG with highest loading contributed by NDRG. Elevation presented a lowest (-0.13) absolute loading value for the first PC. On the other hand elevation and SUMVIS presented large contribution for the second PC relative to the other soil variables. Elevation showed highest (0.92) loading for the second PC. In summary, PCA aggregated the soil parameters in two principal components accounting for the majority of spatial variations in the data used. Elevation and NDRG were selected for each component respectively.

Table 12. Unrotated factor loading of the first two PCs and the communalities of soil variables for field two of scenario III.

		Principal component loadings	
Soil parameters	Communality	PC1	PC2
RGB aerial photographs			
R-B	0.84	0.91	0.01
R/B	0.89	0.94	0.08
NDRG	**0.97**	**0.98**	0.08
SUMVIS	0.76	-0.51	-0.70
DEM			
Elevation(m)	0.88	-0.13	**0.92**

The PCA also produces a plot of the variables used for the current scenario on axes representing the two unrotated components. As shown from (Figure 10), it seems reasonable to tentatively identify the first principal component associated with NDRG, R-B, and R/B, and all have high loadings on it. Elevation and SUMVIS were also identified on the second principal component though in opposite directions.

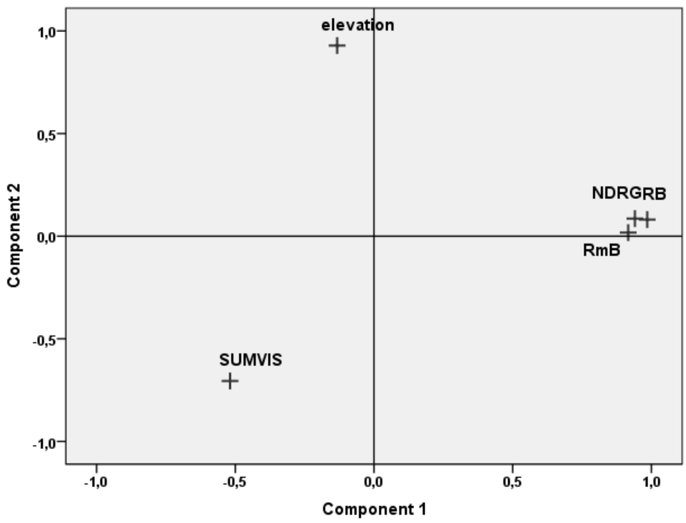


Figure 10. Unrotated loading matrix plots of the first and second principal components.

4.3. Map of selected soil parameters

4.3.1. Scenario I

Semivariograms of the selected variables: NDRG, ECa-H1 and ECa-V1 were computed to examine the spatial structures or correlations after observing a normal distribution for most of the variables. Spherical or exponential models were fitted to the parameters depending on which fits best to the data curve as described by (Liu et al., 2006; Zhu and Lin, 2010). Best fitted models were selected for each soil variable. The experimental and fitted semivariograms models of the selected soil variables are presented (Appendix 3). Each of NDRG and ECa-V1 showed a spatial autocorrelation from their semivariogram cloud and a reasonable model. Both ECa-V1 and NDRG were modelled by exponential and spherical models respectively. ECa-H1was not successfully modelled neither spherical nor exponential and of course less spatially structured. Parameters of models fitted to the computed semivariograms are presented (Table 13). There seems a presence of small nugget variance in three of soil variables. This might probably be due to short range variability and unaccountable measurement error (Li et al., 2007).

Each of the soil variables was then interpolated using ordinary kriging geostatistics method to create continuous soil variables map of 5 x 5 meter resolution same for all variables, which later helps to define the management classes (Van Meirvenne et al., 2013). Extracted NDRG pixel values and nearest co-located ECa-H1 and ECa-V1 variables at 849 (for both fields) locations were used for the interpolation.

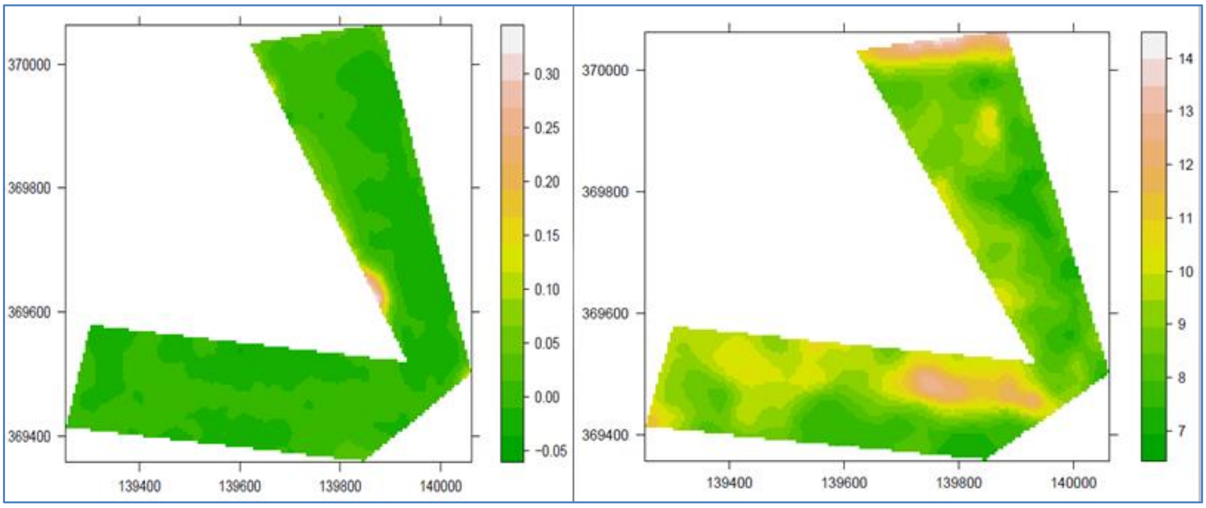
Table 13. Parameters of models fitted to semivariograms of the NDRG, ECa-V1, and ECa-H1 soil variables of scenario I.

Soil Property	Model fit	Nugget	Sill	Range
NDRG	Spherical	0.0002	0.0018	211.5
ECa-V1	Exponential	0.43	2.36	50.23
ECa-H1	Spherical	0.61	2.78	526

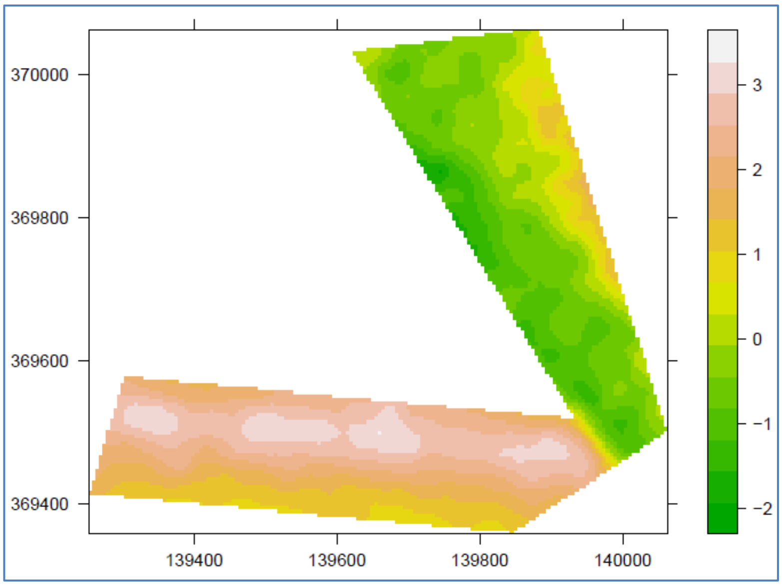
Kriged maps of the three soil variables were produced and are presented in Figure 11. There seems in general very little distinct pattern similarity between NDRG and ECa-V1 although, it is not straight forward. Lower value of NDRG is observed in the eastern part of field one. And relative lower values of NDRG for field two is observed in southern and northern parts of field partly. A higher value of NDRG is also observed in the western parts of field one. This might be probably due to the vegetation edge influence. The presences of vegetation on border of field one might be the cause for high outlier value of NDRG index in the lower western part.

Unlike ECa-V1, ECa-H1 generally showed a clear inter-field distinction between fields than intra-zone distinction both fields. The distinction between the adjacent fields is more pronounced than the distinction within zones of each field (Figure 11 c). The spatial distribution of ECa-H1 seems actually in opposite direction to that of ECa-V1. The noticeable zonal distinction between the two fields probably might be due to soil fertility status differences though this situation is not reflected in the other ECa measurements. The pattern distribution of ECa-H1 and NDRG seems opposite especially for field one as both of them are poorly correlated (Table 4).

Kriged map of ECa-V1 (Figure 11, b) showed a higher value in the western and north part of field one. Eastern and central parts of field one corresponds dominantly to lower values. On the other hand two broadly less distinct zones can be observed (north and south) for field two. The northern part of field two shows higher values while the southern part of the field shows to lower values.



a) b)



c)

Figure 11. Interpolated maps using kriging of, a) NDRG b) ECa-V1 c) ECa-H1 soil variables for scenario I.

4.3.2. Scenario II

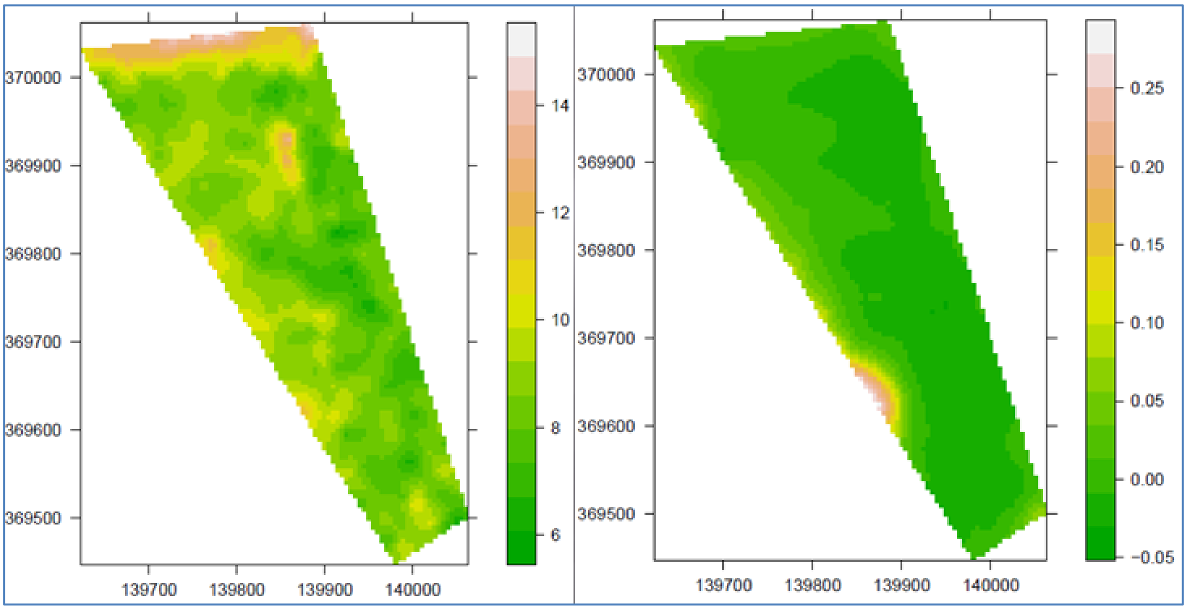
Field one: In similar fashion to section 4.3.1, spatial structures of NDRG, ECa-H1 and ECa-V1 semivariograms were examined. Spherical or exponential models were also fitted to these parameters depending on appropriateness fit which best fits to data curve (Banerjee et al., 2004; Liu et al., 2006; ZHU and Lin, 2010). The experimental and fitted semivariograms models of these soil variables are presented (Appendix 4). Semivariograms of NDRG and ECa-V1 showed good spatial structure and reasonable model. Both ECa-V1 and NDRG were modelled using exponential and spherical models respectively. Parameters of models fitted to the computed semivariograms are presented (Table 14).

The spatial pattern of NDRG for field one in the current scenario showed a similar pattern to scenario I. For the ECa-V1 relatively high measurement values in the northern parts of the field is observed. The central eastern part of the field is dominantly of low ECa-V1 values though irregularity or patchiness of the pattern dominates.

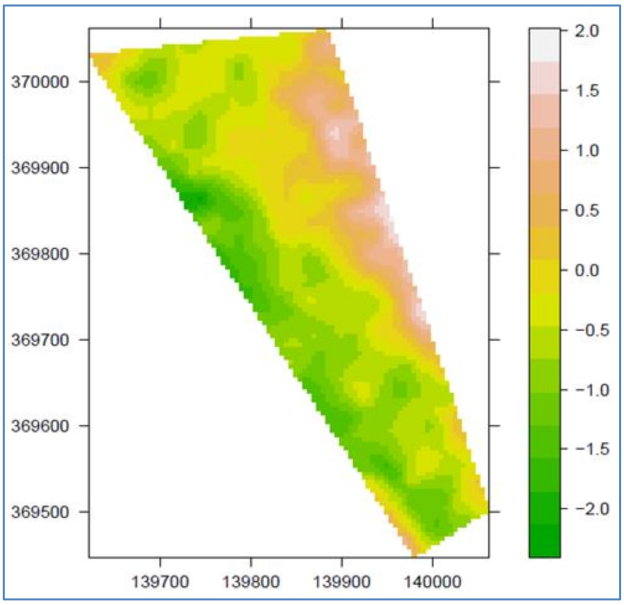
Three major zones can be seen evidently identified from the ECa-H1 kriged map. The eastern part of the field corresponds to high ECa-H1 values in contrast to NDRG. The western part showed lower ECa-H1 values. Another class which corresponds to intermediate values can also be evidently identified in the central part of the field. The correlation coefficient between ECa-H1 and ECa-V1was also found very low (-0.042) Table 6 a.

Table 14. Parameters of models fitted to semivariograms of the NDRG, ECa-V1, and ECa-H1 soil variables of scenario II and field one.

Soil Property	Model fit	Nugget	Sill	Range
NDRG	Spherical	0.0007	0.002	256
ECa-V1	Exponential	0.37	2.04	30.69
ECa-H1	Spherical	0.98	16.25	5324



a) **b)**



c)

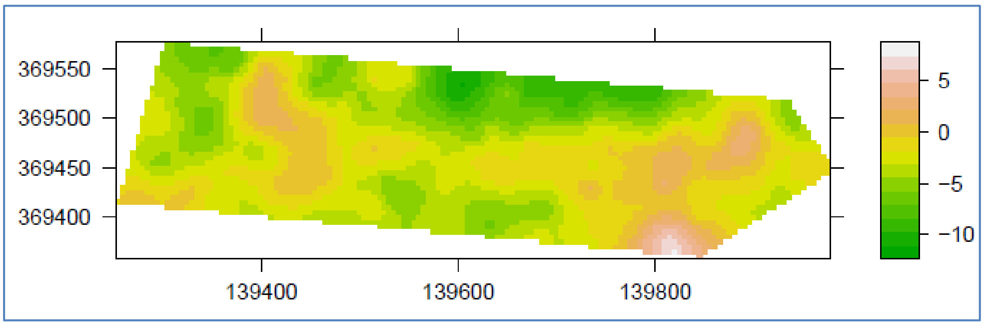
Figure 12. Interpolated maps using kriging of, a) ECa-V1 b) NDRG c) ECa-H1 soil variables for scenario II field one.

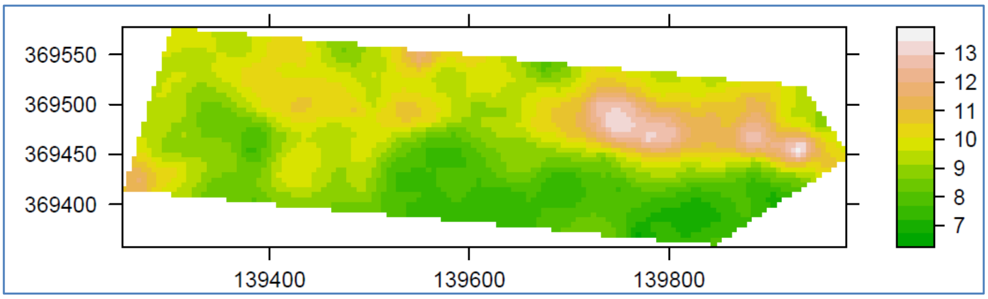
Field two: For RmB, ECa-V1 and Elevation soil variables, semivariograms were also computed. Spherical and Exponential models were also fitted accordingly in similar manner to above sections section 4.3.1. Parameters of models fitted to the computed semivariograms are presented in Table 15. The experimental and fitted semivariograms models of soil variables are presented (Appendix 5). Ordinary kriging was then applied to produce kriged maps (Figure 13) of the soil variables.

The RmB map indicated larger values in the central and eastern parts of the field. Lower value is also indicated in the northern and western parts of the field. As from the ECa-V1 kriged map, lower values are indicated dominantly southern part of the field and larger values dominates the northern part in contrary to RmB. There seems the distribution of RmB is in opposite direction to ECa-V1 distribution.

Table 15. Parameters of models fitted to semivariograms of the ECa-V1, RmB and elevation soil variables of scenario II field two.

Soil Property	Model fit	Nugget	Sill	Range
ECa-V1	Spherical	0.47	2.12	118.74
RmB	Exponential	8.7	13.54	64
Elevation	Spherical	0	3	2500





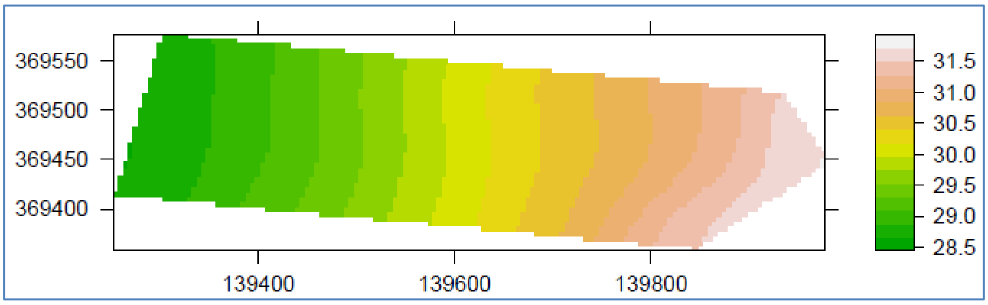


Figure 13. Interpolated maps using kriging of, a) RmB b) ECa-V1 c) elevation soil variables of scenario II field two.

4.3.3. Scenario III

In the current scenario nation's available geo-data elevation and colour aerial photography data sources were used. NDRG index and elevation were selected key variables by explaining most of the soil variations. These parameters used for the fitted semivariogram model for structural analysis are presented in Table 16. And the experimental semivariogram of soil parameters are presented (Appendix 6). Both parameters were also interpolated with ordinary kriging so that a continuous map of each variable was produced in a similar procedure as described for scenario I and II.

Table 16. Parameters of models fitted to semivariograms of the NDRG and elevation

variable	Model fit	Nugget	Sill	Range
NDRG	Spherical	0.866	6.86	1150
Elevation	Spherical	0.99	6.22	6220.8

The spatial distribution of the NDRG in this scenario is same to scenario I and II. The spatial distribution of elevation map regularly increase from west to east and from north west to south east for field two and field two respectively (
Figure 14). Of course NDRG and elevation are poorly (-0.03) correlated.

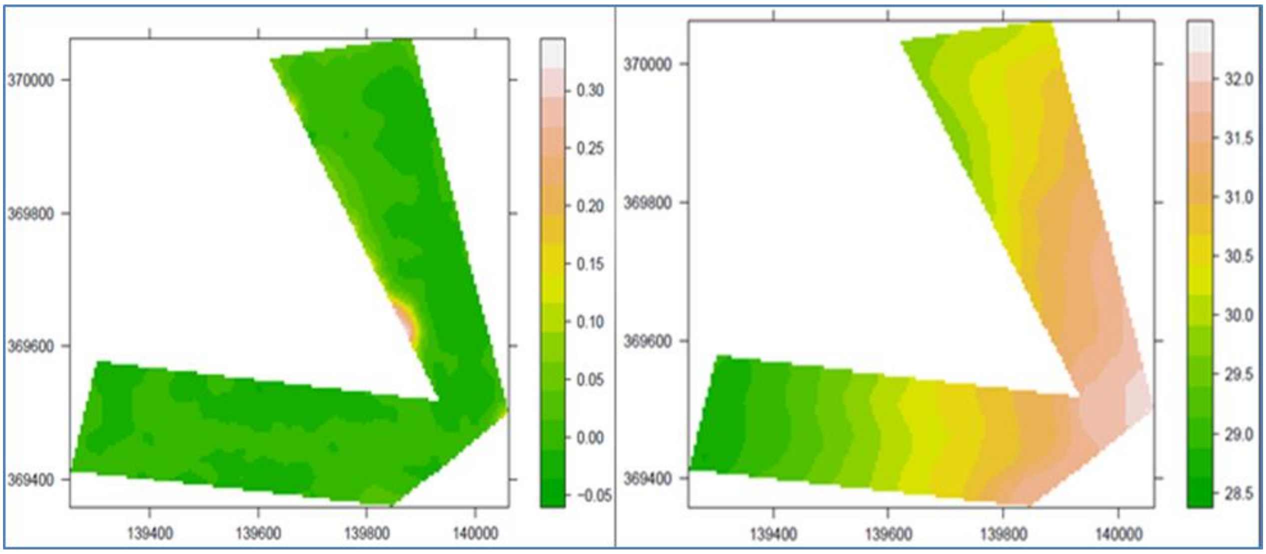


Figure 14. Interpolated maps using kriging of, elevation (right) and NDRG (left) soil variables of scenario III.

4.4. Management zone delineation

4.4.1. Scenario I

K-mean unsupervised clustering method partitioned measurements in to *k* groups such that the sum of squares from points to the assigned cluster centre is minimized. *K-mean* clustering was used to define naturally occurring clusters in the soil variables. Clustering of kriged maps used to define potential management zone.

The number of clusters within a field is basically a function of: 1) natural variability within the field, 2) size of the field and 3) practical use of the management zones It was stated that no advantage can be gained by dividing a field in to more than four or five classes though size of the field needs to be considered (Fraisse et al., 2001; Fridgen et al., 2000). They also indicated the variation of optimum number of zones depending on weather and type of crop to be planted. As the number of classes increase less pronounced, possibly less interpretable and patchy classes were formed. Therefore different class centres numbers were tried for this purpose. Finally three clusters were selected for further applications.

Finally three clusters were decided to define the potential management zones. NDRG, ECa-V1 and ECa-H most variations explaining soil parameters were used as input to the k-mean classification algorithm. The defined potential management classes are presented in Figure 15.

Generally, a spatial distribution pattern similarity seems to exist between NDRG and ECa-V1 kriged maps and the defined potential management class. But the similarity of pattern is stronger with NDRG. However, no such similarity in pattern was observed between ECa-H1 and potential management classes. From area coverage prospective, class 1 and class 2 covers about comparable size while class 3 covers smaller part of the field.

Zone 1, occupied the northern and western borders parts of field one. This class corresponds to relatively high value of NGRG and ECa-V1. However the zone corresponds to lower value of ECa-H1. Small islands of zones were also observed in field two (Figure 15). For this particular zone, high value of ECa-V.5 was also observed though the parameter was not incorporated in the clustering. Relatively lower pH value covers this area. Relatively higher average yield was recorded for this zone though, it is not straight forward to specify spatial patterns between the management zone and yield map.

Zone 2, occupied the central area between zone 1 and zone 3 in field one. Eastern and central western parts of field two are also occupied by this zone. This zone coincides with relatively medium value of NDRG and ECa-V1 for field one. The higher ECa-V1 value of the field two also coincides with this zone. An average medium yield was recorded for the zone.

Zone 3, is located in the central part of field one bordered by zone two. It also is located in the western and central south and north parts of field two. This zone corresponds to lower NDRG and mixed (low and medium) ECa-V1 values. The zone also corresponds to mixed (medium and higher) values of Eca-H1values. In summary there is no straight spatial autocorrelation among soil variables, yield and potential management zones.

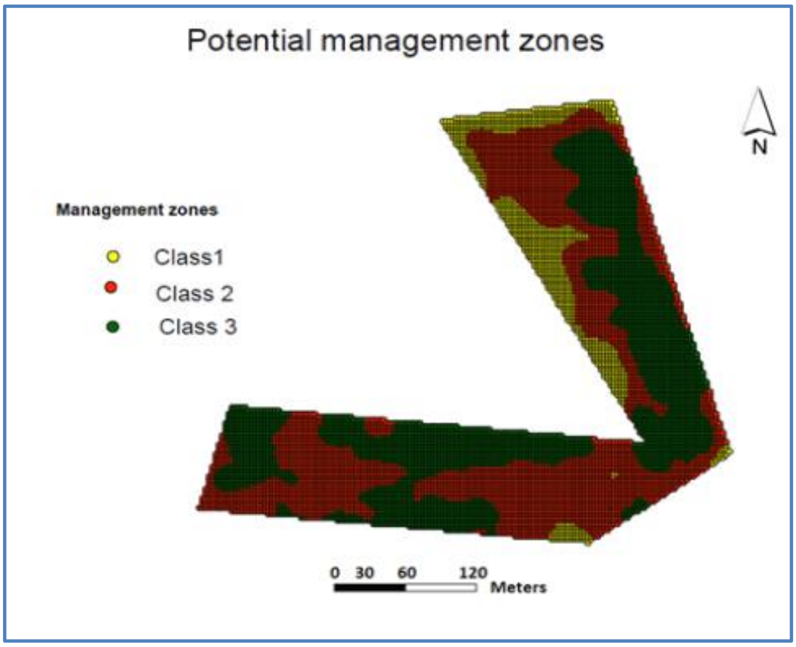


Figure 15. Potential management zones using NDRG, ECa-V5 and ECa-H1 soil variables for field 1 and 2 under scenario I.

4.4.2. Scenario II

Field one: similar to scenario I, three clusters were used to define the potential management zones. NDRG, ECa-V1 and ECa-H1most explaining soil variables were used as input to the k-mean classification. Defined potential management classes are presented in Figure 16. Defined management zones for the current scenario are apparently in different spatial patterns distribution than zones produced in scenario I.

In contrast to scenario I, the defined management zones pattern matches with spatial distribution of ECa-H1. The pattern similarity to some extent extends to NDRG as well. The ECa-V1 soil variable looks poorly coincided in pattern with defined management zones besides the dominance of patchy pattern.

Zone 1, occupied the northern border parts of field and it covers very small portion and of course the zone corresponds to low average potato yield. This class corresponds to high value of ECa-V1 and relatively medium value of ECa-H1 and NDRG. Again the visual spatial pattern similarity between yield map and soil variables is very low.

Zone 2, occupied the eastern part of the field with few isolated islands. This zone coincides with dominantly higher ECa-H1 values and lower NDRG values. The zone also coincides with lower values of ECa-V1. Zone 3, is located in the western part of the field and is a zone with relatively higher average production. This zone corresponds to lower value of ECa-H1. Zone 3 also coincides with relatively higher NDRG and ECa-V1 though patchiness is the barrier in this case.

Field two: Three clusters were also used to define the potential management zones for field two. RmB, ECa-V1 and elevation were used as input to the k-mean classification in this case. Defined potential management classes are presented in Figure 17.

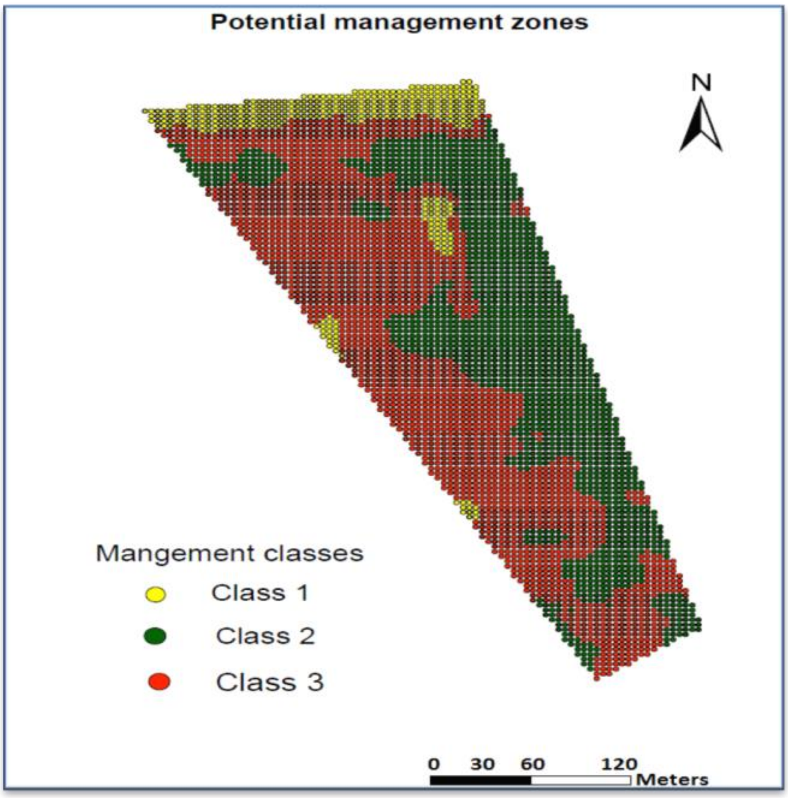


Figure 16. Potential managements zones using ECa-H1, ECa-V1 and NDRG soil variables for field 1 under scenarion II.

The defined management zones pattern seems likely similar to the pattern distribution of kriged RmB soil variable map (Figure 13). The spatial distribution of ECa-V1 map also matches to some extent to the patterns of defined the management zones. Elevation does not match to any spatial patterns of the defined zones.

Zone 1, occupies the northern border parts of field. This area covers relatively small portion of the field in area coverage prospective and it corresponds to low RmB and higher ECa-V1 values. Zone 2 occupies dominantly the western and central parts of the field. This zone corresponds to relative medium RmB values and mixed low and medium ECa-V1 values.

Zone 3, is located in the eastern and western central parts of the field. This zone corresponds to higher RmB values and mixed high and low ECa-V1 values. This zone has high average yield record. In summary the spatial distribution pattern of defined management zone resembles to the spatial distribution of kriged RmB.

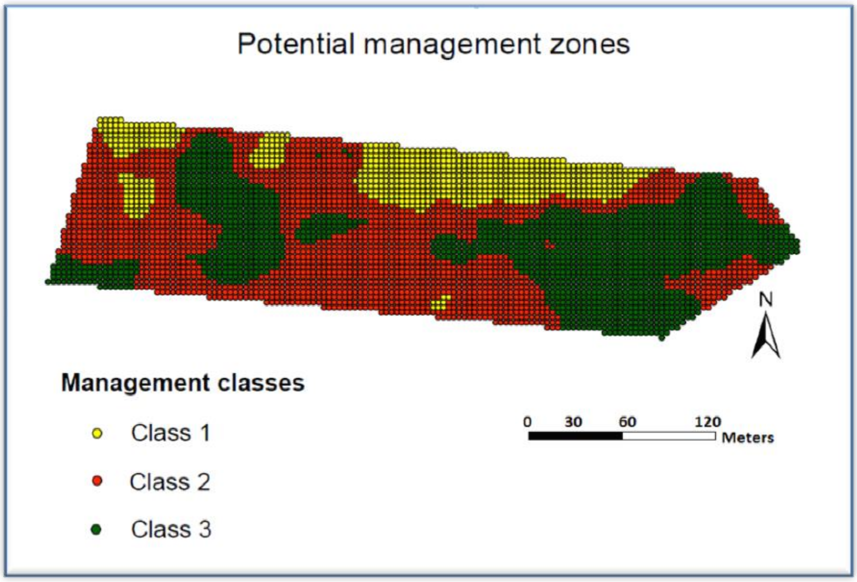


Figure 17. Potential management zones using RmB, ECa-V1 and elevation for field 2 under scenario II.

4.4.3. Scenario III

For this scenario state available elevation and optical soil index were used. The defined management zones showed a regular pattern following the elevation pattern (Figure 18). Elevation does not coincide with any management zones patterns in scenario I and II. However, the created zones in the current scenario followed a similar pattern as the elevation.

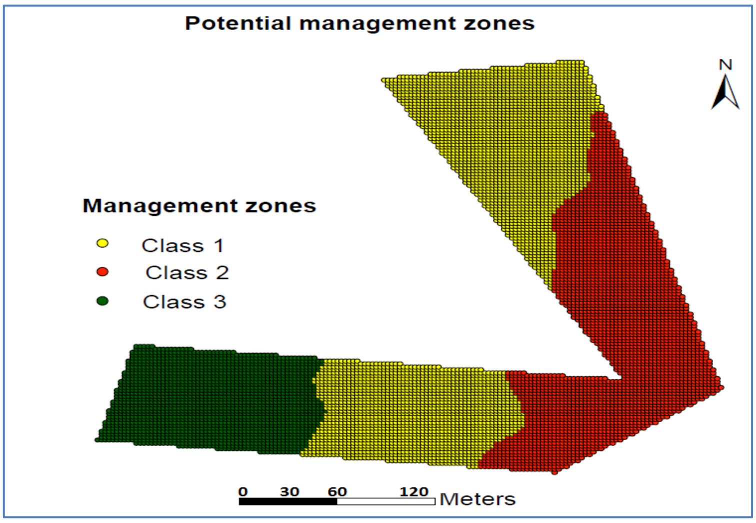


Figure 18. Potential management zones using Elevation and NDRG for field 1 and 2 under scenario III.

4.5. Crop productivity and management zone validation

4.5.1. Potato yield map verse management zones visual comparison

Visual comparison of the defined management zones with one season potato yield distribution patterns might be helpful to assess the goodness or appropriateness of the defined management zones on broad prospective. Ordinary kriging was used to interpolate the yield data in ArcGIS commercial software. Spherical semivariogram model was also used. The kriged map then was classified in to four class intervals. The average yield of the combined adjacent fields was found 62.80 ton/ ha. For visualization purpose separated yield maps for each field was created.

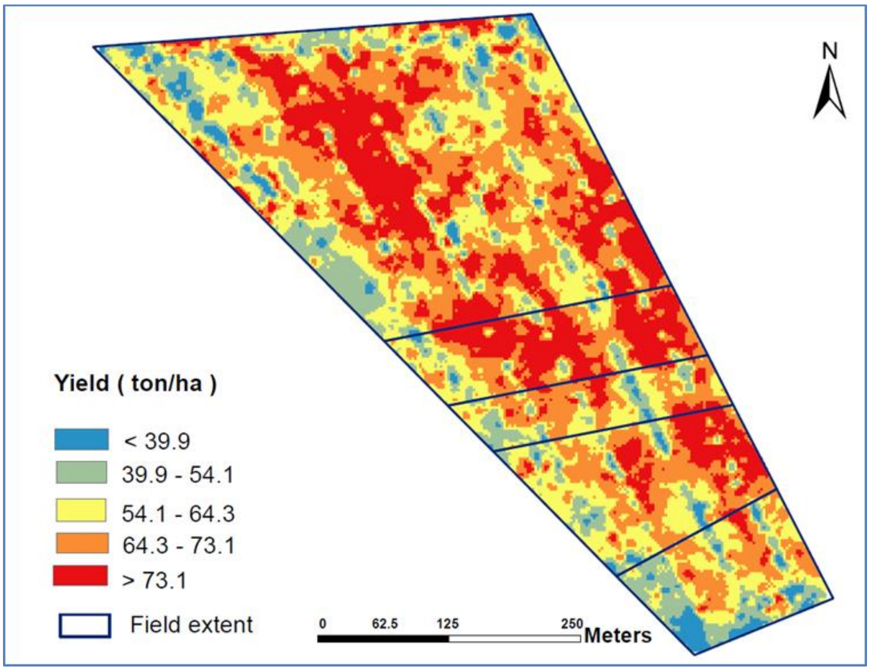


Figure 19. Kriged maps of potato yield for field 1. The lines indicate fertilizer treatments level.

The classified potato yield map of field one and field two are presented in Figure 19 and Figure 20. Poorly distinct pattern similarity between yield maps and management zones (Figure 16 and Figure 17) could be observed in both field one and field two under scenario II. Poor visual similarity patterns between potato yield maps of combined fields and management zones defined for scenario I and scenario III also observed. However, visual comparison between the yield maps and the management zones is not straightforward again. It is not clear to observe any regular structural patterns showing the relationship between defined management zones and kriged yield map. At this point in time, it was found difficult to configure the patterns. Even the yield map itself is full of patchiness except some patterns similarity to some of the soil variables like ECa-V1 for field one and RmB for field two. For instance eastern and central parts of the field one is dominated by higher and the western and eastern parts of field two also corresponds to higher values.

For evident understanding of the patterns among yield map and management zones, georeferenced potato yield measurements were assigned to the defined management zones for each scenario (Fraisse et al., 2001; Van Meirvenne et al., 2013). Mean, standard deviation and standard error descriptive statistics of potato yield measurements for each class under each scenario is presented in section 4.5.2.

4.5.2. Management zones validations using Potato yield

To assess the validity of the method to delineate management zones using sensing based soil variables, the georeferenced potato yield was assigned to the management zones accordingly. After assigning the yield to one of management zones, the attribute data were exported for further analysis of variance. Then statistical analysis to check the statistical mean difference between and among management zones was performed per scenario and the results are presented per scenario.

In addition to potato yield data 22 organic matters samples were also used for validation purpose. The samples points of organic matter were partly located in field one. As far as the samples were

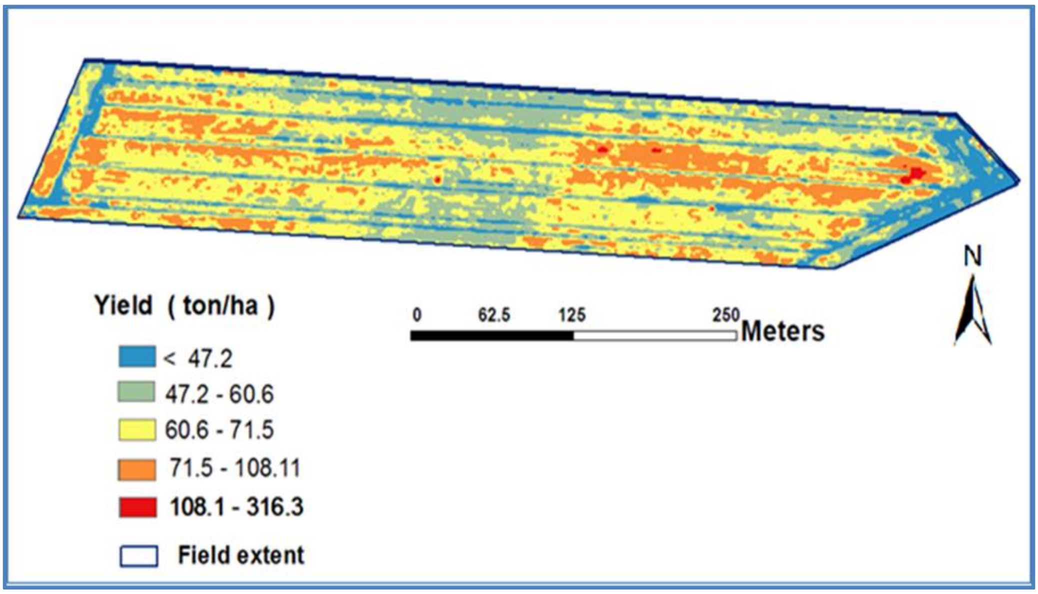


Figure 20. Kriged maps of potato yield for field 2.

located in field one, the validation was done under scenario II of field one and presented under section 4.5.3.

4.5.2.1. Scenario I

Growing season normalized potato yield (ton per hectare) maps are shown in (Figure 19 and Figure 20). As shown from the figures especially for field two there exist strip lines referring to low yield located west-east wards of field two which probably corresponds to the driving or harvesting paths. Similar situations happened north-south wards of field one.

The 53,826 number of points were grouped to three classes according to the management zones. The descriptive statistics of the yield data is presented in Table 17. For the current scenario 62.8 ton/ ha average yield was recorded in the field. Class 3 produced lower average (57.7 ton/ha) potato yield and class 1 produced higher average (64.17 ton/ha) potato yield. Class 2 lies in between with 62.58 ton /ha which is almost very close to the total average mean of the field. The standard deviation around means was also calculated. Higher deviation for class 3 was observed due to the lower number of observations.

Table 17. Descriptive statistics of the potato yield (ton/ha) management zones for scenario I.

Management class	N	Mean	Std. Deviation	Std. Error
1	25417	64.17	19.179	0.120
2	22566	62.58	16.473	0.110
3	5843	57.70	20.707	0.271
Total	53826	62.80	18.377	0.079

A one-way analysis of variance was conducted to evaluate, if there is significant difference level of yield (ton/ ha) mean variance among and between the management classes**.** Descriptive statistics of potato yield assigned for each management zone (Table 17).

The assumption of homogeneity of means variance was tested. The ANOVA was found significant with F (2, 53823) = 301.294, p = 0.000 (Table 18). The p value in this case is by far smaller than 0.05. The values (2, 53823) refers to degree of freedom between classes and degree of within classes respectively. Thus there is evidence that there is significant difference of mean variance among the three classes of the potato yield though the mean difference is lower.

Table 18. ANOVA output for comparison of yield variation between management zones for scenario I.

	Sum of squares	Degree of freedom	Mean square	F-statistics	Significance level
Between groups	201262.786	2	100631.393	301.294	0.000
Within groups	17976719.767	53823	333.997		
Total	18177982.554	53825			

ANOVA gives the overall mean variance difference. Post Hoc comparison to evaluate the pairwise difference among group means for all scenarios were conducted using of Tukey HSD test since equal variance was tenable. Tests for scenario I and scenario II (both field 1 and field 2) revealed a significant pairwise difference between means scores among the three classes of potato yield for each scenario. There is significant pairwise difference among the means scores of the three potato yield combination classes with $p < 0.05$. However, Post Hoc comparison test were not conducted for scenario three since there was no significant mean variance difference among the three classes potato yield.

4.5.2.2. Scenario II

Field one: the mean, standard error and standard deviation distribution of potato yield for management classes of field one are presented in Table 19. Class 3 showed higher (66.0 ton/ha) average yield per hectare while class1showed lower mean averages (61.1 ton/ha) for field one.

A one-way analysis of variance was conducted to evaluate if there is significant mean difference level of potato yield (ton/ha) among the management classes of field.

Table 19. Descriptive statistics of the potato yield (ton/ha) management zones for field 1 under scenario II.

Management class	N	Mean	Std. Deviation	Std. Error
1	2697	61.1	22.9	0.44
2	14582	63.0	17.8	0.14
3	9818	66.0	16.6	0.16
Total	29832	63.9	18.1	0.10

In similar way, the assumption of homogeneity of means variance was tested. The ANOVA was found significant F (2, 29829) = 172.5, p = 0.000 (Table 20). The *p* value in this case is again by far smaller than 0.05. Thus there is evidence that there is significant difference of mean variance among the three classes of the potato yield though the mean difference is lower.

Table 20. ANOVA output for comparison of yield variation between management zones for field 1 under scenario II.

	Sum of squares	Degree of freedom	Mean square	F-statistics	Significance level
Between groups	76187.5	2	38093.7	117.4	0.000
Within groups	8791390.3	27094	324.4		
Total	8867577.9	27096			

Field Two:

Descriptive statistics of the three management yield classes are presented in Table 21.

One-way analysis of variance was conducted to evaluate, if there is significant difference level of yield (ton/ha) mean variance potato yield among the management classes of field two.

Table 21. Descriptive statistics of the potato yield (ton/ha) management zones for field 2 under scenario II.

Management class	N	Mean	Std. Deviation	Std. Error
1	15638	62.0494	17.43419	0.13942
2	10048	64.0557	20.67451	0.20625
3	4146	57.7953	14.85864	0.23076
Total	29832	62.1339	18.38007	0.10642

In similar manner, the assumption of homogeneity of means variance was tested. The ANOVA was found significant F (2, 29829) = 172.5, *p* = 0.000 (Table 22). The p value in this case is by far smaller than 0.05. Thus there is significant evidence that there is significant difference of mean variance among the three classes of the potato yield.

Table 22. ANOVA output for comparison of yield variation between management zones for field 2 under scenario II.

	Sum of squares	Degree of freedom	Mean square	F-statistics	Significance level
Between groups	115264.4	2	57632.2	172.5	0.000
Within groups	9962452.8	29829	333.9		
Total	10077717.2	29831			

4.5.2.3. Scenario III

Descriptive statistics of the three management classes are presented in Table 23. It was observed that the average mean of the three management classes is very close to the overall mean variance. In addition the standard deviation of the classes is comparable.

Table 23.Descriptive statistics of the potato yield (ton/ha) management zones for scenario III.

Management class	N	Mean	Std. Deviation	Std. Error
1	23694	62.96	17.78	0.115
2	10747	62.97	17.61	0.169
3	19397	62.51	19.47	0.139
Total	53838	62.804	18.3770	0.079

A one-way analysis of variance was conducted to evaluate, if there is significant difference level of yield (ton/ha) mean variance potato yield among the management classes.

In similar manner, the assumption of homogeneity of means variance was tested. The ANOVA was found not significant $F(2, 53835) = 3.84$, $p = 0.521$ (Table 24). The *p* value in this case is greater than 0.05. Thus there is no significant evidence of significant difference of mean variance among the three classes and homogeneity of the management classes is accepted.

Table 24. ANOVA output for comparison of yield variation between management zones for scenario III

	Sum of squares	Degree of freedom	Mean square	F-statistics	Significance level
Between groups	2594.7	2	1297.36	3.84	**0.521**
Within groups	18179008.7	53835	337.68		
Total	18181603.4	53837			

4.5.3. Management zone validation using organic matter

Half of the organic matter sample points were located in zone 2 and the remaining half in zone 3 of the field one (Appendix 9). Analysis of ANOVA then was conducted to examine if significant mean variance of organic matter exist between the two zones.

Table 25 . Descriptive statistics of Organic matter management zones for field 1under scenario II.

Management class	N	Mean	Std. Deviation	Std. Error
2	11	3.19	0.54	0.16
3	11	3.75	0.34	0.10
Total	22	3.47	0.53	0.11

Similarly to potato yield, one-way analysis of variance was conducted to evaluate if there is significant difference level of organic matter mean variance between the two classes.

The assumption of homogeneity of means variance was tested. The ANOVA was found significant F (1, 20) = 8.328, *p* = 0.009 (Table 26). Thus there is evidence that there is significant difference of mean variance among the two classes of organic matter.

Table 26. ANOVA output for comparison of organic matter variation between management zones for field 1under scenario II.

	Sum of squares	Degree of freedom	Mean square	F-statistics	Significance level
Between groups	1.747	1	1.747	8.328	0.009
Within groups	4.196	20	0.21		
Total	5.944	21			

Chapter 5. Discussion

This section explores discussions on data sources, methods, and own findings versus finding of previous studies from literature. In addition a quality assessment of the potential management zones in the context of precision agriculture prospective is made.

RQ1: Which sensing based soil variables describes most variations of soils within fields ?

EMI measurements, pH value, colour aerial photograph and elevation sensing based data sources were evaluated in this study. Potato yield from one growing season was used for validation on the delineated management zones. The management zones produced for this study probably might change by using data source from multiple growing seasons due to climatic factors which in return could influence soil properties like soil moisture content and others related properties. Regarding this (Vitharana et al., 2008) found that crop production in the valley floor is likely to be more variable between years than the management classes. Moreover, (Kaspar et al., 2003) added corn yield is negatively correlated with elevation when precipitation is less than the normal growing season and yield is positively correlated with elevation for above normal precipitation season.

PCA was performed on the soil variables for each scenario to identify soil variables explaining most variation. Key soil variables were then selected for each scenario accordingly and are presented in Table 27. Significance level of the ANOVA F-test results, *p* values and percent of total cumulative variance explained are also presented in the same table. Each of identified soil variables explaining most variations for each scenario, percent of total cumulative variance explained, F-test results of each scenario and the *P* values of ANOVA for each scenario are summarized in Table 27.

Beside selection of variables, assessing the relationship of the selected variables to the yield is an important issue. A correlation matrix was performed to evaluate the relationship between potato yield and selected soil variables. ECa-H1 was found significantly influencing crop production of the growing season of the year at 0.05 level of significance. Other selected soil variables were not found significantly influencing yield productivity for the specific growing season of the year.

According to (Van Meirvenne et al., 2013), ECa-H.5, elevation and pH value were selected as key soil parameters on sandy loam soil type. The soil pH in their study was obtained from soil samples and analysed in soil laboratory. However, pH value was not selected as the key variables in any of the scenario in the current study. The variance of pH value was even higher for the current study ranging from 6.26 to 8.77 while it ranges from 4.6 to 5.6 for the previous study. Though the larger the variance, it might not contribute for explaining the variation that exist within the soil in the current fields. In addition the low correlation value between pH and yield. Moreover, an earlier study of (Fraisse et al., 1999) identified soil EC, elevation, and slope as most useful attributes for the delineation of management zones. Two years later, the same authors acknowledged elevation and bulk electrical conductivity as more important attributes than slope and compound topographic index to define management zone for crop yield.

When both field one and field two were combined for PCA in scenario I: NDRG, ECa-H1 and ECa-V1 were identified as key soil variables. However, when the two fields are treated separately under scenario II, different soil variables were selected (Table 27). This selection difference might be due to spatial scale or spatial coverage concept. So spatial scale might be another determining factor which it might influences the delineation and spatial continuity of management zones.

Table 27. Selected soil variables in order of contribution with respective percent of total cumulative variance explained in each scenario. Significance F-test and p-values of ANOVA test are also summarized.

Scenario	Selected key soil variables	% total cumulative variance explained	Significance F-test	p value ANOVA
I	NDRG ECa-H1 ECa-V1	70	Significant	0.000
II - field 1	NDRG ECa-V1 ECa-H1	72.87	Significant	0.000
II - field 2	R-B ECa-V1 Elevation	69.35	Significant	0.000
III	NDRG Elevation	87.1	Not significant	0.521

Beside the selection of different soil variables, the change of patterns of the defined management zones can easily be recognized during the change of spatial scale in scenario I and scenario II (Figure 15, Figure 16, Figure 17). On the other hand elevation and NDRG were identified for scenario III. These two variables defined management zones which basically does n't not reflect spatial continuity and coherence. Patterns of the defined management zones roughly match with elevation in this respect.

To sum up, ten sensing based soil variables were used for this study. The explanatory contribution potential of soil variables might change with spatial scale change as observed in scenario I and scenario II. This might be due to the spatial variations that changes with spatial coverage of fields which actually is the case in most situations. Beside the expected change which comes due change of input data sources, use of same data sources might end up with different key soil variables to explaining most of the soil variations.

RQ2: How to define management classes in precision agricultural farm using soil variables?

The number of management zones in a field depends on a number of factors. Level of variability in the field, variables type (stable or dynamic soil variables), weather and crop type are some of them (Patabendige et al., 2003). Moreover, Li et al. (2007) added the number of management zones depends on measurement sensitivity and intra-field variability. For this current study, three management zones were considered most convenient for each scenario based on separability and overlap of the management classes plots. Better separation and less overlap among the management classes were found when three classes are used. Spatial structure of defined compact management zones for scenario I and II to some extent resembles to the spatial distribution patterns of ECa's soil variables. However, the spatial continuity of defined management zone for scenario III does not match with any of the kriged soil property maps except elevation.

The spatial patterns and distribution of the defined management zones were found different for the each scenario. The defined management zones for scenario I and III are dissimilar in their spatial coherences which might be triggered by the use of different dataset (Figure 15 and Figure 18). Defined management zones for scenario I and II seems not spatially coherent. This spatial pattern dissimilarity might be mainly due to the spatial scale taken in to account as same dataset was used. The change in spatial scale for scenario I and II resulted in selection of different soil variables though the same dataset were used for PCA. Even though the influence of spatial scale on PCA seems not explored Demšar et al. (2013) reviewed how PCA can be used to investigate multiple scales of spatial autocorrelation. Management zones complexity might also increase with increasing geographical scale. Related to this Novembre and Stephens (2008) described PCA applied to spatial data decays with geographical distance. The spatial scale in this regard refers to the spatial area coverage.

Therefore, k-mean clustering based on sensing based soil variables could benefit agriculture system analysis by increasing crop productivity and of course improve environmental quality. Profitability could be maximized through implementation of potential homogenous management zones making sure that the variations within-zones are explored. It is believed that management zones for site-specific management zones should be simple, functional and economical feasible (Fleming et al., 2000; Koch et al., 2004; Taylor et al., 2007). In summary farmers or growers should notice that different management classes could be defined depending on the data source used, spatial scale and of course purpose of the zone delineation.

RQ3: How to evaluate the potential performance of the method used to define management classes?

Spatial analysis of variance showed that means of the defined management zones for scenario III were not significantly different (Table 24) which might be confirmation to the dissimilarity of the defined management zones pattern to majority of soil properties maps. So state available geo-data of elevation and colour aerial photographs data sources resulted in poorly defined and poorly differentiated management zones in which no significant potato yield variation among management zones was found for the specific growing season. This might lead to the question which data source to use to define management zones particularly in the study field. But the mean variance of defined management zones in the other two scenarios (I and II) were found significant both within group means and among group means. Moreover, partly geo-located organic matter samples on field one under scenario II confirmed distinctness of the two zones (Table 26). A significant mean difference of organic matter was found between the two zones.

Moreover, it is also important to take in to account the total cumulative percentages of soil variations explained by the key identified soil variables. For scenario I and II (Table 7) about 70% of the data variation is explained where as 87 % of variation was explained for scenario III (Table 11). The lower percent of variation explained in scenario I and II might contribute for the uncertainty of defined management zone as far as some unexplained variation left in the data. So the goodness of the management zones might also be affected by the percent of the variations explained though no reference was found in my search for it.

Applicability of the method for management zone delineation

The statistical analysis of the this study showed that the defined management zones had different mean yield and organic matter for scenario I and II signifying that the approach implemented might be applicable for the study field. The procedure might be effective in identifying feasible management zones in precision farming operations. Besides the applicability of the approach, in the current advancement of precision farming, spatially detailed ECa with EM38-MK2 (McBratney et al., 2000; Moral et al., 2010; Van Meirvenne et al., 2013), elevation with airborne laser scanning (LiDAR) (Vlaanderen 2003) and colour aerial photograph with Unmanned Aerial Vehicles (Bartholomeus et al. 2011) sensing based soil variables can be recorded for use in the approach.

Furthermore, thinking of management zone delineation, identifying potential sources of variation or potential yield limiting factors is most important. Then the source of variations might be stable over time (e.g. elevation) or dynamic change from year to year due to weather conditions and other factors (e.g. ECa, soil moisture). In line with this (Fridgen et al., 2000) described the appropriate number of zones to use when dividing a field may vary between years and is often dependent on the weather and the crop type to be planted. The authors also added year to year variation of appropriate number of management zones was attributed to weather and crop type. The number of management zones can be decreased if water stress tolerant crops are planted as described by Fraisse et al. (2001). Beside the weather conditions and crop type, soil type and paedogenesis underlying process which may trigger a variation in soil properties in a time frame needs to be considered.

The importance of data sources used, as indicated in section 4.5.1 there exist pattern similarity between created management zones and ECa's measurements. In addition, management zones defined using contribution of ECa measurement as key parameters (scenario I and II) performed well in terms of crop production. This is due to the fact that proximally sensed ECa is a vital integrator of the bulk soil property (Corwin and Lesch, 2005; Saey et al., 2009). Vitharana et al. (2006) added electromagnetically sensed ECa is a promising and cost-effective source of ancillary information for detailed mapping of the heterogeneous subsoil.

Regarding the operational application of the defined management zones for precision agriculture, the method might be effective in identifying operable management zones at field level using sensing based data sources: apparent electrical conductivity, elevation and optical soil indices. However, weather changes and crop type should be considered while implementing the method. The method should also be validated by fully covered spatial ground truthing data before implementation to check the goodness consistency of the method.

Chapter 6. Conclusion

Recent advancements in precision farming technologies help to define potential management zone which can be used for variable rate input application. In the current study ECa measurements, topographic elevation and optical soil index parameters contributed to explain soil and yield variation for the considered scenarios while pH had only limited power to explain soil variations.

Different key soil variables were selected although same data source were used for PCA technique. Analysis at field level (scenario II) and cluster of fields together (scenario I) were performed and different key soil variables were identified accordingly. At this time field level analysis would be advised as it exploits the variations that exist in the fields.

Geostatistical interpolation techniques were used to investigate the spatial coherence and distribution of selected soil key variables and subsequently ordinary kriginged soil property maps were clustered using *k-mean* clustering algorithm. Three management zones were defined per scenario based on separability and overlap. Management zones showed spatial patterns similarity to the key soil variables identified in scenario I and II. The boundary of these defined management zones depends on temporal weather variations and crop specific plantation.

The spatial analysis of variance and goodness of created management zones were validated through a comparison with field measured potato yield. The combination of ECa measurements and optical soil index parameters for scenario I and II produced significantly distinct management zones. However, topographic elevation and optical soil index parameters for scenario III created not separable management zones. It can be concluded that the combination of topographic elevation and soil index parameters was not capable of producing good performing implementable distinct management zones in this current study.

To the end defined management zones should be easy for real operation and practical implementation for the farm manager or grower. After all the application of such an approach should be environmentally sustainable and improve productivity.

.

Chapter 7. Recommendations

Based on the results of the study, some recommendations are suggested for delineation of management zones as indicated below:

- Growers or farmers should delineate management zones at field level as it exploits site specific or local variations.
- During delineation of management zones mapping or spatial scale for variable input rate application needs due attention.
- Availability of technological advancements should be considered while delineating management zones.
- Once management zones are delineated, stability or dynamicity of the zones with weather or crop change should be considered as management zones are site and crop-specific. So updating management zones and use of less temporally varying soil properties might be more feasible.
- The process of potential management zone delineation should be simple, practicable and of course exploit the within field variation as much as possible.

Chapter 8. Reference

Abdi, H., & Williams, L.J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics, 2*, 433-459

Adamchuk, V., Lund, E., Sethuramasamyraja, B., Morgan, M., Dobermann, A., & Marx, D. (2005). Direct measurement of soil chemical properties on-the-go using ion-selective electrodes. *Computers and Electronics in Agriculture, 48*, 272-294

Adamchuk, V., Viscarra Rossel, R., Glinski, J., Horabik, J., & Lipiec, J. (2011a). Precision agriculture: proximal soil sensing. *Encyclopedia of Agrophysics, Springer, New York, New York*, 650-656

Adamchuk, V.I., Rossel, R.A.V., Sudduth, K.A., & Lammers, P.S. (2011b). Sensor fusion for precision agriculture

Al-Gaadi, K.A. (2012). employing electromagnetic induction technique for the assessment of soil compaction. *American Journal of Agricultural & Biological Science, 7*

Auernhammer, H. (2001). Precision farming—the environmental challenge. *Computers and Electronics in Agriculture, 30*, 31-43

Banerjee, S., Gelfand, A.E., & Carlin, B.P. (2004). *Hierarchical modeling and analysis for spatial data*. Crc Press

Bartholomeus, H., & Kooistra, L. (2012). Use of aerial photographs for assessment of soil organic carbon and delineation of agricultural management zones. In, *EGU General Assembly Conference Abstracts* (p. 2509)

Bartholomeus, H., Kooistra, L., Stevens, A., van Leeuwen, M., van Wesemael, B., Ben-Dor, E., & Tychon, B. (2011). Soil organic carbon mapping of partially vegetated agricultural fields with imaging spectroscopy. *International Journal of Applied Earth Observation and Geoinformation, 13*, 81-88

Batte, M.T., & Arnholt, M.W. (2003). Precision farming adoption and use in Ohio: case studies of six leading-edge adopters. *Computers and Electronics in Agriculture, 38*, 125-139

Beaumont, R. 1. Learning outcomes

Bianchini, A.A., & Mallarino, A.P. (2002). Soil-sampling alternatives and variable-rate liming for a soybean–corn rotation. *Agronomy Journal, 94*, 1355-1366

Bongiovanni, R., & Lowenberg-Deboer, J. (2004). Precision agriculture and sustainability. *Precision Agriculture, 5*, 359-387

Boydell, B., & McBratney, A. (2002). Identifying potential within-field management zones from cotton-yield estimates. *Precision Agriculture, 3*, 9-23

Brock, A., Brouder, S., Blumhoff, G., & Hofmann, B. (2005). Defining yield-based management zones for corn–soybean rotations. *Agronomy Journal, 97*, 1115-1128

Brouder, S., Hofmann, B., & Morris, D. (2005). Mapping Soil pH. *Soil Science Society of America Journal, 69*, 427-442

Catalano, M. (2011). Measuring Soil Conductivity with Geonics Limited Electromagnetic Geophysical Instrumentation: Proximal soil sensing In

Cattell, R.B. (1966). The scree test for the number of factors. *Multivariate behavioral research, 1*, 245-276

Chang, C.-W., & Laird, D.A. (2002). Near-infrared reflectance spectroscopic analysis of soil C and N. *Soil Science, 167*, 110-116

Christy, C.D. (2008). Real-time measurement of soil attributes using on-the-go near infrared reflectance spectroscopy. *Computers and Electronics in Agriculture, 61*, 10-19

Corwin, D., & Lesch, S. (2003). Application of soil electrical conductivity to precision agriculture. *Agronomy Journal, 95*, 455-471

Corwin, D., & Lesch, S. (2005). Apparent soil electrical conductivity measurements in agriculture. *Computers and Electronics in Agriculture, 46*, 11-43

Cox, S. (2002). Information technology: the global key to precision agriculture and sustainability. *Computers and Electronics in Agriculture, 36*, 93-111

Crookston, R.K. (1996). Using decision cases to enhance technology transfer in precision agriculture. *Precision Agriculture*, 1117-1122

Davis, J.G., Kitchen, N.R., Sudduth, K.A., & Drummond, S.T. (1997). Using electromagnetic induction to characterize soils. *Better Crops, 81*, 6-8

De Gruijter, J., & McBratney, A. (1988). A modified fuzzy k-means method for predictive classification. *Classification and related methods of data analysis, 97*, 104

Demšar, U., Harris, P., Brunsdon, C., Fotheringham, A.S., & McLoone, S. (2013). Principal component analysis on spatial data: an overview. *Annals of the Association of American Geographers, 103,* 106-128

Doerge, T. (1999). Defining management zones for precision farming. *Crop Insights, 8,* 1-5

Doerge, T.A. (2009). Site-Specific Management Guidelines (SSGM-2). *Management Zone Concepts*

Farahani, H., & Buchleiter, G. (2004). Temporal stability of soil electrical conductivity in irrigated sandy fields in Colorado. *Transactions of the ASAE, 47,* 79-90

Fleming, K., Westfall, D., Wiens, D., & Brodahl, M. (2000). Evaluating farmer defined management zone maps for variable rate fertilizer application. *Precision Agriculture, 2*, 201-215

Fraisse, C., Sudduth, K., & Kitchen, N. (2001). Delineation of site-specific management zones by unsupervised classification of topographic attributes and soil electrical conductivity. *Transactions of the ASAE, 44*, 155-166

Fraisse, C., Sudduth, K., Kitchen, N., & Fridgen, J. (1999). Use of unsupervised clustering algorithms for delineating within-field management zones. *St. Joseph: American Society of Agricultural Engineers*

Frazier, B., Walters, C., Perry, E., Pierce, F., & Sadler, E. (1997). Role of remote sensing in site-specific management. *The state of site specific management for agriculture.*, 149-160

Fridgen, J., Fraisse, C., Kitchen, N., & Sudduth, K. (2000). Delineation and analysis of site-specific management zones. In, *Proceedings of the 2nd International Conference on Geospatial Information in Agriculture and Forestry* (pp. 402-411): Citeseer

Gaston, L., Locke, M., Zablotowicz, R., & Reddy, K. (2001). Spatial variability of soil properties and weed populations in the Mississippi Delta. *Soil Science Society of America Journal, 65*, 449-459

Gebbers, R., & Adamchuk, V.I. (2010). Precision agriculture and food security. *Science, 327*, 828-831

Gibbons, G. (2000). Turning a Farm Art into Science An Overview of Precision Farming. *URL: http://www. precisionfarming. com*

Godwin, R., Wood, G., Taylor, J., Knight, S., & Welsh, J. (2003). Precision farming of cereal crops: a review of a six year experiment to develop management guidelines. *Biosystems Engineering, 84*, 375-391

Jaynes, D., Kaspar, T., Colvin, T., & James, D. (2003). Cluster analysis of spatiotemporal corn yield patterns in an Iowa field. *Agronomy Journal, 95*, 574-586

Johnson, R.A., & Wichern, D.W. (2002). *Applied multivariate statistical analysis*. Prentice hall Upper Saddle River, NJ

Jolliffe, I. (2005). *Principal component analysis*. Wiley Online Library

Kaspar, T.C., Colvin, T.S., Jaynes, D.B., Karlen, D.L., James, D.E., Meek, D.W., Pulido, D., & Butler, H. (2003). Relationship between six years of corn yields and terrain attributes. *Precision Agriculture, 4*, 87-101

Khakural, B., Robert, P., & Hugins, D. (1998). Use of non-contacting electromagnetic inductive method for estimating soil moisture across a landscape. *Communications in Soil Science & Plant Analysis, 29*, 2055-2065

King, J., Dampney, P., Lark, R., Wheeler, H., Bradley, R., & Mayr, T. (2005). Mapping potential crop management zones within fields: use of yield-map series and patterns of soil physical properties identified by electromagnetic induction sensing. *Precision Agriculture, 6*, 167-181

Kitchen, N., Sudduth, K., Myers, D., Drummond, S., & Hong, S. (2005). Delineating productivity zones on claypan soil fields using apparent soil electrical conductivity. *Computers and Electronics in Agriculture, 46*, 285-308

Koch, B., Khosla, R., Frasier, W., Westfall, D., & Inman, D. (2004). Economic feasibility of variable-rate nitrogen application utilizing site-specific management zones. *Agronomy Journal, 96*, 1572-1580

Kuang, B., Mahmood, H.S., Quraishi, M.Z., Hoogmoed, W.B., Mouazen, A.M., & van Henten, E.J. (2012). 4 Sensing Soil Properties in the Laboratory, In Situ, and On-Line: A Review. *Advances in Agronomy, 114*, 155-223

Kuang, B., & Mouazen, A.M. (2013). Effect of spiking strategy and ratio on calibration of on-line visible and near infrared soil sensor for measurement in European farms. *Soil and Tillage Research, 128*, 125-136

Lamb, D., & Brown, R. (2001). PA—Precision Agriculture: Remote-Sensing and Mapping of Weeds in Crops. *Journal of Agricultural Engineering Research, 78*, 117-125

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. *Agronomy Journal, 97*, 524-532

Li, Y., Shi, Z., Li, F., & Li, H.-Y. (2007). Delineation of site-specific management zones using fuzzy clustering analysis in a coastal saline land. *Computers and Electronics in Agriculture, 56*, 174-186

Liu, T.-L., Juang, K.-W., & Lee, D.-Y. (2006). Interpolating soil properties using kriging combined with categorical information of soil maps. *Soil Science Society of America Journal, 70*, 1200-1209

Maohua, W. (2001). Possible adoption of precision agriculture for developing countries at the threshold of the new millennium. *Computers and Electronics in Agriculture, 30*, 45-50

McBratney, A., & Pringle, M. (1999). Estimating average and proportional variograms of soil properties and their potential use in precision agriculture. *Precision agriculture, 1*, 125-152

McBratney, A., Whelan, B., Ancev, T., & Bouma, J. (2005). Future directions of precision agriculture. *Precision Agriculture, 6*, 7-23

McBratney, A.B., Mendonça Santos, M.d.L., & Minasny, B. (2003). On digital soil mapping. *Geoderma, 117*, 3-52

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

McNeill, J. (1980). Electromagnetic terrain conductivity measurement at low induction numbers. In: Geonics Limited Ontario, Canada

Molin, J.P., & Castro, C.N.d. (2008). Establishing management zones using soil electrical conductivity and other soil properties by the fuzzy clustering technique. *Scientia Agricola, 65*, 567-573

Moral, F., Terrón, J., & Silva, J. (2010). Delineation of management zones using mobile measurements of soil apparent electrical conductivity and multivariate geostatistical techniques. *Soil and Tillage Research, 106*, 335-343

Morari, F., Castrignanò, A., & Pagliarin, C. (2009). Application of multivariate geostatistics in delineating management zones within a gravelly vineyard using geo-electrical sensors. *Computers and Electronics in Agriculture, 68*, 97-107

Mouazen, A.M., De Baerdemaeker, J., & Ramon, H. (2005). Towards development of on-line soil moisture content sensor using a fibre-type NIR spectrophotometer. *Soil and Tillage Research, 80*, 171-183

Mulder, V., De Bruin, S., Schaepman, M., & Mayr, T. (2011). The use of remote sensing in soil and terrain mapping—A review. *Geoderma, 162*, 1-19

Muñoz, J.D., & Kravchenko, A. (2011). Soil carbon mapping using on-the-go near infrared spectroscopy, topography and aerial photographs. *Geoderma, 166*, 102-110

Mzuku, M., Khosla, R., Reich, R., Inman, D., Smith, F., & MacDonald, L. (2005). Spatial variability of measured soil properties across site-specific management zones. *Soil Science Society of America Journal, 69*, 1572-1579

Novembre, J., & Stephens, M. (2008). Interpreting principal component analyses of spatial population genetic variation. *Nature genetics, 40*, 646-649

Odeh, I., Chittleborough, D., & McBratney, A. (1992). Soil pattern recognition with fuzzy-c-means: application to classification and soil-landform interrelationships. *Soil Science Society of America Journal, 56*, 505-516

Ortega, J., Foster, W., & Ortega, R. (2002). definicion de sub-rodales para una silvicultura de preci-sion: una aplicacion del metodo fuzzy k-means. *Ciencia e Investigación Agraria, 29*, 35-44

Ortega, R.A., & Santibáñez, O.A. (2007). Determination of management zones in corn (< i> Zea mays</i> L.) based on soil fertility. *Computers and Electronics in Agriculture, 58*, 49-59

Patabendige, D.M., Wong, M., & Bowden, B. (2003). Zone management in precision agriculture by matching fertiliser input to crop demand

Pierce, F., & Elliott, T. (2008). Regional and on-farm wireless sensor networks for agricultural systems in Eastern Washington. *Computers and electronics in agriculture, 61*, 32-43

Pierce, F.J., Nowak, P., & Donald, L.S. (1999). Aspects of Precision Agriculture. *Advances in Agronomy* (pp. 1-85): Academic Press

Plant, R.E. (2001). Site-specific management: the application of information technology to crop production. *Computers and Electronics in Agriculture, 30*, 9-29

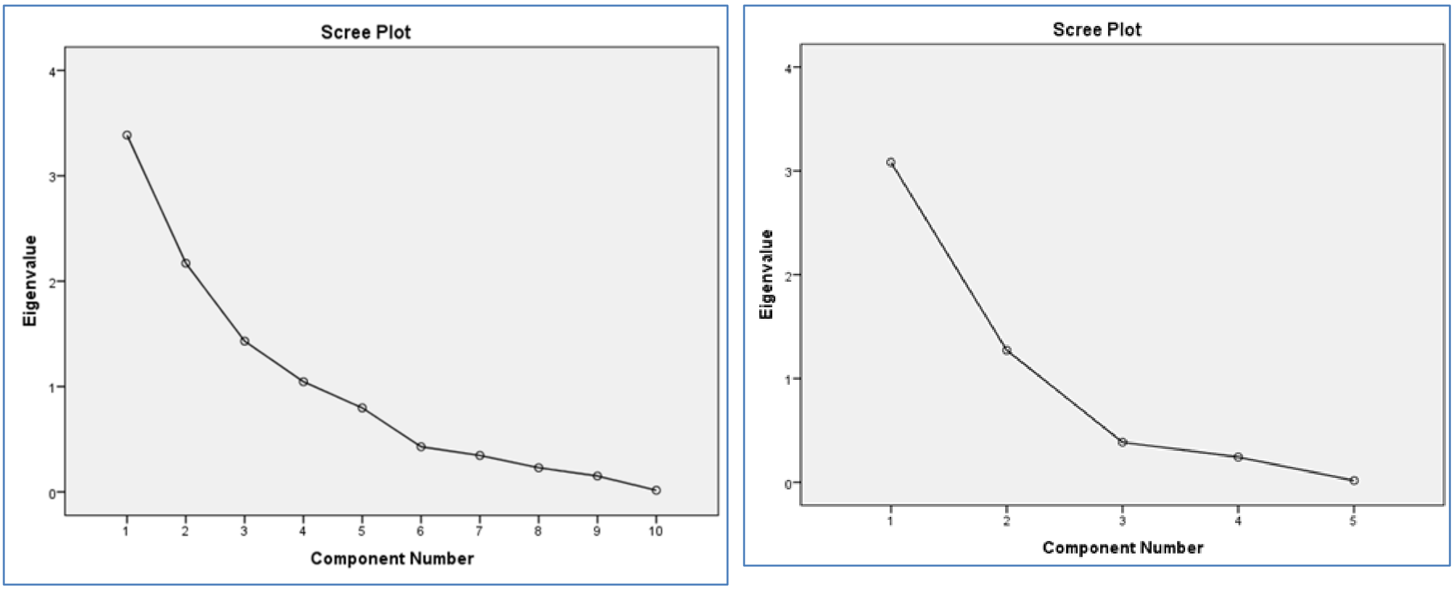
Reyniers, M., Maertens, K., Vrindts, E., & De Baerdemaeker, J. (2006). Yield variability related to landscape properties of a loamy soil in central Belgium. *Soil and Tillage Research, 88*, 262-273
Rhoades, J. (1993). Electrical conductivity methods for measuring and mapping soil salinity. *Advances in Agronomy, 49*, 201-251
Rossel, R.V., Adamchuk, V., Sudduth, K., McKenzie, N., & Lobsey, C. (2011). Proximal soil sensing: an effective approach for soil measurements in space and time. *Advances in agronomy, 113*, 237-282
Saey, T., Van Meirvenne, M., Vermeersch, H., Ameloot, N., & Cockx, L. (2009). A pedotransfer function to evaluate the soil profile textural heterogeneity using proximally sensed apparent electrical conductivity. *Geoderma, 150*, 389-395
Schepers, A.R., Shanahan, J.F., Liebig, M.A., Schepers, J.S., Johnson, S.H., & Luchiari, A. (2004). Appropriateness of management zones for characterizing spatial variability of soil properties and irrigated corn yields across years. *Agronomy Journal, 96*, 195-203
Schepers, J., Schlemmer, M., & Ferguson, R. (2000). Site-specific considerations for managing phosphorus. *Journal of environmental quality, 29*, 125-130
Shaner, D., Khosla, R., Brodahl, M., Buchleiter, G., & Farahani, H. (2008). How well does zone sampling based on soil electrical conductivity maps represent soil variability? *Agronomy Journal, 100*, 1472-1480
Sheets, K.R., & Hendrickx, J.M. (1995). Noninvasive soil water content measurement using electromagnetic induction. *Water Resources Research, 31*, 2401-2409
Shonk, J. (1991). Spectroscopic sensing of soil organic matter content. *T. ASAE, 34*, 1978-1984
Song, X., Wang, J., Huang, W., Liu, L., Yan, G., & Pu, R. (2009). The delineation of agricultural management zones with high resolution remotely sensed data. *Precision Agriculture, 10*, 471-487
Stafford, J.V., Lark, R.M., & Bolam, H.C. (1999). Using yield maps to regionalize fields into potential management units. *Precision Agriculture*, 225-237
Stenberg, B., Viscarra Rossel, R.A., Mouazen, A.M., & Wetterlind, J. (2010). Chapter Five-Visible and Near Infrared Spectroscopy in Soil Science. *Advances in agronomy, 107*, 163-215
Sudduth, K., Drummond, S., & Kitchen, N. (2001). Accuracy issues in electromagnetic induction sensing of soil electrical conductivity for precision agriculture. *Computers and Electronics in Agriculture, 31*, 239-264
Sudduth, K., Kitchen, N., Bollero, G., Bullock, D., & Wiebold, W. (2003). Comparison of electromagnetic induction and direct sensing of soil electrical conductivity. *Agronomy Journal, 95*, 472-482
Sudduth, K., Kitchen, N., Wiebold, W., Batchelor, W., Bollero, G., Bullock, D., Clay, D., Palm, H., Pierce, F., & Schuler, R. (2005). Relating apparent electrical conductivity to soil properties across the north-central USA. *Computers and Electronics in Agriculture, 46*, 263-283
Taylor, J., McBratney, A., & Whelan, B. (2007). Establishing management classes for broadacre agricultural production. *Agronomy Journal, 99*, 1366-1376
Thessler, S., Kooistra, L., Teye, F., Huitu, H., & Bregt, A.K. (2011). Geosensors to support crop production: current applications and user requirements. *Sensors, 11*, 6656-6684
Tilman, D., Cassman, K.G., Matson, P.A., Naylor, R., & Polasky, S. (2002). Agricultural sustainability and intensive production practices. *Nature, 418*, 671-677
Van Meirvenne, M., Islam, M.M., De Smedt, P., Meerschman, E., Van De Vijver, E., & Saey, T. (2012). Key variables for the identification of soil management classes in the aeolian landscapes of north–west Europe. *Geoderma*
Van Meirvenne, M., Islam, M.M., De Smedt, P., Meerschman, E., Van De Vijver, E., & Saey, T. (2013). Key variables for the identification of soil management classes in the aeolian landscapes of north–west Europe. *Geoderma, 199*, 99-105
Viscarra Rossel, R.A., & McBratney, A.B. (1998). Laboratory evaluation of a proximal sensing technique for simultaneous measurement of soil clay and water content. *Geoderma, 85*, 19-39
Vitharana, U., Van Meirvenne, M., Cockx, L., & Bourgeois, J. (2006). Identifying potential management zones in a layered soil using several sources of ancillary information. *Soil Use and Management, 22*, 405-413
Vitharana, U.W., Van Meirvenne, M., Simpson, D., Cockx, L., & De Baerdemaeker, J. (2008). Key soil and topographic properties to delineate potential management classes for precision agriculture in the European loess area. *Geoderma, 143*, 206-215
Vlaanderen, O. (2003). Digitaal hoogtemodel vlaanderen. *Nieuwsbrief GIS-Vlaanderen, 16*

Whelan, B., & McBratney, A. (2000). The “null hypothesis” of precision agriculture management. *Precision Agriculture, 2*, 265-279
Whelan, B., & McBratney, A. (2003). Definition and interpretation of potential management zones in Australia. In, *Proceedings of the 11th Australian Agronomy Conference* (pp. 2-6)
Whelan, B., & Taylor, J. (2013). *Precision Agriculture for Grain Production Systems*. Csiro publishing
Williams, B.G., & Hoey, D. (1987). The use of electromagnetic induction to detect the spatial variability of the salt and clay contents of soils. *Soil Research, 25*, 21-27
Xin-Zhong, W., Guo-Shun, L., Hong-Chao, H., Zhen-Hai, W., Qing-Hua, L., Xu-Feng, L., Wei-Hong, H., & Yan-Tao, L. (2009). Determination of management zones for a tobacco field based on soil fertility. *Computers and Electronics in Agriculture, 65*, 168-175
Zhang, N., Wang, M., & Wang, N. (2002a). Precision agriculture—a worldwide overview. *Computers and Electronics in Agriculture, 36*, 113-132
Zhang, N., Wang, M., & Wang, N. (2002b). Precision agriculture—a worldwide overview. *Computers and Electronics in Agriculture, 36*, 113-132
Zhu, Q., & Lin, H. (2010). Comparing ordinary kriging and regression kriging for soil properties in contrasting landscapes. *Pedosphere, 20*, 594-606

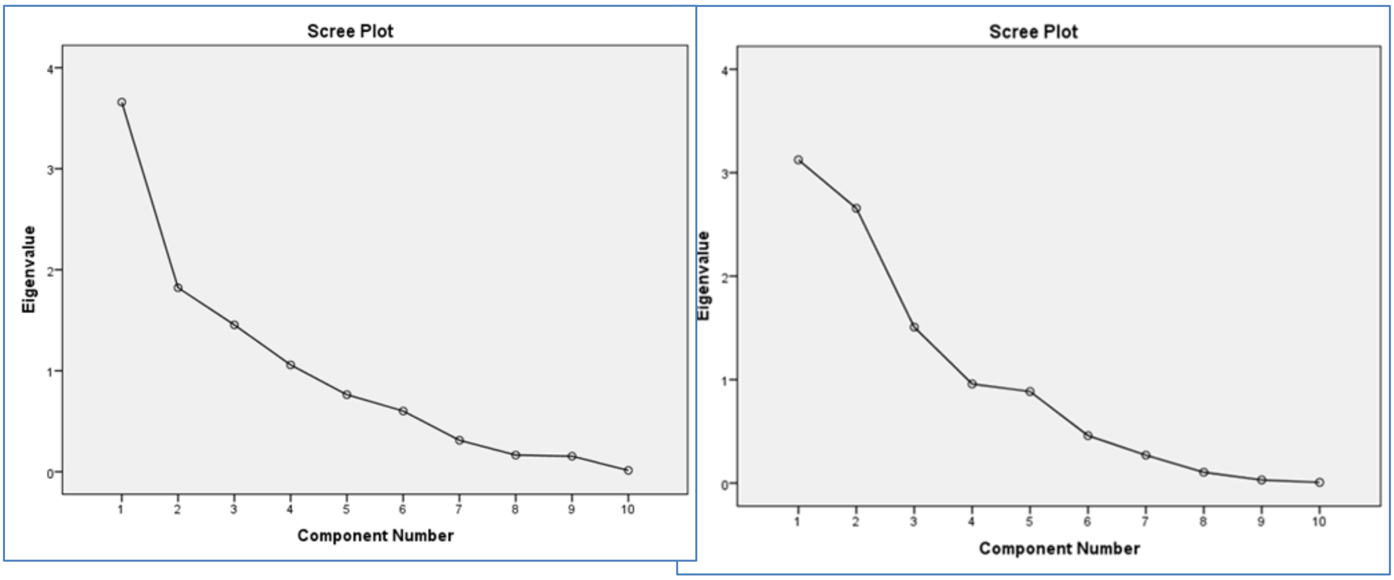
Appendixes

Scree plots PCA:

PCA scree plots of the three scenarios are presented here below. Two scree plots for scenario II for each field one and field two are presented in Appendix 2. Scree plots of scenario I and III are also presented in Appendix 1.



Appendix 1.PCA scree plots of scenario I (left) and III (right)

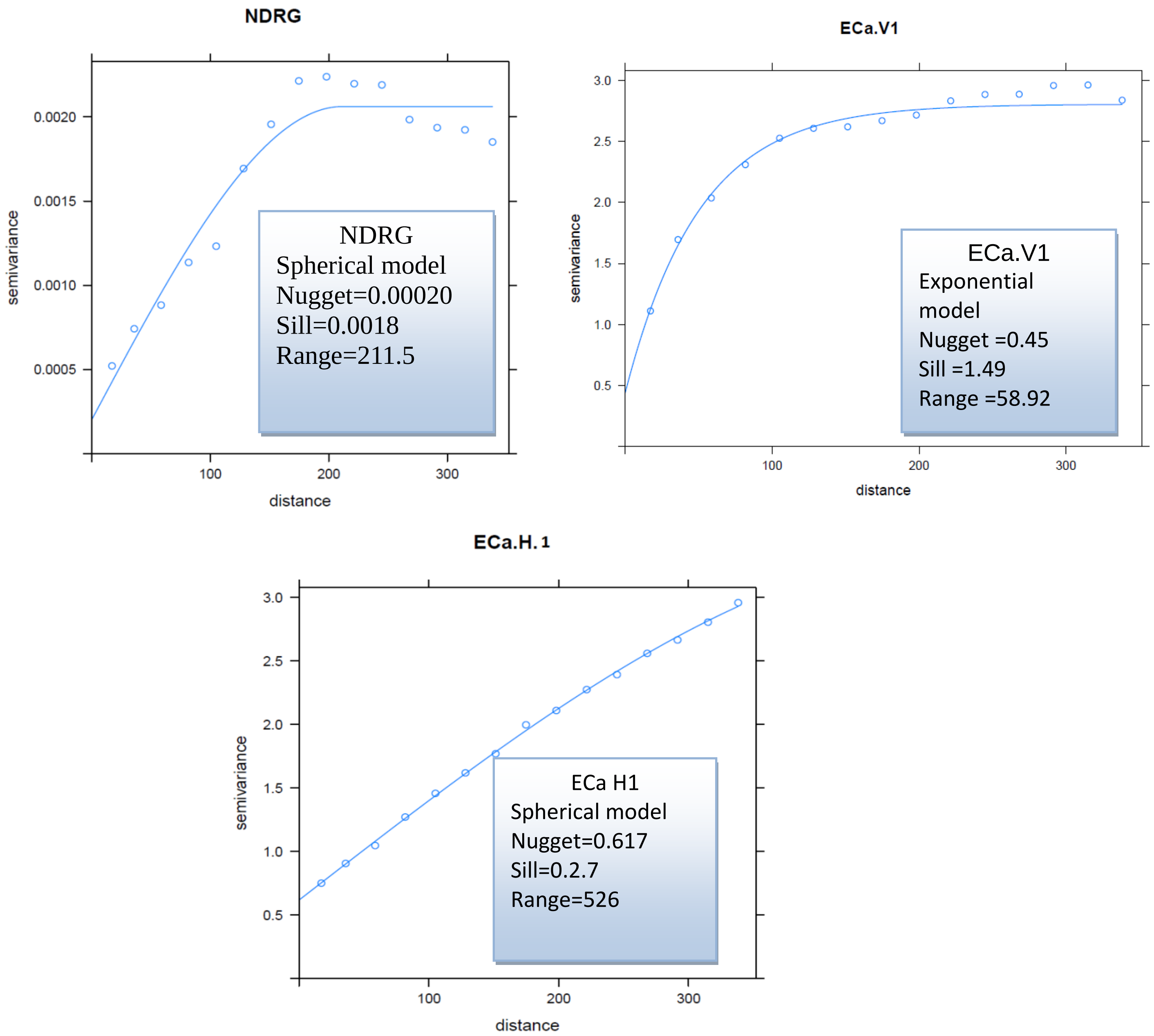


Appendix 2.PCA scree plot of scenario II, field one (left) and field two (right)

Semi-variograms and fitted curves parameters:

Semi-variograms and their fitted curves of the selected soil parameters for each of the three scenarios are presented in Appendix 3, Appendix 4, Appendix 5 and Appendix 6.

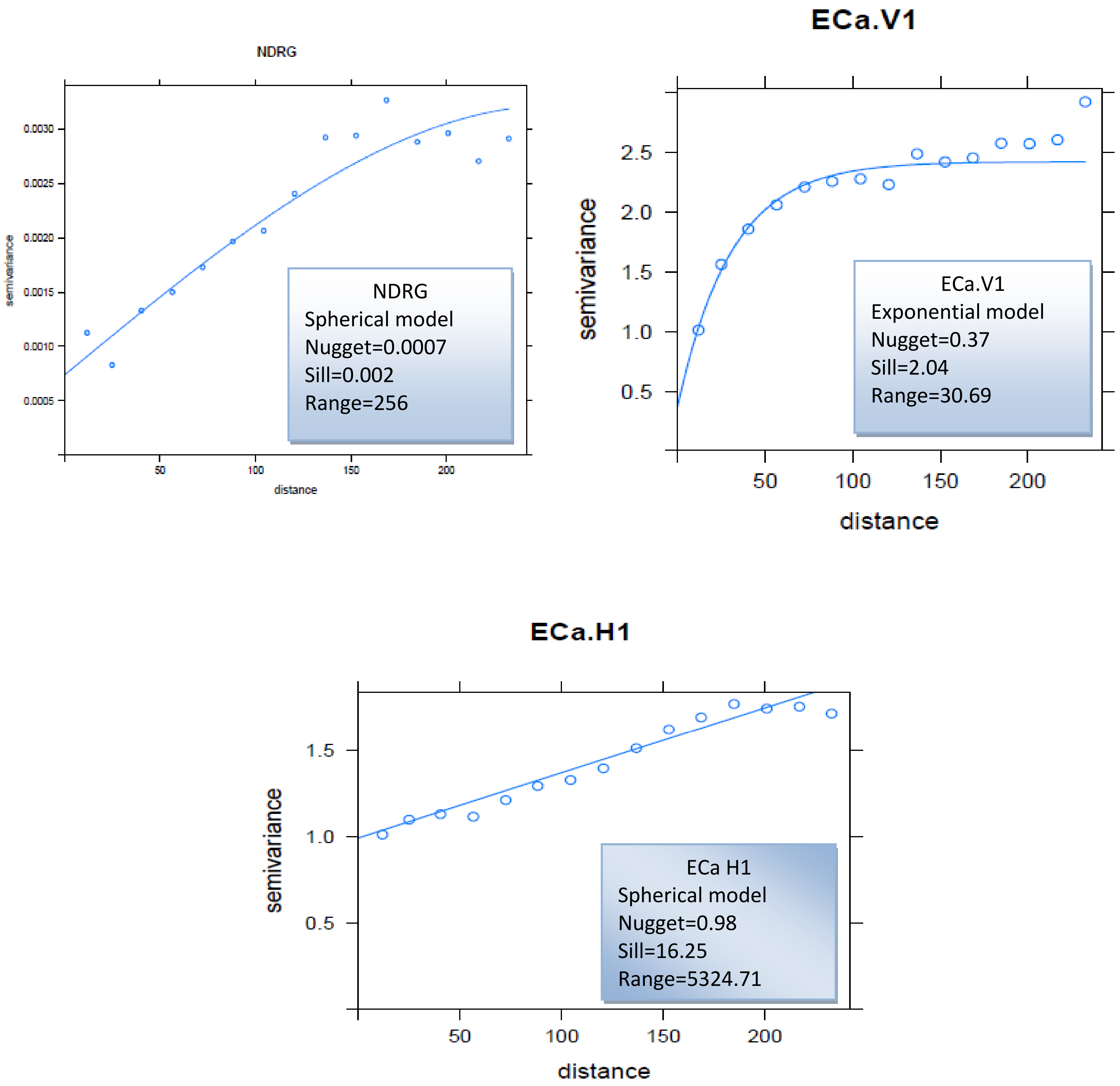
Scenario I: NDRG, ECa-V1 and ECa-H1 soil property are used the selected key soil properties for scenario I and their respected semi-variograms and fitted curves parameters are presented.



Appendix 3. Semivariograms of soil variables and their fitted curves and parameters for scenario I.

Scenario II

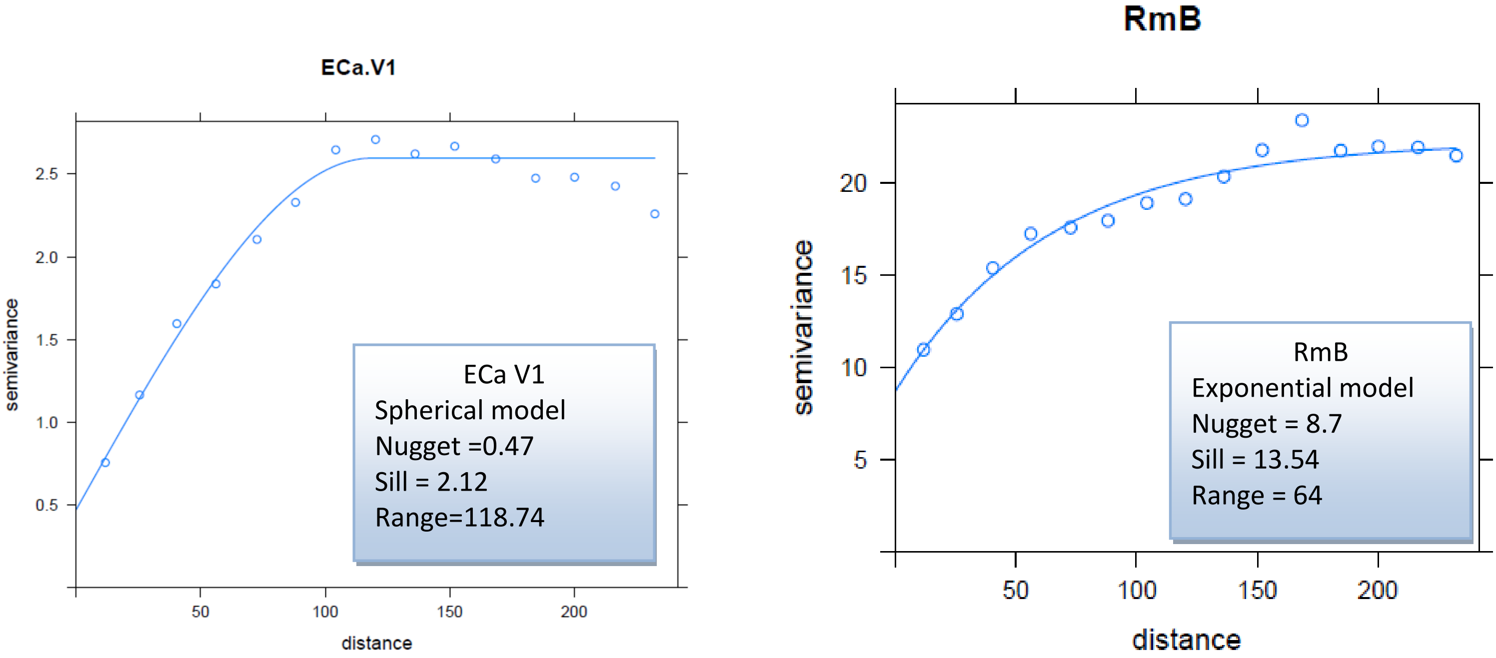
Field one: For field one under scenario II, NDRG, ECa-V1 and ECa-H1 soil variables were selected. Respective Semi-variograms and fitted curves parameters are presented in Appendix 4.



Appendix 4. Semivariograms of soil variables and their fitted curves and parameters for scenario II.

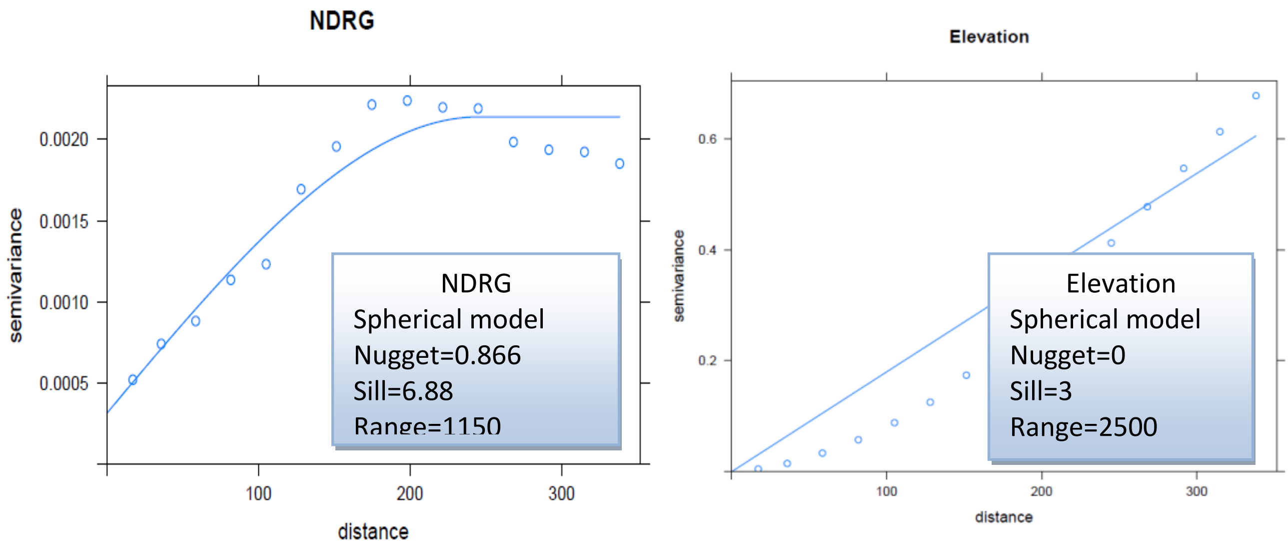
Scenario II

Field two: For field two under scenario II**,** ECa-V1, RmB, and elevation key soil parameters were selected. Semivariograms of ECa-V1and RmB variables and their fitted curves and parameters are presented in Appendix 5.



Appendix 5. Semivariograms of soil variables and their fitted curves and parameters for scenario II.

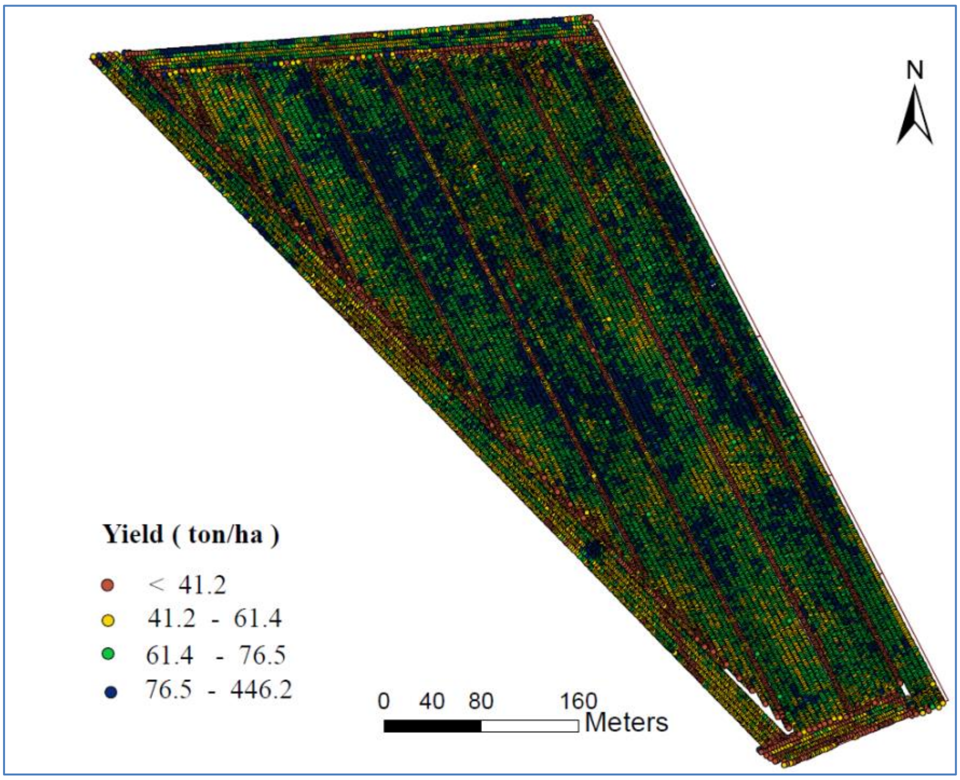
Scenario III: NDRG and elevation soil variables were selected for this scenario. Semi-variograms of these variables and their fitted curves and parameters are presented in Appendix 6.

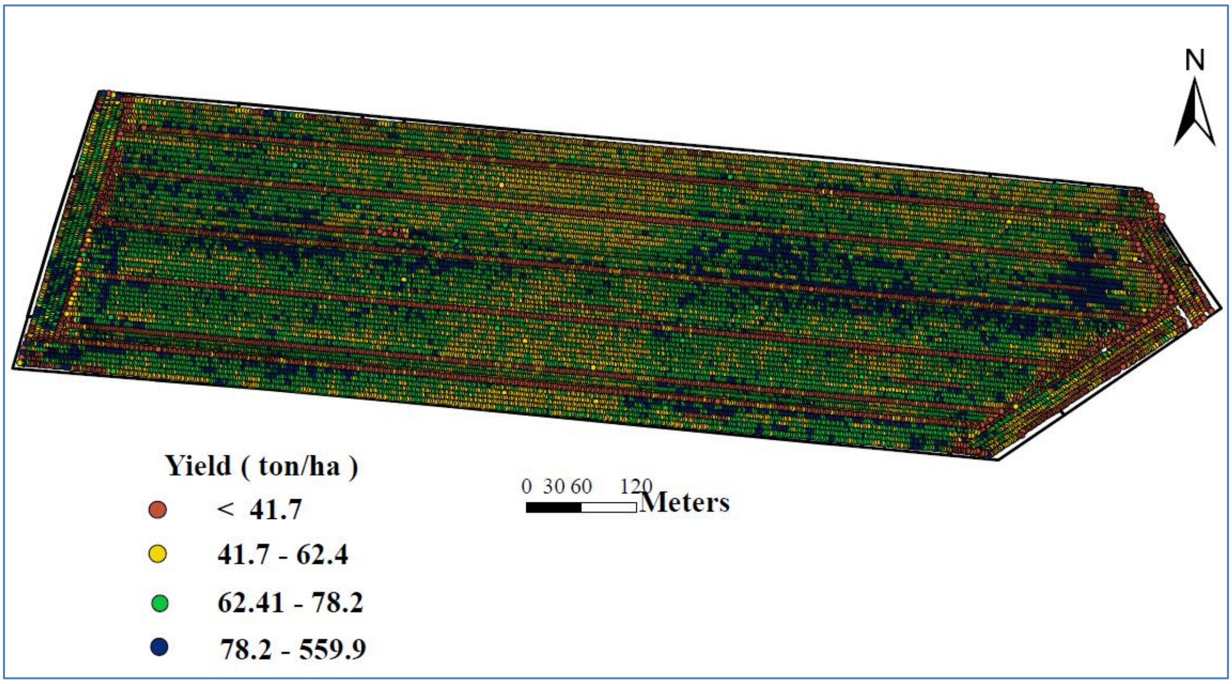


Appendix 6. Semivariograms of soil variables and their fitted curves and parameters for scenario III.

Yield maps:

Normalized potato yield distribution map. This map simply illustrates normalized potato yield is classified in to four categories. The created map is presented for each field in Appendix 7.

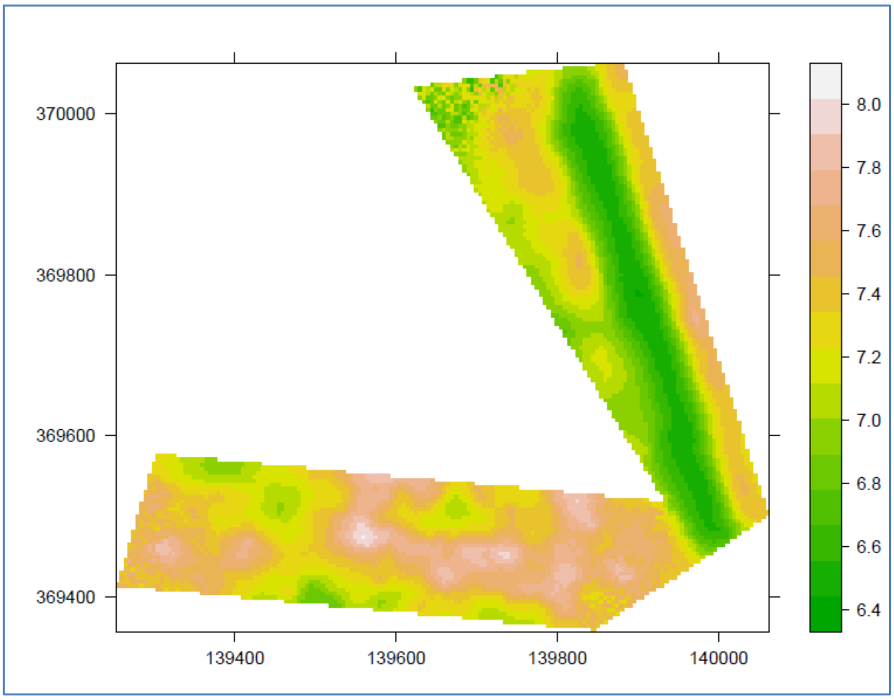




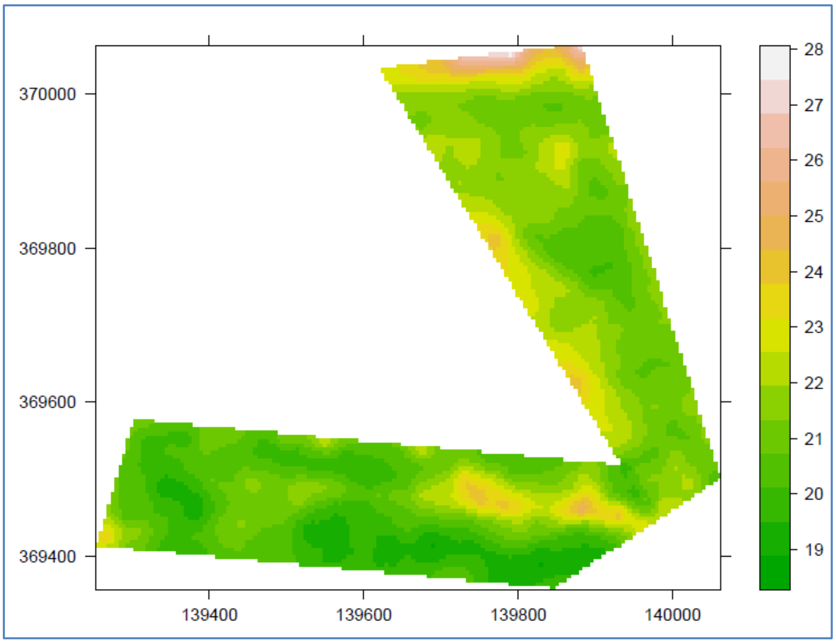
Appendix 7. Normalized potato yield distribution map field one (top) and field two (bottom)

Interpolated soil property maps:

In contrast to other studies the pH value was not selected as key parameter. The kriged soil pH maps presented in Appendix 8 (a) to show its spatial distribution. The kriged map of ECa-V5 is also presented in Appendix 8 (b).



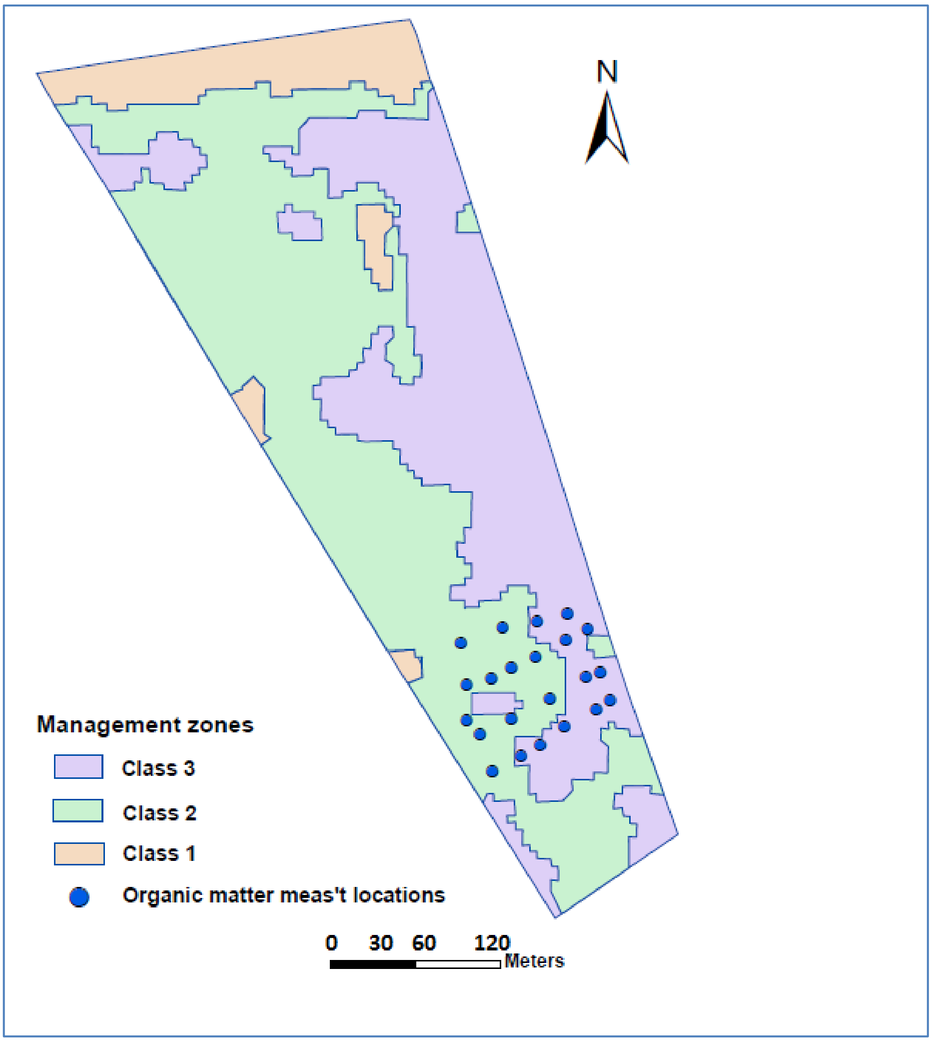
a)



b)

Appendix 8. Interpolated maps of pH value (a) and ECa-V.5 (b) using kriging.

The map below demonstrates the spatial locations of soil organic matter samples. the organic matter samples are located on zone2 and zone 3 of management zones field one defined under scenario II. It can be observed that the samples are partly distributed on the southern part of the field.



Appendix 9. Organic matter measurement location displayed over management zones for field 1.