UNIVERSITÀ DEGLI STUDI DI SASSARI

SCUOLA DI DOTTORATO DI RICERCA

Scienze e Biotecnologie

dei Sistemi Agrari e Forestali

e delle Produzioni Alimentari

Indirizzo: Agrometeorologia ed ecofisiologia dei sistemi agrari e forestali

Ciclo XXVI

Proximal soil sensors and geostatistical tools in precision agriculture applications

dr. Sameh Mohammed Shaddad

Direttore della Scuola	Prof. Alba Pusino
Referente di Indirizzo	Prof. Donatella Spano
Docente Guida	Prof. Maurizio Mulas
Docente Tutor	Prof. Salvatore Madrau
Correlatore	Prof. Abdul Mounem Mouazen
Correlatore	Prof. Annamaria Castrignanò

Anno accademico 2012-2013

Abstract:

Recognition of spatial variability is very important in precision agriculture applications. The use of proximal soil sensors and geostatistical techniques is highly recommended worldwide to detect spatial variation not only in fields but also within-field (micro-scale). This study involves, as a first step, the use of visible and near infrared (vis-NIR) spectroscopy to estimate soil key properties (6) and obtain high resolution maps that allow us to model the spatial variability in the soil. Different calibration models were developed using partial least square regression (PLSR) for different soil properties. These calibration models were evaluated by both cross-validation and independent validation. Results show good to excellent calibration models for most of soil properties under study in both cross-validation and independent validation. The on-line maps created using the collected on-line spectra and the calibration models previously estimated for each soil property were compared with three different maps (measured, predicted, error). The second step uses multivariate geostatistical analysis to develop three different geostatistical models (soil, spectral, fusion). The soil model includes 8 soil properties, spectral model includes 4 soil properties and the fusion model includes 12 soil properties. The three models were evaluated by cross-validation and the results show that the goodness of fitting can be considered as satisfactory for the soil model, whereas the performance of the spectral model was quite poor. Regarding the fusion model, it performed quite well, though the model generally underestimated the high values and overestimated the low values.

An independent validation data set was used to evaluate the performance of the three models calculating three statistics: mean error (ME), as an indicator of bias; mean standardized squared error (MSSE), as an indicator of accuracy, and root mean squared error (RMSE), as an indicator of precision of estimation. Synthetically, the two, soil and fusion, models performed quite similarly, whereas the performance of the spectral model was much poorer.

With regard to delineation of management zones (MZs), the factor cokriging analysis was applied using the three different models. The first factor (F1) for the soil and fusion models was related to soil properties that affect soil fertility, whereas for the spectral model was related to P (-0.88) and pH (-0.42). Based on the first factor of the soil and fusion models, three management zones were delineated and classified as low, medium and high fertility zones using isofrequency classes.

Spatial similarity between the yield map and delineated MZs maps based on F1 for the soil and fusion models was calculated. The overall accordance between the two maps was 40.0 % for the soil model and 38.6 % for the fusion model. The two models performed quite similarly. These results can be interpreted as more than 50% of the yield variation was ascribable to more dynamic factors than soil parameters not included in this study, such as agro-meteorological conditions, plant diseases, nutrition stresses, etc. However, the results are quite promising for the application of the proposed approach in site-specific management.

Keywords: vis-NIR on-line sensor, PLSR, Multi-collocated Cokriging, Factor cokriging, Management Zones.

Riassunto:

Il riconoscimento della variabilità spaziale è molto importante nelle applicazioni di agricoltura di precisione. L'uso dei sensori di rilevamento prossimale di suolo e delle tecniche geostatistiche è attualmente raccomandato a livello globale per rilevare variazioni spaziali non solo a scala di campo, ma anche a livello puntuale (micro-scala). Questo studio ha richiesto, ai fini di una indagine preliminare finalizzata alla stima delle proprietà chiave dei suoli, un analisi spettroscopica nel visibile e vicino infrarosso (vis-NIR) per la realizzazione di mappe ad alta risoluzione spaziale. Sono stati stimati i modelli di calibrazione per le diverse proprietà del suolo, utilizzando la regressione parziale ai minimi quadrati (PLSR). I modelli sono stati valutati sia mediante cross-validation che utilizzando un data set indipendente di validazione.

I risultati indicano per la maggior parte delle proprietà del suolo considerate un adattamento da buono a eccellente. Le mappe delle proprietà del suolo, ottenute dagli spettri registrati on-line e dai modelli di calibrazione precedentemente stimati, sono state valutate mediante il calcolo di varie statistiche dell'errore sperimentale.

Successivamente è sta applicata un'analisi geostatistica multivariata per sviluppare tre diversi modelli geostatistici, denominati: suolo, spettrale, fusione. Il modello suolo interessa 8 proprietà del suolo, lo spettrale 4, ed infine quello di fusione 12. I risultati dei tre modelli sono stati valutati mediante cross-validazione ed indicano che la bontà del *fitting* può essere considerata soddisfacente per il modello suolo e di fusione, mentre la performance del modello spettrale è risultata piuttosto scarsa. Il data set indipendente di validazione è stato utilizzato per valutare le prestazioni dei tre modelli, calcolando tre statistiche: errore medio (ME), come indicatore di bias; errore quadratico medio (MSE), come indicatore di precisione, e deviazione standard dell'errore standardizzato (RMSE), come indicatore di accuratezza della stima. I risultati mostrano che i modelli suolo e di fusione hanno operato in modo simile, mentre la performance del modello spettrale è stata decisamente inferiore.

Per la delineazione delle zone di gestione differenziata (MZ), è stato impiegato il *factor cokriging*, utilizzando i tre modelli citati. Il primo fattore regionalizzato (F1), strutturato spazialmente nei modelli suolo e di fusione, è risultato correlato alle proprietà del suolo che ne influenzano la fertilità, mentre per il modello spettrale il primo fattore relativo alla struttura spaziale con range più lungo era correlato a P (-0.88) e pH (-0.42). Il primo fattore per i modelli suolo e di fusione è stato mappato, utilizzando tre classi di iso-frequenza, che individuavano tre aree a differente livello di fertilità: bassa media e alta. È stata altresì

calcolata la similarità spaziale tra una mappa di resa della coltura di orzo e quelle ottenute sulla base dei valori F1 per i modelli suolo e di fusione. La conformità generale tra le due mappe è pari al 40,0 % per il modello suolo e al 38,6 % per il modello di fusione. Questi ultimi risultati indicano come più del 50 % della variazione della resa in campo sia ascrivibile a fattori più dinamici dei parametri del suolo e non inclusi in questo studio, tra cui condizioni agro-meteorologiche, condizioni fitosanitario delle colture, stress nutrizionali e idrici, ecc.

Benché sia stata evidenziata l'influenza sulla produttività anche di fattori diversi da quelli strettamente pedologici, i modelli proposti possono rappresentare un valido strumento per la gestione sito- specifica in ambito aziendale.

Acknowledgment

*I would like to express my deepest gratitude to my supervisor **Prof. Annamaria Castrignanò**, Consiglio Nazionale per la Ricerca e la Sperimentazione in agricoltura (Counsil for Research in Agriculture)(CRA), Italy for her precise supervision and valuable guidance during the whole work, especially during the geostatistical analysis.*

*Sincere thanks are due to **Prof. Abdul Mounem Mouazen**, Senior lecturer at School of Applied Sciences, Cranfield University, United Kingdom for giving me the opportunity to use his internationally patented multiple-sensor platform to conduct this study and precise supervision during the period I spent at Cranfield University, UK.*

*Thanks are also extended to **Prof. Salvatore Madrau,** associate professor of soil pedology, Sassari University, Italy for helping me to have the opportunity to travel to Cranfield University, United Kingdom, and for his continuous encouragement during the whole work.*

*Sincere thanks are given to **Prof. Maurizio Mulas**, associate professor of agrometeorology, Sassari University, Italy for his valuable guidance and continuous encouragement during this study.*

*I also gratefully acknowledge all technical support staff in the Environmental Science and Technology Department, Cranfield University for their kind helps on field work and laboratory scanning analysis. I gratefully thank my colleague **Dr. Boyan Kuang,** Cranfield University for helping me during the elaboration of spectral data. Sincere thanks are given to **Mr. Ian Truckle,** Cranfield University for providing me different types of softwares and maps of the studied field.*

*Sincere thanks are also extended to **Mr. Mario Deroma**, PI of Soil laboratory Analysis, Sassari University for his supporting during the physical and chemical analysis of soil samples. Thanks are due to all colleagues at **Department of Agricultural Sciences, Sassari University, Italy** for their help and encouragement.*

Special thanks are given to the referees for their very useful comments and suggestions for improving my thesis quality.

I am also indebted to my family members in Egypt for their patience and encouragement and wish their all the best.

TABLE OF CONTENT

List of figures

List of tables

List of abbreviations

CEC	Cation-exchange capacity
DGPS	Differential global positioning system
K	Potassium
LV	Latent variable
MC	Moisture content
MIR	Mid infrared
MZ	Management zone
NIR	Near infrared
P	Phosphorus
PA	Precision agriculture
PCR	Principal component analysis
PLSR	Partial least square regression
r	Coefficient of correlation
R^2	Coefficient of determination
RMSE	Root mean square error
RMSECV	Root mean square error of cross-validation
RMSEP	Root mean square error of prediction
RPD	Ratio of prediction to deviation
SOM	Soil organic matter
TC	Total carbon
TN	Total nitrogen
UK	United Kingdom
vis	visible

1. Introduction

Soils are important natural resources for the existence of mankind. They have to be carefully understood to optimise productivity and agricultural production in general. Agricultural soils are highly complex and heterogeneous materials that vary greatly at both spatial and temporal scales. Spatial variation of soil properties is attributed to the complex interaction of many factors including climate that impacts over large distance (Jenny, 1980), whereas soil weathering impacts over long time periods. But both previous factors are modified by other processes which operate more locally (e.g. erosion and deposition of parent materials) or more frequently (e.g. weather). This nested nature of soil variability indicates that type and cause of variation depend largely on the scale and frequency of observation (Castrignanò, 2011).

No two soils are exactly alike and variations occur over short distances, both vertically and horizontally. Given the importance of soils, there is a need for regular monitoring to detect changes in its status so as to implement appropriate management in the event of degradation. Soil surveying may be performed at national levels for the inventory of soil resources, or for agriculture at regional, farm or field scales. Monitoring of carbon, nutrient status, pH, and salinity are among soil properties that need to be quantified for agriculture applications (Stenberg et al., 2010).

Precision agriculture is a site-specific management method explicitly taking within field variation of soil and crop into consideration. It is a system-driven technology that provides spatial and temporal information (where, how much, and when to apply) about the application of farm inputs such as tillage, irrigation, fertilizers, pesticides, etc., in a field (Corwin and Lesch, 2005; Gebbers and Adamchuk, 2010; Pierce and Nowak, 1999). Knowledge concerning within field variability is essential for the successful implementation of precision agriculture (Bullock and Bullock, 2000). However, recommendations on application of different inputs into soils were set, until recently, by considering agricultural fields as uniform elements, despite the evidence of within-field spatial variability. Ignoring this variability can result in poor land and crop management, leading to yield losses or inefficient use of inputs. Apart from the costs associated with excess use of chemicals, for example, fertilizers, pesticides, and herbicides, environmental impact may be considerable, due to the increased contamination of water resources.

In order to describe the spatial and temporal variability accurately, direct soil sampling and lab analyses should be carried out which are expensive and time-consuming. Therefore, they are being complemented with soil sensing techniques which have high spatial resolution and fast performance. So far, there are two approaches to collect auxiliary data to determine soil variability. The first one is remote sensing, which is based on contactless approach and does not require access to the field. But, it still suffers from inadequate spatial and temporal resolution (McBratney et al., 2003) and interference from clouds and crop residue cover. It is therefore expected that precision agriculture will have to rely on data obtained in the field at least in the near future. The second approach is proximal soil sensing that involves the use of ground-based sensors to collect soil information from close by (say within 2 m) or within, the soil body. Proximal soil sensors can be active or passive; they are invasive, when there is a direct sensor-to-soil contact, or non-invasive, measuring soil properties from above the surface. They can measure the soil property either directly or indirectly by finding a proxy that is easier and cheaper to measure and developing a pedotransfer function. Frequently the sensors are mounted on vehicles for on-the-go measurements (Viscarra Rossel et al., 2010).

Several sensing techniques are being tested and explored or are already available in the market for commercial applications (Adamchuk et al., 2005; Kuang et al., 2012). Visible and near-infrared (vis-NIR) spectroscopy sensors are one of those techniques that have been reported to successfully measure soil properties either in laboratory or on-line. The vis-NIR spectrum is characterized as the region from 400 to 2500 nm of the electromagnetic spectrum (EMS) corresponding to 25000 to 4000 cm^{-1} (wave numbers). There are many reasons for researchers being interested in the implementation of vis–NIR spectroscopy. For example, sample preparation involves only drying and crushing; the sample is not affected by the analysis in any way; no (hazardous) chemicals are required; measurement takes a few seconds; several soil properties can be estimated from a single scan; and the technique can be used both in the laboratory and in situ (Viscarra Rossel et al., 2006). Although laboratory methods are very useful for soil and crop management, on-line methods are essential for precision agriculture applications, since they able to provide detailed information about soil parameters quickly and in a cost-effective way (Mouazen et al., 2007).

Due to the overlapping absorption of soil constituents, diffuse reflectance spectra of soil in vis–NIR are largely nonspecific. This characteristic lack of specificity is compounded by scatter effects caused by soil structure or specific constituents such as quartz. All of these factors result in complex absorption patterns that need to be mathematically extracted from

the spectra and correlated with soil properties (Stenberg et al., 2010). Hence, the analysis of soil diffuse reflectance spectra require the use of multivariate calibrations (Martens and Naes, 1989). The most common calibration methods for soil applications are based on linear regression, namely stepwise multiple linear regression (SMLR) (Ben-Dor and Banin, 1995; Dalal and Henry, 1986), principal component regression (PCR), and partial least squares regression (PLSR). The main reason for using SMLR is the inadequacy of more conventional regression techniques such as multiple linear regression (MLR) and lack of awareness among soil scientists of the existence of full spectrum data compression techniques such as PCR and PLSR. Both of these techniques can cope with data containing large numbers of predictor variables that are highly collinear. PCR and PLSR are related techniques and in most situations their prediction errors are similar. However, PLSR is often preferred by analysts because it relates the response and predictor variables so that the model explains more of the variance in the response with fewer components, it is more interpretable and the algorithm is computationally faster. The use of data mining techniques such as neural networks (NN) (e.g., Daniel et al., 2003; Fidencio et al., 2002), multivariate adaptive regression splines (MARS) (Shepherd and Walsh, 2002), and boosted regression trees (Brown et al., 2006) is increasing. Viscarra Rossel (2007) combined PLSR with bootstrap aggregation (bagging-PLSR) to improve the robustness of the PLSR models and produce predictions with uncertainty. MLR, PCR, and PLS are linear models, while the data mining techniques can handle nonlinear data. Viscarra Rossel and Lark (2009) used wavelets combined with polynomial regression to reduce the spectral data, account for non linearity and produce accurate and parsimonious calibrations based on selected wavelet coefficients. Mouazen et al. (2010) compared NN with PCR and PLS for the prediction of selected soil properties. They found combined PLSR-NN models to provide improved predictions as compared to PLSR and PCR. Viscarra Rossel and Behrens (2010) compared the use of PLSR to a number of data mining algorithms and feature selection techniques for the prediction of clay, organic carbon and pH. They compared MARS, random forests (RF), boosted trees (BT), support vector machines (SVM), NN and wavelets. Their results suggest that data mining algorithms produce more accurate results than PLSR and that some of the algorithms provide information on the importance of specific wavelength in the models so that they can be used to interpret them.

Some soil properties have a direct spectral response in the NIR range such as moisture content, Total or organic nitrogen, clay content and organic carbon (Stenberg et al., 2010). The prediction of soil C and N depends upon the forms to be measured and the most

successful is reported for organic, and total forms (Yang et al., 2012). Other soil properties have indirect spectral response such as pH, Ca and Mg which can be measured through co-variation with other soil properties having direct spectral responses (Stenberg et al., 2010). It was reported that the worst measurable properties with indirect spectral responses are K and Na (Kuang et al., 2012). Literature states the potential of Vis-NIR spectroscopy for the assessment of heavy metals, e.g. Fe, Ni, Cu and Zn with acceptable accuracy (Kuang et al., 2012).

Soil properties are continuous variables whose values at any location are expected to vary according to the direction and distance of separation from neighbouring samples (Burgess and Webster, 1980). Therefore, the classical approach for the interpolation of spatially dependent variables is inadequate, because it assumes random variation within units and no correlation between units and takes no account of relative location of samples. An alternative approach is to treat the soil as a random function and to describe it using the methods of Matheron's regionalized variable theory (known as geostatistics) (Matheron, 1973). The theory expresses the idea that values of a soil property at near places are likely to be similar, whereas those at places far from one another are not. It does so quantitatively and in a way that can be used for interpolation (Castrignanò, 2011). Geostatistics is one of statistics' branches focusing on spatial or spatiotemporal datasets. Developed originally to predict probability distributions of ore grades for mining operations, it is currently applied in diverse disciplines including petroleum geology, soil science, hydrogeology, hydrology, meteorology, geochemistry, oceanography, geography, forestry, environmental control, landscape ecology, and agriculture especially in precision farming.

Geostatistics provides descriptive tools such as semivariograms to characterize the spatial pattern of continuous and categorical soil attributes. Various interpolation ((co)kriging) techniques capitalize on spatial correlation between observations to predict attribute values at unsampled locations using information related to one or several attributes. An important contribution of geostatistics is the assessment of the uncertainty about unsampled values, which usually takes the form of a map of the probability of exceeding critical values, such as regulatory thresholds in soil pollution or criteria for soil quality. This uncertainty assessment can be combined with expert knowledge for decision making such as delineation of contaminated areas where remedial measures should be taken or areas of good soil quality where specific management plans can be developed. Last, stochastic simulation allows one to generate several models (images) of the spatial distribution of soil attribute values, all of

which are consistent with the information available. A given scenario (remediation process, land use policy) can be applied to the set of realizations, allowing the uncertainty of the response (remediation efficiency, soil productivity) to be assessed (Goovaerts, 1999).

Geostatistics is applied in many aspects of precision agriculture (PA) including sampling, prediction, mapping, decision making, variable-rate applications, economics and so on. Contributions from experts in several fields of study illustrate how geostatistics can be applied advantageously in the handling of data of different type, such as yield, soil, crops, pests, aerial photographs, remote and proximal imagery. Geostatistical techniques include variography, simple-, ordinary-, disjunctive-, indicator-, regression-, and space-time- kriging, factorial kriging and co-kriging, and stochastic simulation. (Oliver, 2010).

2. Aims

Visible and near infrared spectroscopy (vis-NIRS) is used as one of proximal soil sensing technologies for laboratory and on-line measurement conditions. This is fast, cost effective, easy and can provide high resolution data on spatial variation in soil properties. the main aim of this study is the use of vis-NIR on-line sensor data and geostatistical tools to delineate management zones. Recognition of within-field spatial variability allows us to manage the soil in the best way, reduce tillage costs, decrease the environmental impact of fertilization and pesticide use and decide the optimal destination (cultivated or wild land), in order to maximize their productivity.

3. Objectives

In order to fulfil the general aims of the current research, the following objectives were achieved:

- To develop calibration models of soil properties under consideration by linking the soil measured data with laboratory-based spectral information using multivariate statistical tools.
- To carry out on-line measurement in one selected field using a vis-NIR on-line sensor.
- To develop high resolution maps of soil properties of interest using on-line collected data.
- To delineate management zones using multivariate geostatistical analysis applied on laboratory measured values of soil properties.
- To delineate management zones using multivariate geostatistical analysis applied on key soil properties predicted from calibration models using spectral data as regressors.
- To delineate management zones using multivariate geostatistical analysis applied on fused data of laboratory measured and on-line predicted soil properties.
- To evaluate and compare between three different methods for management zones delineation through a validation test.
- To assess the spatial association of the yield map of one year with each of the three maps of management zones delineation with different statistical methods.

Figure 1 shows the overall structure of the thesis where there can be seen in an easy way all steps performed in this research.

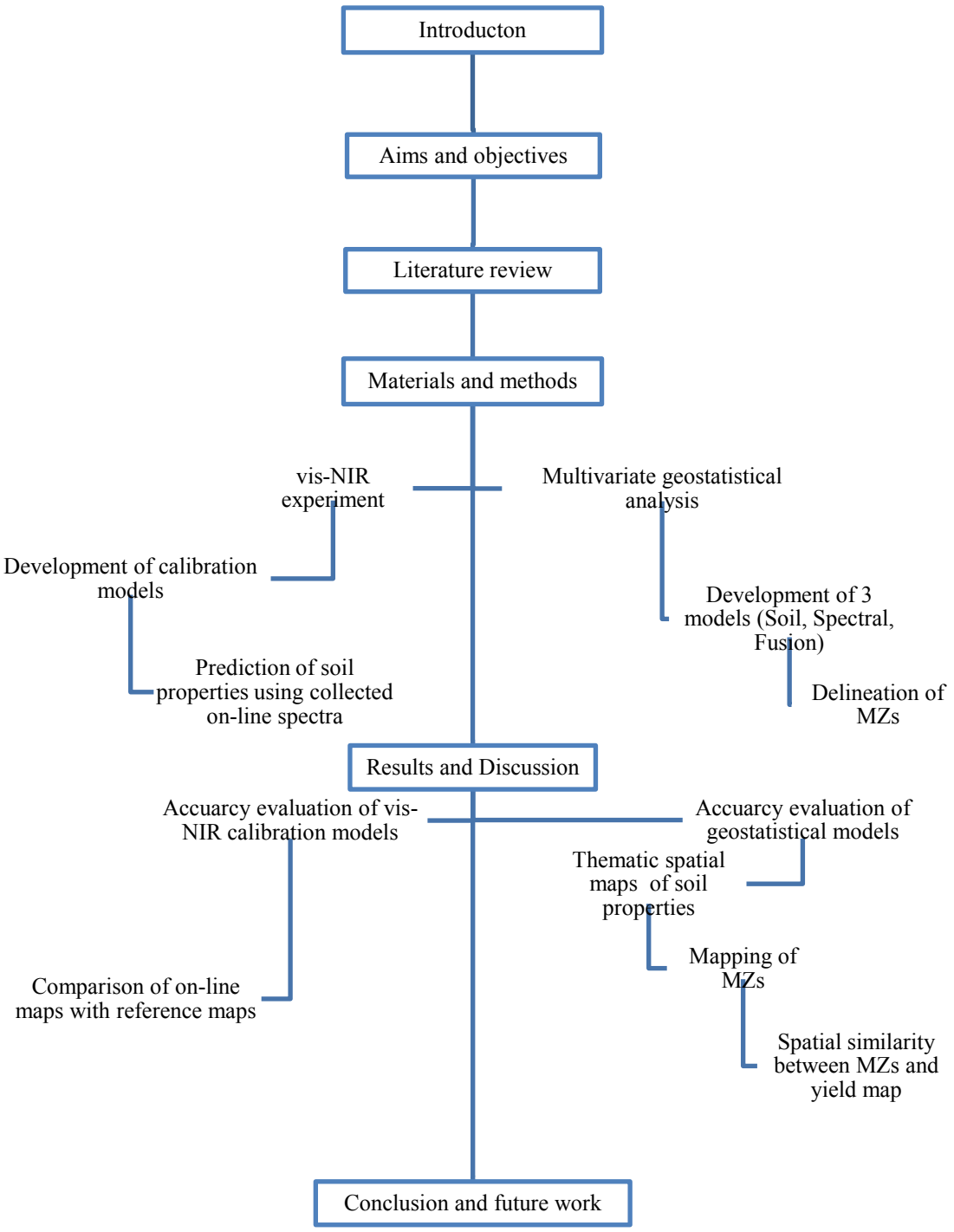

Figure 1. Flow chart of thesis structure

4. Literature review

This literature review focuses on proximal soil sensing particularly on visible and near infrared spectroscopy and its important role in assessing soil properties. In addition, it reviews previous works on the use of geostatistical tools for delineation of site-specific management zones using soil data as well as the fusion of soil spectra and soil data.

4.1 Proximal soil sensing

In precision agriculture, the use of proximal soil sensors is highly recommended, as these can provide high resolution data on spatial variation in soil properties (Stenberg et al., 2010), which enables the management of land at field and sub-field scale.

Potentially, proximal or ground-based (invasive or non-invasive) soil sensors have the ability to collect high-resolution data rapidly, and in certain cases even allowing real-time analysis and processing, by taking measurements as frequently as one per second (Viscarra Rossel and McBratney, 1998). Sensor-based soil analysis potentially provides several advantages over conventional laboratory methods such as lower cost, increased efficiency, more timely results, and collection of dense datasets while just traversing a field.

The following five categories are suggested by Kuang et al., (2012) for laboratory, *in situ* non-mobile, and on-line mobile measurement conditions:

1. Conductivity, resistivity, and permittivity based soil sensors
2. Passive radiometric based soil sensors
3. Strength based soil sensors
4. Electrochemical based soil sensors
5. Reflectance based soil sensors

4.1.1 Conductivity, resistivity, and permittivity based soil sensors

It has been reported (Kuang et al., 2012) that by assessing soil electrical conductivity, resistivity and permittivity, soil properties can be measured directly or indirectly. There are different types of sensors used under this category such as electrical conductivity (EC), time domain reflectance (TDR), frequency domain reflectance (FDR), electrical resistivity (ER) and electromagnetic induction (EMI).

4.1.1.1 EMI sensors

Usage of EMI sensors in agriculture was first introduced in the late 1970s for assessing soil salinity (Corwin and Rhoades, 1982; de Jong et al., 1979; Rhoades and Corwin, 1981; Williams and Baker, 1982). Today EMI is used worldwide to assess within field variability. The EMI device involves a transmitter coil and a receiver coil installed on both ends of a nonconductive bar. Different EMI sensors are available in global markets for ECa measurement (e.g. EM31, EM34, EM38, etc), however, according to literature (Kuang et al., 2012) the most frequently used EMI device is EM38.

Several studies have been conducted to determine soil parameters, some of those were carried out *in situ* to assess soil salinity with R^2 values range between 0.5 – 0.98 (McNeill, 1992; Rhoades et al., 1999; Herrero et al., 2003; McLeod et al., 2010). Water content as well as soil texture were measured *in situ* with R^2 values ranging between 0.37 and 0.99 and between 0.20 and 0.90, respectively (Williams and Hoey, 1987; Domsch and Giebel, 2004; Jung et al., 2005; Saey et al., 2009). Other researchers carried out on-line measurements to assess the three previous soil properties with R^2 values ranging between 0.40 and 0.70 for the first property (Triantafilis et al., 2002; Corwin and Lesch., 2003; Arriola-Morales et al., 2009), between 0.23 and 0.70 for the second property (Sudduth et al., 2005; Hezarjaribi and Sourell., 2007) and between 0.47 and 94 for the third property (Kitchen et al., 1996; Sudduth et al., 2005).

4.1.1.2 Ground penetrating radar (GPR)

Basically GPR is a geophysical technique that allows to image soil in two or three dimensions reaching to several meters in depth with high spatial resolution. The working principle of GPR is similar to reflection seismic and sonar techniques (Davis and Annan, 1989). Electromagnetic (EM) waves are transmitted towards the soil and from the reflections of this wave; properties of the soil can be extracted. The theoretical aspects of radar

components and their working principles can be found in detail in Daniels (2007) and Jol (2009). GPR systems work in a frequency range of 10–5000 MHz (e.g., VHF-UHF). The main characteristics of a GPR system are its operating frequency (centre frequency), resolution, and depth of penetration. The GPR resolution is the ability of the system to distinguish two signals that are close to each other in time. Usually, the resolution of a GPR increases with increasing operating frequency (Davis and Annan, 1989; Huisman et al., 2003). As the penetration depth reduces with increasing frequency, the choice of an operating frequency is always a trade-off between resolution and penetration depth, as higher frequencies permit higher resolution but lower penetration depth (Davis and Annan, 1989). The depth range of GPR is also strongly influenced by the electrical conductivity of the soil. GPR is a very promising tool for imaging primarily the subsurface features (Annan, 2002; Van den Bosch et al., 2006). GPR can be used to measure different soil properties such as moisture content (e.g., Chanzy et al., 1996; Van Overmeeren et al., 1997; Weiler et al., 1998; Minet et al., 2010), soil salinity (e.g., Al Hagrey and Müller, 2000), soil compaction (Petersen et al., 2005), water table (e.g., Smith et al., 1992).

4.1.2 Passive radiometric sensing

It is commonly known that all objects above the temperature of absolute zero (-273.15C) radiate EM waves to their surrounding environment. These emissions are of different types identified on the basis of wavelength, for example, ultraviolet, visible, infrared, radio waves, gamma rays, etc. In the following section gamma-ray spectrometers will be described briefing their potential use in precision agriculture and soil science.

4.1.2.1 Gamma-ray spectrometers

Gamma ray (γ-ray) spectrometry is a relatively new soil sensing technique that can be used to carry out measurements from either ground or airborne platforms. The spectrometer measures natural γ-emissions from emitters, such as ^{40}K, the daughter radionuclides of ^{238}U and ^{232}Th, and total emissions from all elements. Association of these radionuclides with clay, gravels, and soil-forming materials leads to variations in the concentrations of these emitters that can be used to estimate soil properties such as clay, potassium, organic carbon and iron contents, soil depth, and soil pH (Viscarra Rossel et al., 2007; Wong and Harper, 1999; Wong et al., 2008, 2009).

The presence of radioisotopes such as K, Th, and U in soils and rocks is associated with certain constituents. Gamma rays emitted from the surface are related to the mineralogy and geochemistry of the bedrock and weathered materials, for example, soils, saprolite, alluvial, and colluvial sediments. Understanding the bedrock and regolith responses has proven invaluable not only for mapping regolith materials but also for understanding geomorphic processes (Wilford, 2002; Wilford et al., 1997). Wilford and Minty (2006) explained briefly how these radioisotopes occur in rock minerals. For instance, the concentration of K, Th, and U contents in soils and rocks generally increases with increasing silica content. The concentration of K decreases with increased weathering. This is because K is soluble under most weathering environments and tends to be leached from a soil/regolith profile. On exceptional occasions, the K is incorporated into potassic clays such as illite. Otherwise, it is either absorbed onto clays such as montmorillonite and kaolinite, or associated with either large K-feldspar phenocrysts or mica that take time to weather. In contrast, U and Th are associated with more stable weathering products in soil profiles. U and Th released during weathering are readily absorbed onto clay minerals, Fe, Al oxyhydroxides, and organic matter in soils. In addition, U and Th also reside in resistate minerals that persist for a long time in the soil. It is therefore not uncommon for relative concentrations of U and Th to increase in highly weathered soils, as other more soluble minerals are lost in solution (Wilford and Minty, 2006). Detailed information about the geological and geochemical laws governing the behavior of radioisotopes in radiometrics can be found in Dickson and Scott (1997) and Hyvönen et al., (2005).

In soil science applications, portable gamma radiometers gained interest during the past decade for mapping individual soil properties (Pracilio et al., 2005, 2006; Wong and Harper, 1999). These ground-based gamma spectrometers were used to estimate soil texture (Roberts et al., 2003; Taylor et al., 2002; Viscarra Rossel et al., 2007; Mahmood et al., 2011), plant available K (Wong and Harper, 1999), and other minerals (Van Egmond et al., 2010; Viscarra Rossel et al., 2007). To incorporate the other useful information together with region of interest (ROI), a full spectrum analysis is recommended and being used for minimal loss of information from gamma spectra (Hendriks et al., 2001; Viscarra Rossel et al., 2007). The ground-based gamma spectrometers are used as an on-line system to measure gamma counts. There are some small handheld gamma spectrometers to be used in situ as well as in the laboratory.

4.1.3 Soil strength sensors

Soil strength changes with time under influence of climate, soil management, and plant growth (Koolen and Kuipers, 1983). Soil mechanical properties under external loads of tillage tools, and agricultural machinery are influenced by several factors including bulk density (BD), MC, SOM, and soil texture type (Mouazen et al., 2002). Methods for the measurement of soil strength include laboratory, in situ, and on-line measurement techniques. Soil shear strength is soil resistance to deformation by applied external shear forces, for example, during soil cutting process with different tillage tools. Shear failure occurs when shear forces exceed a maximum limit called yield strength. However, shear failure can also occur under compression load, for example, under tyres, which make soil as a bulk material to behave differently than metals under compression load (McKyes, 1989). The soil shear strength is represented as sum of soil cohesion (C) and internal frictional angle (j). Cohesion is, contrary to friction, independent of loading on soil particles. By determining the maximum shear stress at corresponding normal stresses, one can determine cohesion and internal friction angle. Some methods for measuring soil shear strength are explained below.

4.1.3.1 Laboratory measurement methods of shear strength

4.1.3.1.1 Direct shear box

When carrying out a test, each sample is placed within two square rings. During each test, a normal pressure is applied to the upper part of the soil, while the bottom part is moved horizontally. The relative displacement versus shear force is recorded and the soil shear strength properties are estimated based on Coulomb's criterion.

4.1.3.1.2 Triaxial compression test

The triaxial compression apparatus allows remolded or undisturbed soil samples to be tested. Remolded soil specimens are prepared in cylindrical shape after controlling the BD and MC. The soil cylinders are then surrounded by a rubber membrane and confined by water pressure in a water-filled load cell. During the tests, an axial principal stress is generated on the top of the cylinder. Differences between lateral and axial principal stresses generate shear stresses on various planes in the soil cylinder. The cylindrical sample will start to deform after a certain axial displacement continuing with changes in volume and shape. Using a Mohr's Circle diagram C and j can be determined (Koolen and Kuipers, 1983).

4.1.3.2 In situ measurement methods

4.1.3.2.1 Shear methods

Soil shear strength can also be determined in situ by means of torsional shear box, annular grouser plate and shear vane. They are explained in details by Gill and Vandenberg (1967). These methods basically apply similar measuring principles as those of the direct shear box. However, soil is being sheared by means of forces applied by rotation instead of transitional horizontal displacement of the direct shear box.

4.1.3.3 Draught sensors

Draught of a soil cutting tool is commonly used to map soil resistance, incorrectly referred to as compaction. Draught is measured by commercially available or specifically designed load cells or strain gauges. Godwin (1975) used an extended octagonal ring transducer (EORT) to measure tillage forces. Richards (2000) used an EORT to measure draught of a tine and to map soil resistance. The results showed that soil texture type and other soil physical properties were not correlated to draught and that the soil type variability throughout the field could not be predicted using the draught force results. Al-Janobi (2000) combined an EORT with data logging system to measure and record on-line draught. The on-line measured draught only showed soil resistance variability and no correlation with soil physical properties. Mouazen et al., (2003) carried out draught measurements with a soil sensor to measure BD as an indicator of soil compaction. They used a commercially available 5 ton single ended shear beam load cell. Mouazen and Ramon (2006) found draught alone cannot be used to produce a map of within field variability of soil compaction. According to their findings, it is necessary to measure other influencing parameters during the on-line measurement of soil compaction, and they have proven the need for a model of soil BD indicating soil compaction as a function of draught, MC, and depth.

Although laboratory methods for the measurement of soil strength are time-consuming, they provide essential information for soil and land management. In situ measurement methods can easily be used but their main flaw is that they were not developed to account for all parameters affecting soil strength measurement, namely, MC, BD, SOM, and texture. Therefore, when spatial variation in field soil compaction has to be assessed, any soil strength measurement must be accompanied with measurements of the other influencing parameters, unless there is proof of spatial uniformity of any of these parameters (Kuang et al., 2012).

4.1.4 Electrochemical based sensors

Electrochemical sensors have been developed which can provide quick information of nutrient status and pH in the soil, not only for laboratory use but also for in situ or on-line field measurement. Among various classes of electrochemical sensing methods, ion-selective electrodes (ISEs) and ion-sensitive field-effect transistors (ISFETs) are the most frequently used potentiometric sensors. Both, ISEs and ISFETs measure a voltage difference between sensing and reference parts of the system, which is directly related to the concentration and the activity of specific ions such as H^+, K^+ , NO_3^- , Na^+ , etc.

4.1.4.1 Ion-selective electrodes

An ISE is defined as an electro-analytical sensor of the activity of a specific ion in a solution. This activity is converted into an electrical potential, which can be measured by a voltmeter. The sensing part of the electrode is usually made as an ion-specific membrane, along with a reference electrode. Theory, principles, and applications of ISEs are discussed in detail in Pungor (1998, 1999, 2001), Birrell and Hummel (2000, 2001), Adamchuk et al., (2005), Kim et al., (2006), and Sinfield et al., (2010). ISEs are ion specific, each needing its own membrane permitting the passage of certain ions only. An electrical potential is established between two electrodes placed in contact with a moist soil sample or liquid soil solution. Since nitrate ISEs are highly selective to NO_3 ions in solution they were first used around 1967 as quick and reliable alternatives to chemical-based laboratory methods for nitrate measurements (Dahnke, 1971). The pH electrode is the most well-known and simplest type of ISE measuring the concentration of hydrogen ions [H+].

4.1.4.2 Ion-sensitive field-effect transistors

ISFETs are based on the same chemical principle as ISEs. The main technical difference between ISFETs and ISEs is that ISFETs do not contain an internal solution and the ion-selective membrane is affixed directly on the gate surface of the ISFET (Birrell and Hummel, 2000; Sinfield et al., 2010). Bergveld et al., (1998) have presented a review of theory of ISFET sensors.

ISEs and ISFETs can be affected by two factors: first—disturbances from the measurement system, for example, membrane, reference electrode, amplifier, etc.; second-soil factors hindering the attainment of electrochemical equilibrium in the measuring cell. The role of second type of factors is least studied in the literature due to complexity of measuring

media. Electrode aging and mechanical wear of the sensitive membrane may result in significant changes in output. Periodic calibration in solutions with known ion activity is needed.

Interference from other undesired ions is limiting the use of ion-selective electrodes. ISEs are not completely ion-specific but are sensitive to other ions having similar physical properties. The relative sensitivities of each type of ion-specific electrode to various interfering ions are generally known but the degree of interference depends on many factors, preventing precise correction of readings. For instance, the nitrate electrode has various ionic interferences, that is, perchlorate, iodide, chloride, and sulfate. Due to key features of ISFETs, their application as potentiometric sensors has great advantages over conventional ISEs. For instance, small size and a solid state nature, low output impedance that reduces interference from external EM fields, mass fabrication and low cost, the possibility of integrating compensation and data processing circuits in the same chip with the sensor, high signal-to-noise ratio, low sample volume, and a short response time. ISFETs might be integrated with a flow injection analysis (FIA) system for real-time soil analysis. In contrast to many other analytical methods, ISEs are capable of determining ion activities, rather than total concentration. The analyte is also not consumed in the course of the measurements, and instruments are less expensive compared to the ones needed in other methods.

4.1.5 Reflectance sensors

To generate a soil spectrum, radiation containing all relevant frequencies in the particular range is directed to the sample. Depending on the constituents present in the soil the radiation will cause individual molecular bonds to vibrate, either by bending or stretching. These vibrations lead to absorption of light, to various degrees, with a specific energy quantum corresponding to the difference between two energy levels. As the energy quantum is directly related to frequency, the resulting absorption spectrum produces a characteristic shape that can be used for analytical purposes (Stenberg et al., 2010). The fundamental vibrations in the mid-infrared (MIR) region result in overtones and/or combinations in the near infrared (NIR) region. In the visible (vis) range (400–780 nm), absorption bands related to soil colour are due to electron excitations, which assist the measurement of soil organic matter content (SOM) and moisture content (MC). However, in the NIR range, the overtones of OH and overtones and/or combinations of C–H, C–C, OH^+ minerals, and N–H are

important for the detection of SOM, MC, clay minerals, and nitrogen (Mouazen et al., 2010), with all having direct spectral responses in the NIR spectral range.

4.1.5.1 Visible and near infrared spectroscopy sensors

When NIR radiation interacts with a soil sample, it is the overtones and combinations (Figure 2) of fundamental vibrations in the mid-infrared (mid-IR) region that are detected. Molecular functional groups can absorb in the mid-IR, with a range of progressively weaker orders of overtones detected in both the mid-IR and NIR regions. Generally, the NIR region is characterized by broad, superimposed, and weak vibrational modes, giving soil NIR spectra few, broad absorption features. In the visible region, electronic excitations are the main processes as the energy of the radiation is high (Stenberg et al., 2010).

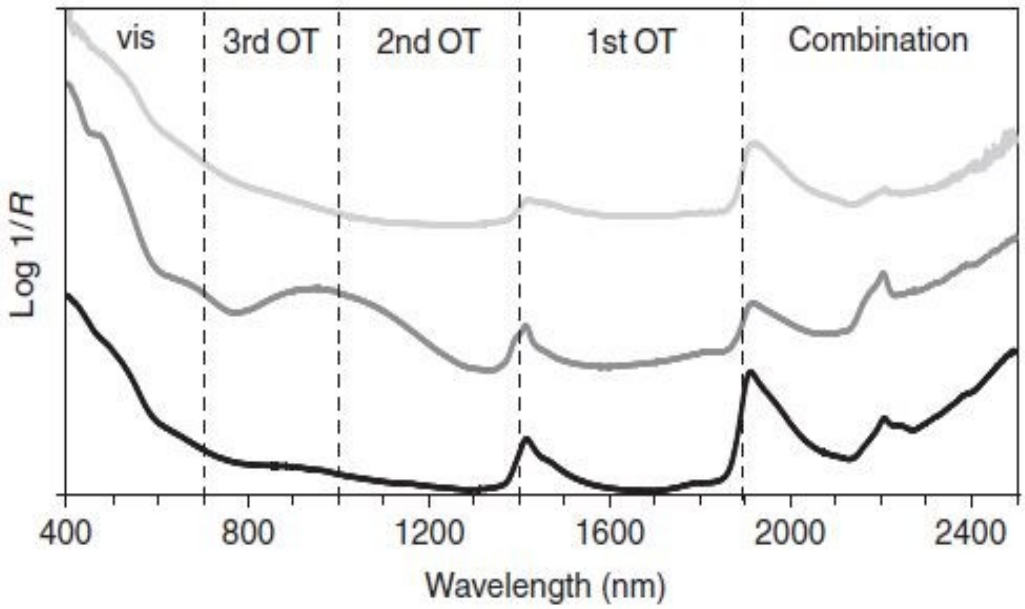

Figure 2. Soil visible and near infrared (vis–NIR) 400–2500 nm spectra showing approximately where the combination, first, second, and third overtone (OT) vibrations occur, as well as the visible (vis) range (Stenberg et al., 2010).

Due to the broad and overlapping bands, vis–NIR spectra contain fewer absorption bands than the mid-IR and can be more difficult to interpret. Nevertheless, this region contains useful information on organic and inorganic materials in the soil. Absorptions in the vis region (400–780 nm) are primarily associated with minerals that contain iron (e.g., haematite, goethite) (Mortimore et al., 2004; Sherman and Waite, 1985). SOM can also have broad absorption bands in the vis region that are dominated by chromophores and the

darkness of organic matter. Absorptions in the NIR region (780–2500 nm) result from the overtones of OH, SO_4, and CO_3 groups, as well as combinations of fundamental features of H_2O and CO_2 (e.g., Clark, 1999). Clay minerals can show absorption in the vis–NIR region due to metal-OH bend plus O–H stretch combinations (Viscarra Rossel et al., 2006). Carbonates also have weak absorption peaks in the near infrared (Hunt and Salisbury, 1970). Water has a strong influence on vis–NIR spectra of soils (Mouazen et al., 2005). The dominant absorption bands of water around 1450 and 1950 nm are characteristic of soil spectra, although there are weaker bands in other parts of the vis–NIR range (Liu et al., 2002), e.g. at 950 nm.

4.1.5.1.1 Past and current role of vis-NIR in soil science

Over the last few decades a large number of attempts have been made to predict soil properties with vis–NIR spectroscopy. Calibrations for total and organic carbon are probably most frequent, followed by clay content. According to a review published by Viscarra Rossel et al., (2006), these two properties, together with total soil N, are also those with the best chance of success. This makes sense because both clay minerals and SOM are the fundamental constituents of the soil and have well-recognized absorption features in the vis–NIR region. Some other frequently reported properties include pH, extractable P, K, Fe, Ca, Na, Mg, and CEC, as well as properties that are dependent on combinations of other soil properties, such as lime requirement and mineralisable N (Stenberg et al., 2010). Results for these properties with indirect spectral responses in the NIR range are typically moderate and often highly variable. This makes sense as the covariations to constituents that are spectrally active cannot be expected to be globally stable.

Literature proves moisture content to be the most accurately measured property with NIR with excellent accuracy (Chang et al., 2001; Mouazen et al., 2006a) which can be attributed to the presence of the clear water absorption band in the second overtone region at 1450 nm, which resulted in a large correlation around 1450 nm. Clay content was also reported to be accurately measured with NIR, which is attributed to the direct spectral response of clay minerals around 2300 nm (Viscarra Rossel et al., 2006).

4.1.5.1.1.1 Soil heavy metals

Pure metals do not absorb in the vis–NIR region (Stenberg et al., 2010). However, they can be detected because of covariation with spectrally active components. For example, they

can form complexes with organic matter, associated with hydroxides, sulfides, carbonates, or oxides that are detectable in the vis–NIR, or adsorbed to clay minerals. Malley and Williams (1997) first attempted predictions of heavy metals in freshwater sediments using vis–NIR (1100–2500 nm). They reported R^2 values between vis–NIR-predicted and chemically-analyzed metal concentrations of 0.63 for Cd; 0.91 for Cu; 0.93 for Zn; 0.81 for Pb; 0.88 for Ni; 0.93 for Mn; and 0.86 for Fe. The authors attributed most of the variance in heavy metal concentrations to organic matter content. Kooistra et al., (2001) researched the use of vis–NIR spectroscopy for the assessment of soil Cd and Zn contamination in river floodplains and reported good predictions, with R^2 of 0.94 for Cd and 0.95 for Zn. They attributed these good results to the associations of the metals with SOM and clay content, which are spectrally active in vis–NIR. Although there are other studies on the use of vis-NIR spectroscopy for the detection of metals in soils, they are not referred in detail since they are not of core importance for the current work.

4.1.5.1.1.2 Soil mineralogy

Soil minerals generally account for half of the soil volume (Schulze, 2002). Their type, proportions and concentrations ultimately determine important properties such as texture, structure, and CEC. These properties may in turn have a significant effect on many other soil properties. For example, potassium availability for plant uptake is dependent on its release from the weathering of primary soil minerals. Soil minerals absorb light in the UV, vis, NIR, and mid-IR portions of the electromagnetic spectrum. Iron oxides absorb strongly in the UV and absorb weakly in the vis–NIR region, while clay minerals such as phyllosilicates have distinct spectral signatures in the vis–NIR region. Comprehensive accounts of the processes that produce these absorptions can be found in Hunt (1977), and Clark (1999). Figure 3 shows the continuum-removed reflectance spectra of these minerals.

The reflectivity of goethite (α-FeOOH) is relatively high at longer wavelengths in the vis–NIR region from the absorption band that occurs near 930 nm. Three other absorption bands for goethite are discernible in the vis–NIR region, one near 660 nm, another near 480 nm and one near 420 nm. The spectrum of goethite also shows weak absorption near 1700 nm, which is due to the first overtone of a stretching vibration of OH that is present in the crystal structure of goethite (Morris et al., 1985). The spectrum of haematite (α-Fe_2O_3) is characterized by high and nearly constant reflectivity at longer wavelengths in the vis–NIR, a

reflectivity minimum near 880 nm, a shoulder centred near 620 nm and a band with very low reflectivity near 510 nm.

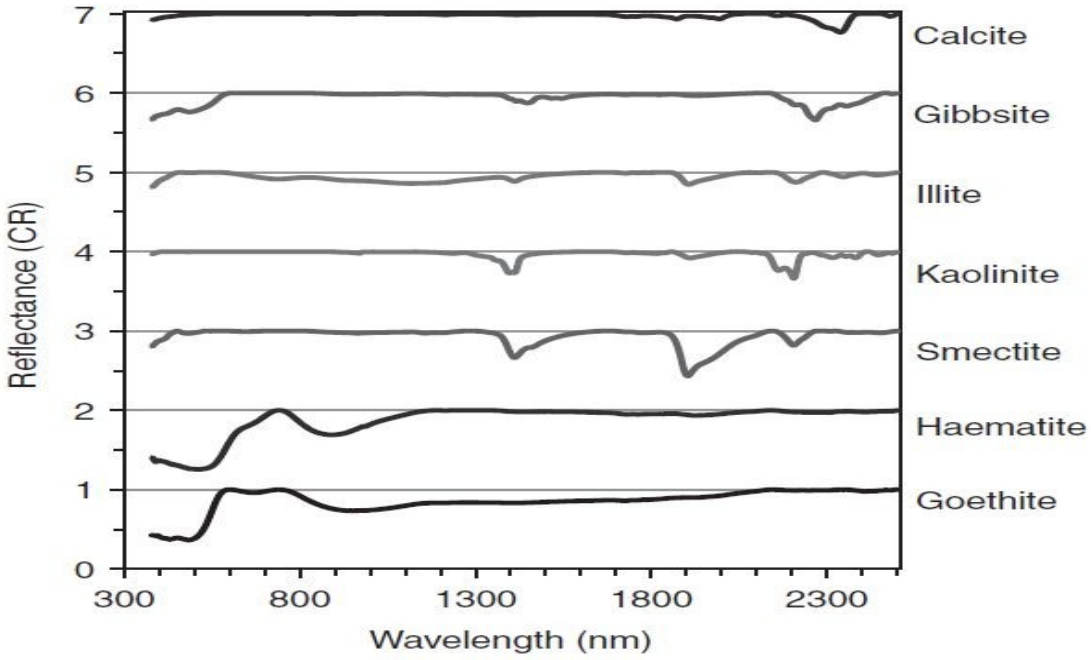

Figure 3. Continuum-removed spectra of common soil minerals offset by 1 unit for each spectrum (Stenberg et al., 2010).

In both goethite and haematite the absorptions near 930 and 880 nm, respectively, may be assigned to ligand field transitions that involve excitations from a ground state to the first higher energy state (Sherman and Waite, 1985). Their absorptions near 660 nm and the shoulder near 620 nm may also be assigned to charge transfer absorptions producing transitions from a ground state to a higher energy state. Assignment of bands near 480 nm and 420 nm in goethite and near 510 nm in haematite are generally attributed to the absorption edges (or wings) of intense charge transfer absorptions that occur in the UV (Sherman and Waite, 1985). Absorption in the vis range causes the vivid colours of iron oxides, for example, red haematite and yellow goethite.

Combination vibrations involving O–H stretch and metal–OH bend occur in the 2200–2500 nm region. It is generally understood that absorptions near 2200 nm are due to Al–OH, as in kaolinite, montmorillonite, and illite. But if the absorption is near 2290 nm it is due to Fe–OH and if near 2300 nm it is due to Mg–OH in, e.g., illites and montmorillonites (Clark et al., 1990; Post and Noble, 1993). There are exceptions to this, e.g. in gibbsite ($AlOH_3$) the combination absorption occurs near 2268 nm instead of 2200 nm.

Clay mineral absorptions are mostly due to OH, H_2O, and CO_3 overtones and combination vibrations of fundamentals that occur at longer wavelengths in the mid-IR region. Kaolin has characteristic absorption doublets near 2200 nm and 1400 nm. The absorption wavelengths near 1400 nm (1395 and 1415 nm) are due to overtones of the O–H stretch vibration near 2778 nm (3600 cm $^{-1}$), while those near 2200 nm (2165 and 2207 nm) are due to Al–OH bend plus O–H stretch combinations.

Smectite has strong characteristic absorptions near 1400, 1900, and 2200 nm. The band near 1400 nm can to one part be attributed to the first overtone of structural O–H stretching mode in its octahedral layer. The 1400 nm and 1900 nm bands are also due to combination vibrations of water bound in the interlayer lattices as hydrated cations and water adsorbed on particle surfaces (Bishop et al., 1994). Such water is not present in kaoline supporting a diagnostic feature for dry kaolinitic soils that will absorb very weakly near 1900 nm. The combination bands that are due to vibrations of bound water occur at slightly shorter wavelengths near 1400 nm and 1900 nm, while those of adsorbed water appear as shoulders near 1468 nm and 1970 nm. By remoistening, these shoulders will dominate (Bishop et al., 1994). Illite has absorptions near 1400, 1900, and 2200 nm too, but generally weaker than smectite. Illite also has additional absorptions near 2340 nm and 2445 nm (Post and Noble, 1993). These bands may diagnostically distinguish between illite and smectite. They are, however, weak and especially the former may be confused with organic matter absorption. The previous review about soil mineralogy indicates the importance of vis-NIR spectroscopy to detect the type of clay minerals, as their assessing is time-consuming, expensive and need expert operators to carry it out.

4.1.5.1.1.3 Soil moisture content

Apart from the obvious point of water being a dominant regulator of plant growth and soil biology as a whole, more specifically it regulates important soil processes such as nitrification/denitrification and hydrological processes such as leaching and erosion. Various forms of water absorb strongly in the near infrared region. This is due to overtones and fundamentals of the three fundamental vibration frequencies of H_2O; symmetric and asymmetric O–H stretching and O–H bending (Hunt, 1977). Water incorporated into the lattice of some clay minerals absorbs strongly near 1400 and 1900 nm and is obviously directly related to the mineralogy of the sample (Stenberg et al., 2010). Water adsorbed as a thin layer onto exposed surfaces and free liquid water filling pore spaces has the general effect

of decreasing the albedo. This effect can be largely attributed to a change in the real refractive index of the medium surrounding the soil particles from that of air to that of water, which is higher and closer to soil particles. Mouazen et al., (2005) measured the soil moisture content using a NIR on-line sensor. They used partial least squares analysis in order to establish a statistical model relating soil light spectra with gravimetric moisture content in the 0.005–0.26 kg kg^{-1} range. This model was validated with the full cross-validation method resulting in a small root mean square error of cross validation (RMSECV) of 0.0175 kg kg^{-1} and a high validation correlation (r) of 0.978. Further validation of the model developed in the laboratory under stationary state showed also a small root mean square error of prediction (RMSEP) of 0.0165 kg kg^{-1} and a prediction correlation of 0.982. When the NIR sensor-model system was used to determine moisture content, based on on-line field measurement, a relatively larger RMSEP of 0.025 kg kg^{-1} and lower prediction correlation of 0.75 were found. However, a reasonably similar spatial distribution of moisture content was found between the on-line NIR measurement and oven drying methods. In another study, Mouazen et al., (2007) using a vis-NIR on-line sensor reported a good results (RPD=3, RMSEP= 0.024 kg kg^{-1}) of the developed PLSR model coupled with cross-validation. Also, they obtained a small average error of 0.374% between the on-line and oven dry methods which is a good indication of the accuracy obtained. This indicates the potential of the on-line measurement system h to provide accurate detailed data about soil moisture content, needed not only for on-line detection of soil compaction but for variable seeding and irrigation rates in the arid and semi-arid areas.

4.1.5.1.1.4 Soil organic matter (SOM) and soil total nitrogen

SOM, often approximated to 1.72 times soil organic carbon (SOC), is the property most frequently estimated by vis–NIR spectroscopy. Overtones and combination bands in vis–NIR due to organic matter result from the stretching and bending of NH, CH, and CO groups (Ben-Dor et al., 1999; Bokobza, 1998; Goddu and Delker, 1960). Bands around 1100, 1600, 1700 to1800, 2000, and 2200 to 2400 nm have been identified as being particularly important for SOC and total N calibration (Ben-Dor and Banin, 1995; Dalal and Henry, 1986; Henderson et al., 1992; Krishnan et al., 1980; Malley et al., 2000; Martin et al., 2002; Morra et al., 1991; Stenberg, 2010). Mouazen et al., (2007) used a vis-NIR on-line sensor to predict soil organic and total carbon and obtained a good calibration model with RPD values of 1.97 and 1.92 for soil organic carbon and total carbon respectively. Values of the root mean square error of prediction were of 0.480 % and 0.268 % for both forms of soil carbon mentioned above. Both

carbon and nitrogen have direct spectral responses in the NIR region, which can be attributed to overtones and combinations of N–H, C–H + C–H and C–H + C–C. Therefore, successful measurement of these two properties with NIR is expected. However, C and N have different forms in the soil, such as, mineral, organic, and inorganic, which have an influence on the accuracy. Chang et al., (2001) reported successful estimation of TC, TN, and MC with R^2 > 0.84 and RPD more than 2.47. Reeves and McCarty (2001) also stated successful estimation of TC and TN with R^2 of 0.92, 0.90 and root mean square error (RMSE) of 0.15% and 0.0132%, respectively. Nevertheless, mineral forms of nitrogen as ammonia and nitrate are very difficult to be measured by NIR (Stenberg et al., 2010).

4.1.5.1.1.5 Plant nutrients

Due to their often direct relationship to plant nutrition, plant nutrients such as N, P, K, Fe, Ca, Na, Mg, and methods for their measurement attract much interest in agriculture (Stenberg et al., 2010). In PA practices such as variable rate fertilization, data on plant-available P and K, which are fairly stable parameters over time, are important for high resolution soil mapping. Nitrogen is by far the most important nutrient in most agricultural systems. Plant N uptake occurs mainly in the form of nitrate or ammonium, but as these sources are very dynamic, estimates of their concentrations are very variable. Plant nutrients are not expected to have direct spectral absorption features in the vis–NIR region. Correlations found to vis–NIR spectra are often weak, but there are exceptions, for example, Chang et al., (2001); Ehsani et al. (1999); Groenigen et al., (2003); Krischenko et al., (1992); Moron and Cozzolino (2003); Mouazen et al., (2006; 2007); Pereira et al., (2004); Shibusawa et al., (2001); Udelhoven et al., (2003). These authors report highly variable coefficients of determination (R^2; in parentheses) for mineral N (0.20–0.99), available K (0.56–0.83), exchangeable K (0.11–0.55), Ca (0.75–0.89), Fe (0.64–0.91), Na (0.09–0.44), Mg (0.53–0.82), and P (0.23–0.92). The occasionally successful calibrations may be attributed to locally present covariation to spectrally active constituents. Such covariations may of course vary between data sets. The potential for vis–NIR to predict extractable P in soil has been relatively well studied as P is the second most important plant nutrient after N and is as well a limited natural resource. Results reported for P are, as indicated above, among the most variable. Some authors report successful calibrations and others absolute failure. Udelhoven et al., (2003) failed to predict CAL-extractable P (Schuller, 1969) at the regional scale, but at the field scale it was fairly well predicted. Mouazen et al. (2009) conducted on-line measurement

of available P with remarkable accuracy (RPD = 1.42; R^2 = 0.62). Shaddad et al., (2013) reported good on-line prediction of extractable P with RPD and R^2 values of 1.72 and 0.64.

4.1.5.1.2 Nonmobile (in situ) field visible and near infrared spectroscopy

Although the application of vis–NIR spectroscopy has reduced considerably the labour and time for the analysis, soil sample preparation for laboratory analysis including drying, grinding, and sieving is still tedious. For in situ and on-line measurement with vis–NIR, calibration models developed from dried, ground, and sieved samples cannot be utilized, since measurement is performed with fresh soil samples. As early as more than two decades ago, using an integrating cylinder and two narrow band interference filters, Barrett (2002) developed a spectrophotometric colour measurement for in situ well-drained sandy soils, reporting a moderately strong correlation. Fystro (2002) confirmed the ability of vis–NIR spectroscopy for measurement of OC, TN, and their potential mineralization in grassland soil samples, arriving at moderate accuracy (R^2 > 0.7 and RPD > 1.5). Udelhoven et al., (2003) evaluated the ability of NIR spectroscopy to estimate soil Fe, Mn, Ca, Mg, and K, and they found that only Ca and Mg (R^2 = 0.67 and 0.69, respectively) were predictable under in situ conditions. Chang et al., (2005) attempted to predict TC, OC, TN, CEC, pH, texture, MC, and potential mineralizable N and indicated that NIR was able to measure these soil attributes with reasonable accuracy using fresh soils (R^2 > 0.74). Maleki et al., (2006) developed a calibration model of available P (Pavl) with acceptable prediction accuracy (R^2 > 0.73) based on fresh soil samples with the intention to be used for on-line variable rate P_2O_5 application system. Combining vis–NIR spectroscopy and laser induced breakdown spectroscopy (LIBS), Bricklemyer et al., (2005) reported moderate prediction accuracy (R^2 = 0.70) of TC and C_{inorg} under in situ conditions. Melendez- Pastor et al., (2008) identified optimal spectral bands to assess soil properties with vis–NIR radiometry in a semi-arid area and estimated SOM with worse accuracy (R^2 = 0.73, RPD = 1.92, and RMSEP = 0.52%) than generally reported under laboratory condition. Literature proves that laboratory vis–NIR methods provide better accuracy than *in situ* field measurement (Fystro (2002), Udelhoven et al., 2003; Mouazen et al., 2010; Kuang and Mouazen., 2011; Zornoza et al., 2008; Wetterlind et al., 2010), which can be attributed to the influence of MC and structure that were eliminated under laboratory conditions by drying, grinding, and sieving.

Successful measurements of soil properties without direct spectral response in the NIR range is due to co-variation through other properties that have direct spectral responses in the

NIR, for example, carbon, nitrogen, and clay (Stenberg et al., 2010). Literature shows that only few successful reports on phosphorous (P) determination by vis–NIR spectroscopy are available. Up to now, the most significant reports on successful measurement of P are those of Bogrekci and Lee (2005, 2005a). Bogrekci and Lee (2005) obtained probably the best R^2 value of 0.92 between P concentrations and spectral absorbance using a vis–NIR spectroscopy. Literature (e.g., Chang et al., 2001; Mouazen et al., 2006) proves that the worst properties to be measured with NIR are K and Na. Measurement of pH, Ca, and Mg were reported to be more successful as compared to K and Na (Kuang et al., 2012). Therefore, further research is recommended to understand and probably improve the calibration accuracy of soil properties without direct spectral responses in the NIR range.

4.1.5.1.3 Mobile (on-line) field vis–NIR sensors

So far there are only three on-line vis-NIR systems in the world. They are of Christy, (2008), Mouazen et al., (2005) and Shibusawa et al., (2001). The beginning of these systems dates back to 1991, when Shonk et al. (1991) developed a system to measure SOM and MC, which utilized a single wavelength (660 nm) of light. Shibusawa et al., (2001) developed an on-line vis–NIR (400–1700 nm) sensor to predict MC, pH, SOM, and NO_3–N. Although this system is highly technically instrumented, it is rather expensive. Christy (2008) developed a prototype soil reflectance mapping unit equipped with a vis–NIR spectrophotometer, which is commercially available in the market. The sapphire glass of the optical probe makes direct contact with soil and stones. A simpler design to the one of Shibusawa et al., (2001) without sapphire window optical configuration was developed by Mouazen et al., (2005). The system was successfully calibrated for MC, TN, TC, pH, and available P in different soils in Belgium and northern France (Mouazen et al., 2005, 2007, 2009).

Literature proves that both the laboratory and *in situ* non-mobile vis–NIR methods provide better accuracy than the on-line method (Shibusawa et al., 2001; Mouazen et al., 2007; Bricklemyer and Brown., 2010), which is attributed to other factors influencing the latter methods. These factors include among others noise associated with tractor vibration, sensor-to-soil distance variation (Mouazen et al., 2009) stones and plant roots and difficulties of matching the position of soil samples collected for validation with corresponding spectra collected from the same position. This point needs to be borne in mind throughout the manuscript, when evaluating accuracy achieved for different soil properties under consideration.

4.1.6 Geostatistics

Soil properties are continuous variables whose values at any location are expected to vary according to the direction and distance of separation from neighbouring samples (Burgess and Webster,1980). Therefore, the classic approach is inadequate for interpolation of spatially dependent variables, because it assumes random variation within units and no correlation between units and takes no account of relative location of samples. An alternative approach is to treat soil as a random function and to describe it using the methods of Matheron's regionalized variable theory (known as geostatistics) (Matheron, 1973). In this view there is no underlying mathematical relation between soil properties and their position on the ground. Even if there is, it is likely to remain unknown and in any case knowledge of it is unnecessary; relationships are instead expressed in terms of separation regardless of absolute position (Castrignanò, 2011). Geostatistics became quite well established in precision agriculture (PA) and the PA community has embraced geostatistics to explore the many kinds of data that farmers work with, mainly because the data are suitable for geostatistical analyses (Oliver, 2010).

Geostastistics is a wide-reaching field of spatial statistics, offering powerful tools for geo-spatial analysis. Most often it is used to interpolate estimates at locations where measurements have not or could not have been taken. As well as an interpolator, geostatistics provides a way of understanding spatial structure and can support the process of designing sample surveys.

There are different definitions of geostatistics. According to Deutsch, (2002), geostatistics is the study of phenomena that vary in space and/or time. Geostatistics can be regarded as a collection of numerical techniques that deal with the characterization of spatial attributes, employing primarily random models in a manner similar to the way in which time series analysis characterizes temporal data**.** (Olea, 1999). Geostatistics offers a way of describing spatial continuity of natural phenomena and provides adaptations of classical regression techniques to take advantage of this continuity (Isaaks and Srivastava, 1989). Geostatistics deals with spatially autocorrelated data.

Autocorrelation means correlation between elements of a series and others from the same series separated from them by a given interval (Oxford American Dictionary).

In order to understand and process data geostatistically, the following fundamental elements should be well known:

1- Spatial dependency (or spatial autocorrelation)
2- Semi-variograms (which are used to assess 'spatial dependency')
3- The three components of spatial variation:
    - Structural or deterministic variation
    - Stochastic and spatially correlated variation
    - Spatially uncorrelated 'noise' or random variation
4- Anisotropic variation (occurrence of directional spatial structures)
5- (co)kriging and its various alternative approaches

4.1.6.1 Spatial dependency

Spatial dependency describes the phenomenon where things that are close to one another are more likely to have similar values or properties than things that are further apart.

The theory of regionalized variables forms the basis of procedures for analysis and estimation of spatially dependent variables, known collectively as geostatistics. Geostatistics assumes that a spatial variation of any variable can be expressed as the sum of three major components. These are:

1) a deterministic component associated with a constant mean value or a long-range trend;
2) a spatially correlated random component;
3) a white noise or residual error term that is spatially uncorrelated.

A regionalized variable is a random variable varying in space. A regionalized variable $z(\mathbf{x})$ can be considered as a particular realization of a random variable $Z(\mathbf{x})$ for a fixed position $\mathbf{x}$ within the area. If all values of $Z(\mathbf{x})$ are considered at all locations within the area, $Z(\mathbf{x})$ becomes a member of an infinite set of random variables, called a random function $Z(\mathbf{x})$. All the random variables have the same probability distribution function $F(z)$, independent of $\mathbf{x}$.

4.1.6.2 Semivariogram

The Semivariogram is the pillar of geostatistics and is a way to assess spatial correlation in observations measured at sample locations. It is commonly represented as a graph that shows the semi-variance of all observation pairs sampled at a distance. Such a graph is helpful for building a mathematical model that describes the variability of the observations as a

function of separation distance. Modelling of such a relationship is called semivariogram modeling. It is used in applications involving estimating the value of a property at a new location. Semivariogram modeling is also referred to as variogram modeling.

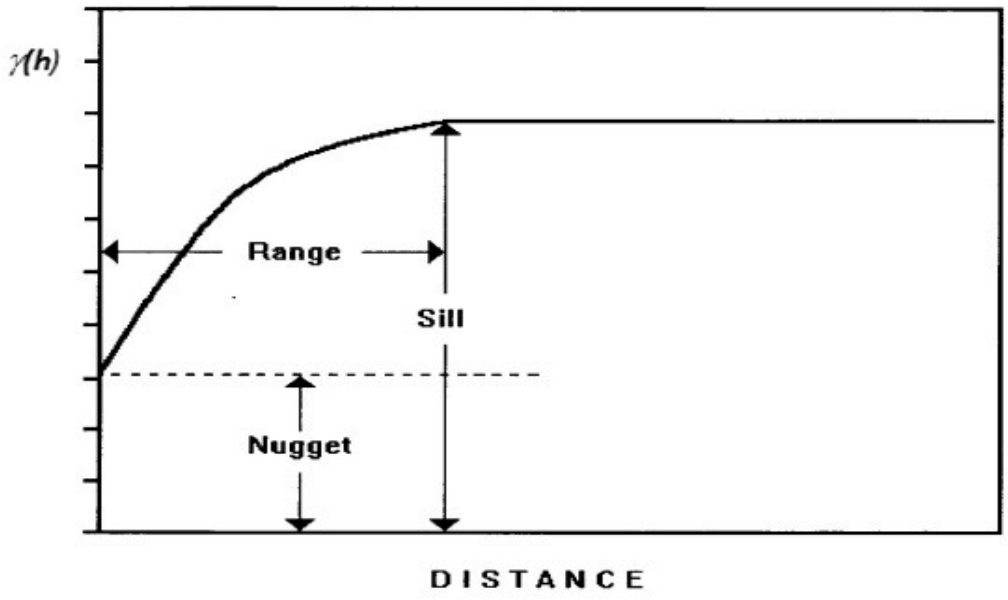

Figure 4. Theoretical semivariogram (Castrignanò, 2011)

There are three important features of an upper bounded semivariogram (Figure 4):

Range: describes the spatial limits of spatial dependency, or the distance beyond which points are spatially uncorrelated. If the distance between two adjacent observations exceeds this distance, then traditional (regressive) techniques of interpolation can be used.

Sill: describes the part of the variogram where it levels off at a distance equal to or greater than range.

Nugget: represents the spatially uncorrelated noise (including the variance of measurement error and micro- variation at a scale smaller than sampling scale).

4.1.6.3 Isotropic and anisotropic variation

Often there are directional effects in the spatial structure of observed measurements.

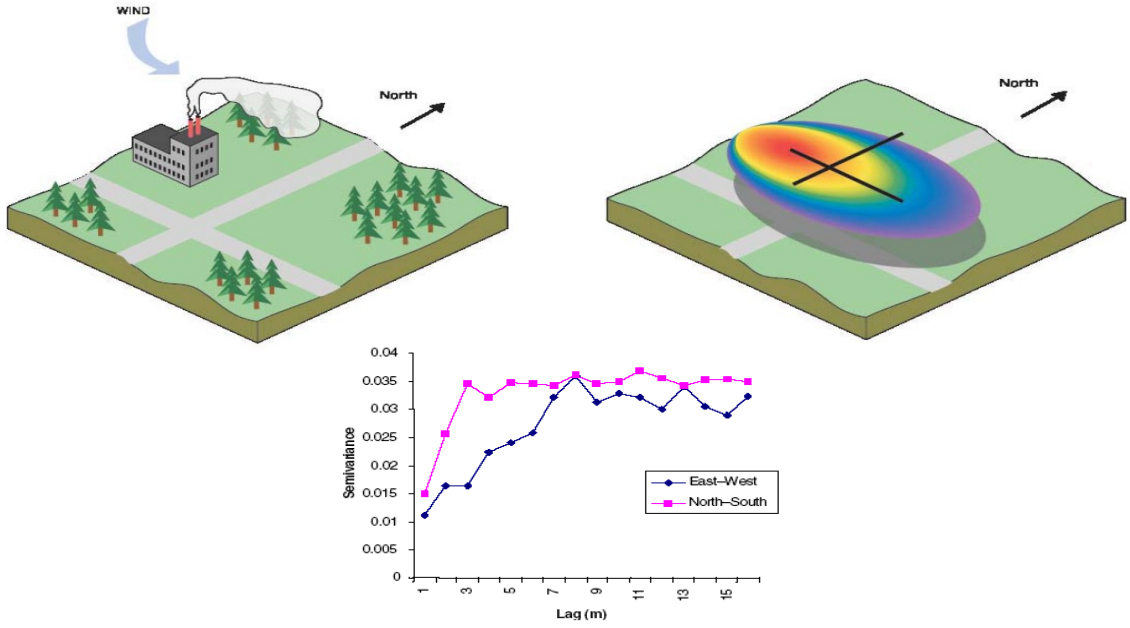

Figure 5. Isotropic and anisotropic variation

In Figure 5 a soil surveyor may go out and measure the sulphur content of the soil in the adjacent fields to the power station. Because of the emissions of sulphur from the chimney and the directional effect of the wind, the soil will comprise additional elements of sulphur that are likely to increase with distance from the power station, but only in the direction of the predominant wind direction.

This kind of variation is said to be anisotropic. If directional effects do not occur, the variogram is said to be isotropic. It is of importance to the surveyor to consider the use of directional variograms and explore spatial structure in a range of directions. An obvious starting point would be to consider the direction of the prevailing wind (in this case) and the perpendicular direction. The figure illustrates two different variogram structures.

The range of the north-south spatial dependency is over twice than in the east-west direction but the sills are approximately the same. This is a case of geometrical anisotropy. When the sill varies as a function of direction, the anisotropy is said to be of zonal type.

4.1.6.4 Kriging

Kriging is a type of interpolation technique. The procedure is similar to averaging techniques of interpolation but the weights are derived from the spatial arrangement as well as

from the distance between nearby points, i.e. from the variogram. The fitted variogram, or the directional variograms (for anisotropic variation) is/are used to determine the weights λ_i needed for local interpolation.

The weights are chosen so that the estimate is unbiased, and that the estimation variance is less than for any other linear combination of the observed values.

Mapping the spatial distribution of a soil property involves interpolation or spatial prediction. Geostatistical interpolation uses the variogram to optimize prediction by kriging. The most basic form of kriging is ordinary punctual kriging in which the unknown value $z(\mathbf{x}_0)$ of a given realization of $Z(\mathbf{x}_0)$ in an unsampled point $\mathbf{x}_0$ is predicted from the known values $z(\mathbf{x}_i)$ *i= 1, 2, . . ., N*, at the support points $\mathbf{x}_i$ of the same realization using a so-called "best linear unbiased estimator" (BLUE). The best linear unbiased predictor $z^*(\mathbf{x}_0)$ of $Z(\mathbf{x}_0)$ is given by a linear combination of the observations:

$$z^*(\mathbf{x}_0) = \sum_{i=1}^{N} \lambda_i z(\mathbf{x}_i)$$

Where λ_i are weights. The weights are chosen in such a way that the estimator is unbiased:

$$E[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] = 0$$

and the estimation is minimized.

Using a Lagrangian multiplier μ, minimization of the estimation variance under the constraint of unbiasedness yields a set of N+1 linear equations:

$$\begin{cases} \sum_{j=1}^{N} \lambda_i \gamma(\mathbf{x}_i, \mathbf{x}_j) + \mu = \gamma(\mathbf{x}_i, \mathbf{x}_0) & i = 1,\ldots,N \\ \sum_{j=1}^{N} \lambda_j = 1 \end{cases}$$

from which the λ_j and μ can be calculated. The estimation variance is then given by:

$$\sigma^2(\mathbf{x}_0) = \mu + \sum_{i=1}^{N} \lambda_i \gamma(\mathbf{x}_i, \mathbf{x}_0)$$

Or equivalently:

$$\sigma^2(\mathbf{x}_0) = 2\sum_{i=1}^{N}\lambda_i\gamma(\mathbf{x}_i - \mathbf{x}_0) - \sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j\gamma(\mathbf{x}_i - \mathbf{x}_j)$$

and represents the uncertainty in the prediction in: $\gamma(\mathbf{x}_i, \mathbf{x}_j)$ and $\gamma(\mathbf{x}_i, \mathbf{x}_0)$ are the semi-variances between the observed locations $\mathbf{x}_i$ and $\mathbf{x}_j$ and between the observed location $\mathbf{x}_i$ and the interpolated location $\mathbf{x}_0$, respectively. In the case of spatial dependence, semi-variance tends to increase with the distance between observations, therefore errors decrease with the density of data and not just with their total number, as is the case with traditional statistical models. It needs to be pointed out that kriging is optimal and unbiased only on the condition that the model is correct. However, predictions are only slightly affected by the choice of the model, provided it is reasonable of course. This is one of the strengths of kriging that is robust enough in this sense; however, error variances can be seriously affected by the model.

Kriging has many useful properties:

- The interpolated value is the most precise in terms of mean squared error
- The interpolated value can be used with a degree of confidence, because an error term is calculated together with the estimation
- The estimation variance depends only on the semi-variogram model and on the configuration of the data locations in relation to the interpolated point and not on the observed values themselves
- The conditions of unbiasedness ensure that kriging is an exact interpolator, because the estimated values are identical to the observed values when a kriged location coincides with a sample location. In this case the weights within the neighbourhood are all zero and the estimation variance equals the observation.
- Generally only the nearest few samples are spatially correlated to the kriged location and therefore they are the most weighted. Two important advantages become clear: firstly, the variogram needs to be accurate only in the first few lags; secondly whatever is gained from introducing distant points is limited also because sample locations interposed between the kriged point and more distant samples screen the distant ones by reducing their weights.
-

4.1.6.5 Geostatistics and precision agriculture

The principle of precision agriculture (PA) has an even longer history than geostatistics. It has been carried out by farmers since the early days of agriculture. They divided their landholdings into smaller areas, fields, to grow crops where the conditions were most suitable (Oliver, 2010).

Modern PA appears to have its origins in the mid 1970s to early 1980s; farmers were becoming more aware of the potential benefits of better farm record keeping and understanding of soil and crop input requirements (Robert, 1999). There was a better awareness of within-field variation in the properties of soil and crop, and of the potential benefits of within field management by zones. Associated with this there was a mushrooming of technology and services in response to the needs of this approach to agriculture, e.g. yield monitors and variable-rate spreaders. The concept of modern precision agriculture has also been driven forward and is underpinned by technological changes based on information technology (Schueller, 1997). One of the most significant steps was the introduction of a yield meter by Massey Ferguson in 1982. This device was mounted on a clean grain elevator of one of their combine harvesters. It meant that yield could be recorded continuously for the first time, but at that time without the benefit of GPS. In the 1990s with the advent of GPS, yield mapping became commonplace. The first GPS devices were available on tractors in 1991, but they had an accuracy of only about 100 m which was not good enough for mapping. By the mid-1990s differential GPS (DGPS) accuracy improved to 5–10 m, and this has improved further since 2000 when the US Department of Defence turned off selective availability. Now it is a few centimetres only. By the mid-1980s microprocessors made possible the development of computers for farm equipment and controllers.

The term precision agriculture appears to have been used first in 1990 as the title of a workshop held in Great Falls, Montana, sponsored by Montana State University. The focus was on more precise local management, i.e. site-specific management (SSM), and consequently the unit of management became the field and its intrinsic variation was of interest. This reflects a change in the scale of operation from the farm to the field, but there is more to it than this. With the increase in size of machinery being used in agriculture in the developed countries, farmers removed field boundaries and merged fields into increasingly larger units. The original fields, which had probably been created because of a particular set of soil or landscape conditions, were now parts of larger fields and their inherent variation

was added together. The increase in field size was accompanied, therefore, by an increase in within-field variability.

Before the 1990s, maps other than of the soil and possibly landscape played little part in agricultural management. Schafer et al., (1984) said at this time that maps of soil type and topography could be used to control fertilizer and pesticide applications and tillage operations. A major stumbling block to the wider spread and adoption of PA is the sparsity of soil and crop information, although there have been examples of on-the-go measurement of pH (Viscarra Rossel and McBratney, 1997). The National Research Council (1997) also made the point that "current mapping techniques are limited by a lack of understanding of geostatistics necessary for displaying spatial variability of crops and soils", and "An increased knowledge base in geostatistical methods should improve interpretation of precision agriculture data".

It is clear from the quotations given above that the value of geostatistics had already been established in PA. In fact, the early applications came from scientists already conversant with geostatistical methods. The marriage of geostatistics and PA was an easy one because geostatistics requires enough data at an interval that resolves the variation adequately to compute reliable variograms. These demands can be satisfied by the kinds of data widely available in PA, apart from some soil and crop data **(**Oliver, 2010**)**.

4.1.6.6 Site-specific management zones

Geostatistical techniques have been adopted with some enthusiasm in PA because of their suitability for quantifying and predicting the spatial variation of soil, crop and landscape properties. Natural systems in the environment usually show structured or periodic variation in time or space (i.e. spatial or temporal dependence). This is particularly true for soil systems where patterns develop as a result of variation in topography, parent material, climate and biology. The consequence of spatial dependence is that samples separated by small distances tend to be more similar than those further apart. Classical statistical procedures on the other hand assume that data are spatially independent. Geostatistics is a collection of statistical methods that have been used for some time in the geosciences. The basis of the methods is to describe and model spatial dependence or autocorrelation among sample data, and to use this information for various types of spatial prediction. There is some overlap with GIS (geographic information systems) and spatial statistics more generally. There are two major components of a geostatistical analysis: modelling spatial dependence in the form of a correlogram or variogram, and predicting variable values at unsampled locations with

techniques such as kriging or cokriging. Geostatistics can provide accurate maps for the successful implementation of variable-rate prescription for site-specific nutrient management and other applications in precision agriculture, such as irrigation. If the objective is to quantify the spatial variation in a given field, the sample design used to obtain data is important. Geostatistics places a different emphasis on the approach to sampling from that used in conventional statistics. Classical methods of sampling based on randomization of the sampling positions aim to avoid spatial correlation because of the assumptions that underpin many conventional statistical techniques. In geostatistics, the aim of sampling is to ensure that the data will be spatially correlated and randomization is no longer a requirement. Randomization in geostatistics is a feature of the model rather than a property of the phenomenon of interest. Furthermore, geostatistics changes the emphasis from the estimation of regional averages in classical statistics to the local estimation of spatially distributed variables using techniques such as kriging and cokriging (Khosla et al., 2010).

4.1.6.7 Delineation of site specific management zones

The delineation of management zones (MZ) is a cost-effective tool for site specific applications of inputs. The traditional methods to delineate management zones don't consider the spatial relationships between the observations. Geostatistics treats multivariate indices of spatial variation as continua in a joint attribute-geographical space.

Castrignanò et al., (2009) reported that one of the major advantages of geostatistics over simpler spatialization methods is that sparse observations of the primary attribute can be complemented by secondary attributes that are more densely sampled. They applied two methods to incorporate dense secondary information: multicollocated cokriging, which restricts the neighbourhood to the only secondary data, collocated with the target location and with the available data of the primary variable and simple cokriging with varying local means related to crisp classes. Their objective was to find the method that best improves the estimation of primary attributes through dense secondary information for the study area (province of Siena in central Italy), of about 3820 km^2. After the two methods were compared in terms of precision, through cross-validation, and accuracy, through a validation test, using an independent data set of 170 soil depth measurements. The results did not show clear differences among the methods.

Castrignanò et al., (2010) proposed a quantitative approach to unambiguously locate, characterise and visualise agro-ecozones and their boundaries which can be associated to

different environmental conditions. They used environmental parameters, including climatic and soil characteristics, hypothesized to be generally relevant to many crops in Capitanata-Foggia (South Italy). Cokriged environmental estimates at 500 m scale were used in a clustering algorithm based on non-parametric multivariate density estimation. The proposed approach produced the delineation of the study area in five compact classes in the space of environmental attributes that were also contiguous in geographic space. The resulting agro-ecozones may provide a framework for useful application in land use decision making.

De Benedetto et al., (2012) stated that proximal sensing (Electromagnetic Induction (EMI), Ground Penetrating Radar (GPR), hyperspectral spectroscopy (HS)) and remote sensing (RS) can complement direct sampling. However, sensor data fusion techniques, jointly analysing data from different sources, are still being developed. They performed a work with the intention to define a multivariate and multi-sensor approach by combining EMI, GPR, RS and HS data, without any previous weighing, in order to differentiate an 1.5 ha arable field into homogenous zones. The multi-sensor data were split into four groups: 1) bulk electrical conductivity (EC) from EMI data, 2) amplitude of GPR signal data, 3) the first principal components relating to five bands (green, yellow, red, red edge, near-infrared (NIR) PCs) of hyperspectral reflectance data and 4) the vegetation indices (NDVI, NDRE and NIR/Green) calculated from the remote sensing images. The data of each group were separately analysed and interpolated at the nodes of the same grid by using cokriging or kriging. To obtain spatially contiguous clusters, a combined approach was used, based on multivariate geostatistics and a non-parametric density function algorithm of clustering, applied to the overall multi-sensor data set of the estimates. The full approach allowed identifying three homogenous areas. In particular cluster 1, in the NW part of the field, with the lowest values of bulk electrical conductivity and GPR amplitude, and the highest red PC values. The other two clusters were delineated in the SE part of the field, with the highest values of green, yellow, red edge and NIR PCs for cluster 2, and the highest values of bulk electrical conductivity and vegetation indices for cluster 3. The delineation might be related to the intrinsic spatial variability of soil and the health status of plants and be used to produce a prescription map for site-specific management.

Yan et al., (2007) studied five soil and landscape attributes, including a NDVI image, soil electrical conductivity, total nitrogen, organic matter and cation exchange capacity acquired for a coastal saline land, and their spatial variability and spatial distribution maps were constructed with geostatistics techniques. Principal component analysis and fuzzy c-

means clustering algorithm were then performed to delineate the management zones, fuzzy performance index (FPI) and normalized classification entropy (NCE)was used to determine the optimal cluster number. To assess whether the defined three management zones can be used to characterize spatial variability of soil chemical properties and crop productivity, 139 georeferenced soil and yield sampling points across the field were examined by using variance analysis. It was found that the optimal number of management zones for the study area was three and there existed significantly statistical differences in the chemical properties of soil samples and yield data between the management zones. Management zone 3 presented the highest nutrient level and potential crop productivity, whereas management zone 1 the lowest. The results revealed that the given five variables could be aggregated into management zones that characterize spatial variability in soil chemical properties and crop productivity. The defined management zones not only can direct soil sampling design, but also provide valuable information for site-specific management in precision agriculture.

Aggelopooulou et al., (2013) studied several factors for the MZ delineation including crop and soil characteristics. They applied multivariate analysis to delineate MZs. Soil and crop data, collected over 3 years from a Precision Agriculture project in an apple orchard in Greece, were used. The collected data were categorized in three groups, namely soil properties, yield and fruit quality. All data were analyzed for descriptive statistics and their distribution. Maps of the spatial variability for the 3 years were presented. Data were jointly analyzed for management zone delineation using a combination of multivariate geostatistics with a non-parametric clustering approach, and the orchard was divided in four zones which could be differently managed. However, further research and experimentation are needed before precision horticulture being confidently adopted in Greece.

Diacono et al., (2013) proposed a combined approach of multivariate geostatistics and non-parametric clustering to delineate homogeneous zones that could be potentially managed with the same strategy. In a durum wheat field of Southern Italy, in organic farming, some soil physical and chemical properties (electrical conductivity; pH; exchangeable bases; total nitrogen; total organic carbon; available phosphorous), elevation and the Normalized Difference Vegetation Index were determined and interpolated by using geostatistics. The clustering approach, applied to the (co)kriged estimates of the variables, produced the delineation into four sub-field zones. A significant relation between soil fertility and yield was not found in such zones. Despite this, the proposed approach has the potential to be used in

future applications of precision agriculture. Further work could focus on site-specific nitrogen fertilization with suitable machinery.

4.1.6.8 Data fusion

According to Durrant-Whyte (2001) data fusion is the process of combing information from a number of different sources to provide a robust and complete description of an environment or process of interest. Data fusion is of special significance in any application where large amounts of data must be combined, fused and distilled to obtain information of appropriate quality and integrity on which decisions can be made. Data fusion finds application in many military systems, in civilian surveillance and monitoring tasks, in process control and in information systems. Data fusion methods are particularly important in the drive toward autonomous systems in all these applications. In principle, automated data fusion processes allow essential measurements and information to be combined to provide knowledge of sufficient richness and integrity that decisions may be formulated and executed autonomously. Definitely, having an additional data set must help, particularly if it is measuring the same thing. If not, it will still help if the second measurement is correlated with the first.

Piikki et al., (2012) studied proximal sensor data fusion for topsoil clay mapping on a 22 hectare agricultural field in southwest Sweden. Eight different predictor sets and two different prediction methods were tested in an orthogonal design. The predictor sets were different combinations of proximally measured gamma (γ) ray spectrometry and apparent electrical conductivity (ECa), four terrain attributes (elevation, slope and the cosine and the sine of aspect) and the digital numbers (DNs) of an aerial photo. The two prediction methods were partial least squares regression (PLS-R) and k nearest neighbour prediction (kNN). It was found that the γ ray spectrometry variables (^{232}Th, ^{40}K and total count of decays) were good predictors of topsoil clay content (mean absolute error of about 1.5% clay) and predictions were neither much improved nor deteriorated by addition of any of the other predictors. The ECa measurements, which are affected also by the subsoil, did not perform as well. Predictions were improved when the ECa data were integrated with the aerial photo DN but were deteriorated by addition of elevation data. The kNN method yielded slightly better predictions than the PLS-R method but overall it was more important which input data were used than how the predictions were made. It was observed that even though dense soil

sampling was used for calibration (three samples per hectare), use of proximal soil sensor data was almost always better than mere interpolation of the calibration samples.

Geo-electrical sensors are often used as auxiliary variables with sparse direct measurements to estimate soil properties. Using a single sensor is not ideal in some circumstances. For example, sandy, sandy gravelly, sandy salt-affected and clayey soils are poorly identified using an EMI or gamma-ray sensor singularly. The complementary use of these sensors should improve interpretation in landscapes containing these soils. Analysis of multi-sensor data is however problematic. Several methods have been developed to integrate multi-sensor data but there is currently no unequivocally accepted methodology (Castrignanò et al., 2012).

Castrignanò et al., (2012) used EMI, gamma-ray emission and GPS height as multi-sensor auxiliary data for soil characterisation. Their objectives were: 1) to define a combined approach of geostatistics and sensor data fusion to integrate field data from electromagnetic induction (EMI) measured with EM38 and EM31, gamma (γ)-ray and RTK GPS sensors with soil sample data for delineating areas of homogeneous soil; 2) to show the potential of gamma radiometric sensor by estimating a relationship for crop available soil potassium (K) from the γ-ray signal. The geophysical survey was carried out on an 80-ha cropping field in Corrigin, Western Australia. Seventy-seven soil samples were collected at the nodes of a 100×100m-mesh grid and analysed for different properties. The EM38 and EM31 data were strongly correlated with each other and so were γ-radiometric counts from thorium (Th), uranium (U) and all elements (TC). The multi-sensor data were split into 4 subgroups, based on their similarities: 1) EMI data; 2) γ-radiometric counts from potassium (emitted from all forms of K including readily plant available, non-exchangeable and structural K); 3) γ-radiometric counts from Th, U and TC and 4) RTK GPS height. Each group of data was separately analysed using different geostatistical techniques. The soil samples and geophysical data were jointly interpolated using multi collocated cokriging. The EMI data showed anisotropy and an anisotropic Linear Model of Coregionalization was fitted before cokriging. The EM31 and EM38 maps looked quite similar. The maps of γ-U, Th and TC were also similar, suggesting that they reflected the same soil properties, but were somewhat different from the γ-K maps. High values of EMI coincided with both low γ- radiometric values at the valley bottom, due to moist sandy salinity-prone soil of varying depth to texture contrast, and high γ-radiometric values at the elevated areas of the field due to emission from finer textured soil. High γ-radiometric values coincided also with low values of EMI over gravelly sands. Only the use

of a multi-sensor platform could discriminate soils that gave similar outputs to one sensor. The first two principal components of the geophysical data were used to partition the field into homogeneous areas.

So far, no published work on the use of vis-NIR spectroscopy coupled with multivariate geostatistics and then this study may be the first work on the marriage between vis-NIR on-line sensing and multivariate geostatistical techniques.

5. Materials and Methods

5.1 Experimental site

This study was conducted in a 18 ha field (52° 05’ 51’’ N, 0° 26’ 50’’ W), named as Cotton End field located in Duck End farm, Wilstead, Bedfordshire, United Kingdom. The soil texture over the field down to 0.20 m is non-homogeneous, including three textures (Table 1) of sandy loam, loam and sandy clay loam according to the United State Department of Agriculture texture classification system. Directly after harvesting the previous wheat crop, on-line soil vis-NIR spectral measurements were carried out.

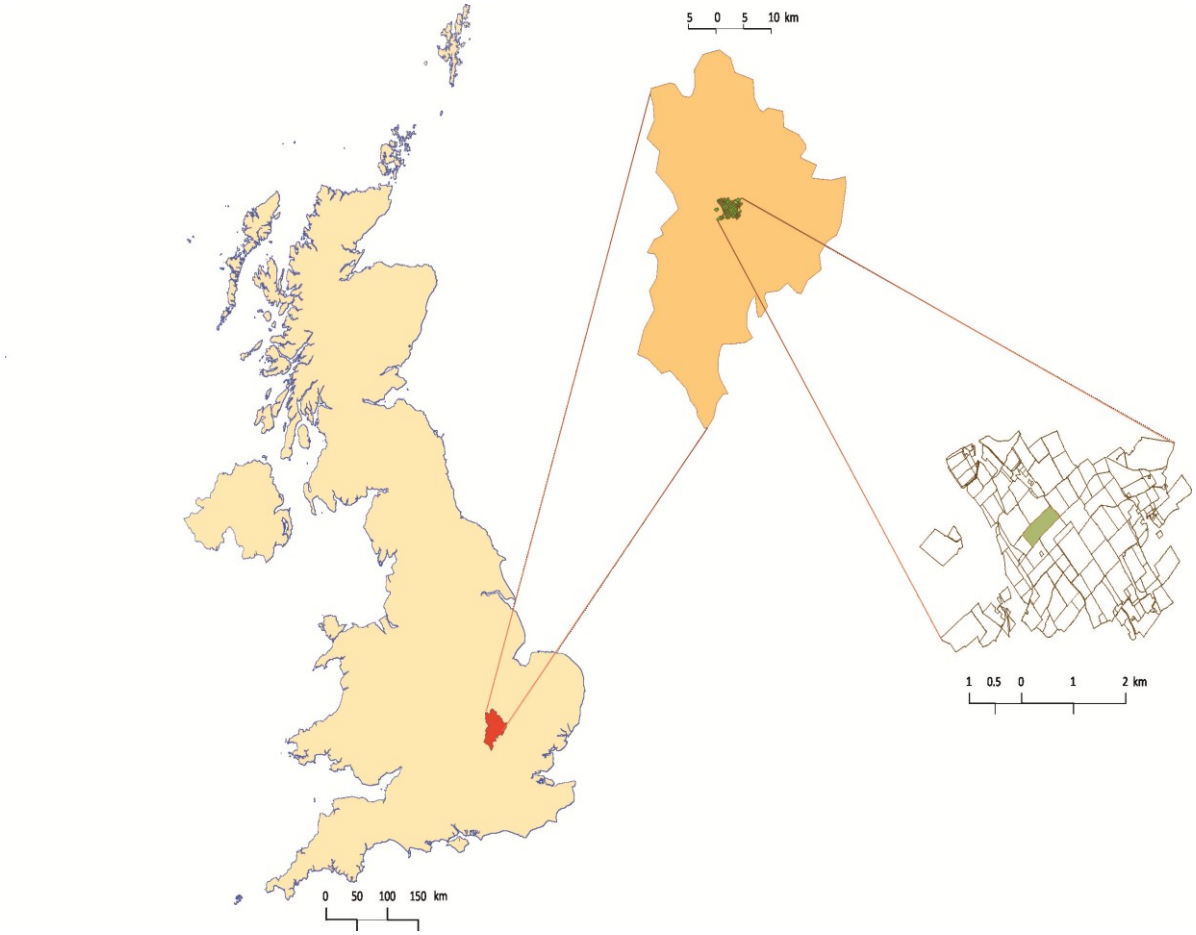

Figure 6. Location of the Cotton End field, where on-line visible and near infrared (vis-NIR) measurement was carried out in summer 2012

5.2 On-line visible and near infrared sensor

The on-line vis-NIR sensor designed and developed by Mouazen (2006) was used (Figure 7) to carry out the field measurement. It consisted of a subsoiler that penetrates the soil to the required depth, making a trench, whose bottom is smoothened due to the downwards forces acting on the subsoiler (Mouazen et al., 2005). The optical probe, housed in a steel lens holder, was attached to the back side of the subsoiler chisel to acquire soil spectra in diffuse reflectance mode from the smooth bottom of the trench. The subsoiler, retrofitted with the optical unit, was attached to a frame that was mounted onto the three point linkage of the tractor (Mouazen et al., 2005). An AgroSpec mobile, fibre type, vis-NIR spectrophotometer (tec5 Technology for Spectroscopy, Germany) with a measurement range of 305-2200 nm was used. The spectrophotometer was IP66 (ingress protection) protected for harsh working environments. Although this spectrophotometer does not cover the entire wavelength range in the NIR region, it was selected in this study as it uses diode array detectors, which have been proven to be stable under on-line measurement conditions (Mouazen et al., 2009). A 20 W halogen lamp was used as a light source. A 100% white reference was used before scanning, and this was repeated every 30 min. A differential global positioning system (DGPS) (EZ-Guide 250, Trimble, USA) was used to record the position of the on-line measured spectra with sub-metre accuracy. A Panasonic semi-rugged laptop was used for data logging and communication. The spectrometer system, laptop and DGPS were powered by the tractor battery. A New Holland T5000 tractor with 100 Ah battery was used. The total power consumption for all electrical parts of the on-line vis-NIR sensor was around 60 W. The on-line sensor was used to measure the Cotton End field in the UK (Figure 6) in summer 2012.

Table 1. Soil texture defined according USDA soil classification.

Soil texture class	Number of Samples	Sand (> 50 µm) range (%)	Silt (50- 2 µm) range (%)	Clay (< 2 µm) range (%)
Sandy loam	42	52.10-61.25	24.02-31.64	12.74-18.96
Loam	28	42.52-52.00	28.89-34.25	15.11-24.74
Sandy clay loam	2	51.10-52.06	26.95-26.98	20.99-21.91

Figure 7. The on-line visible and near infrared (vis-NIR) soil sensor (Mouazen, 2006)

5.2.1 On-line measurement and soil sampling

The Cotton end field, which was of a rectangular shape of 260 m by 700 m (Figure 8), covering 18 ha of land was measured by the on-line system. Each measured line was 700 m long with 15 m intervals between adjacent transects. The forward speed of the tractor was around 1.5 km h^{-1} and the measurement depth was set at 20 cm. During each line measurement, twenty three soil samples were collected from the bottom of the trench and the sampling positions were carefully recorded.

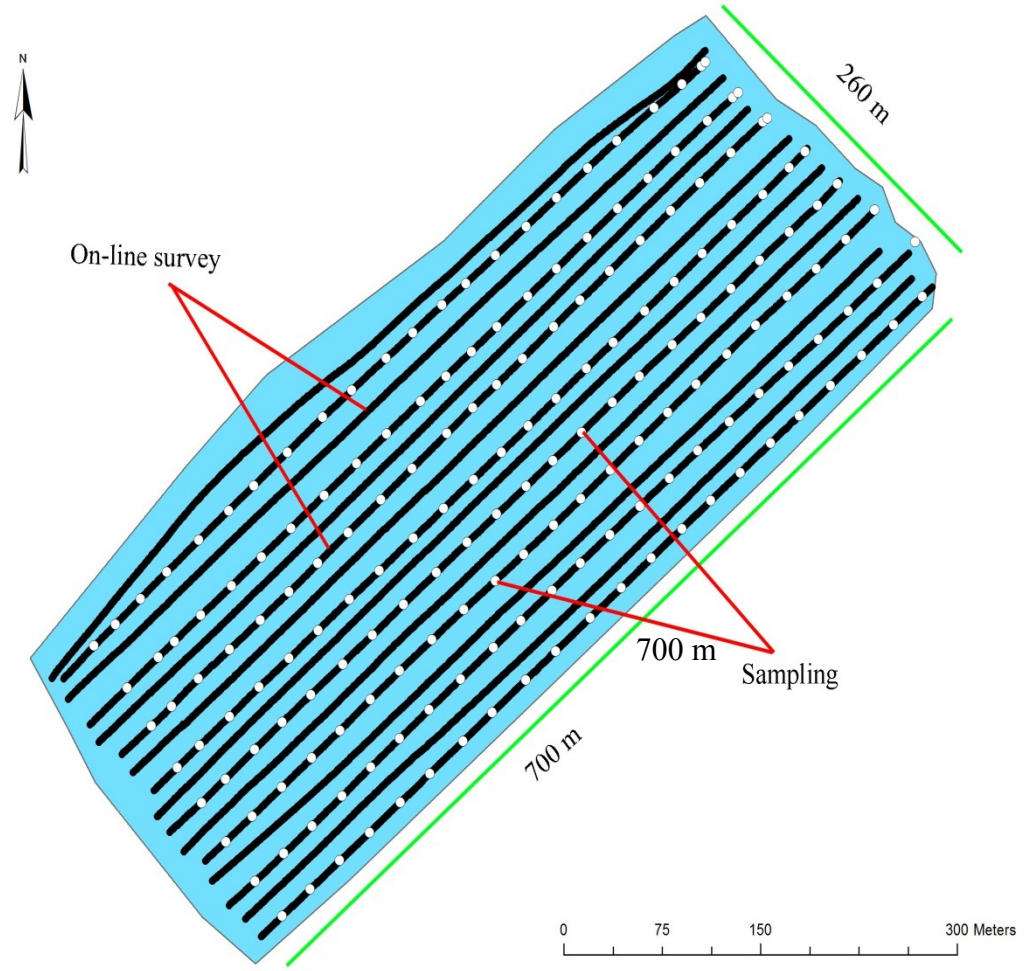

Figure 8. On-the-go field survey and locations of soil samples

A total of 183 soil samples based on 30 m x 30 m grid (Figure 8) were collected. Around 700 g soil was taken for each soil sample and placed into a tightly sealed plastic bag to hold field moisture. The soil samples were then transported to the soil laboratory in Cranfield University, where they were stored deep frozen (-18 °C) until analysis. After defrosting and thorough mixing, each soil sample was divided into two parts. A quantity of the first part (fresh soil) was used for MC measurement by oven drying at 105 °C for 48 h. The rest of fresh soil was scanned with a vis-NIR spectrophotometer. This scanning was designated as laboratory scanning of fresh soil samples. The other part of the soil was dried at 40 °C, grinded and sieved with a 2mm sieve. This part was subjected to physical and chemical soil analysis at soil laboratory in Department of Agriculture, Sassari University, Italy.

5.2.2 Laboratory optical measurement

The optical scanning under laboratory measurement conditions was carried out using the same mobile, fibre type, vis-NIR spectrophotometer (AgroSpec from tec5 Technology for Spectroscopy, Germany), as that used for on-line scanning. A 100 % white reference was used before scanning. Small stones and plant debris were first removed from the fresh soils. Each soil sample was replicated three times in three cups. Soil in cups was levelled by spatula. This resulted in a smooth soil surface that assures a maximum reflection and thus a good signal-to-noise ratio (Mouazen et al., 2005). A total of 10 scans were collected from each cup, and these were averaged in one spectrum.

5.2.3 Laboratory analysis

Procedures used for determining the physical and chemical properties of soil samples were as follows:

- Soil texture fractions were determined according to the international pipette method; sodium hexametaphosphate was used as a dispersing agent (Piper, 1950). Soil texture classes were identified according to USDA classification
- Phosphorus was extracted by 0.5 N in sodium hydrogen carbonate solution as described by Olsen et al., (1954), then the extracted P was measured calorimetrically using a UV-VIS-NIR spectrophotometer according to the method of Murphy and Riley (1962)
- Total nitrogen and total carbon were analysed with CHN 628 analyzer based on dry combustion of sample with helium and oxygen gases (LECO, USA).
- Cation exchange capacity was measured according to barium chloride (pH 8.2) method
- Exchangeable cations were extracted by 1N ammonium acetate solution at pH 7 (Piper, 1950), then the exchangeable calcium, magnesium, potassium and sodium were determined using Atomic absorption spectrometer AA analyst 200 Perkin Elmer Instruments.
- pH was measured using a glass electrode in a 1:5 (volume fraction) of soil in distilled water. After shaking for 2 h and equilibration, pH was measured in the supernatant.

5.2.4 Model development

The first step in developing calibration models is the pre-treatment of the spectral data. Soil spectra were first reduced to 371 – 1662 or 2150 nm to eliminate noise at both sides of spectra. Spectra were further reduced by averaging three successive points in the vis range, and 6 or 15 points in the NIR range to reduce the number of wavebands and smooth the spectra. The Savitzky-Golay (S–G) smoothing, maximum normalisation and first derivation were successively implemented. Normalisation is typically used to get all data to approximately the same scale, or to get a more even distribution of the variances and the average values. The maximum normalisation method adopted in this study is a normalisation that "polarises" the spectra. Details of spectra pre-processing of soil elements are listed in Table 2.

Several pre-treatments were considered and the best performing pre-treatment was withheld for each parameter. The selection criteria of any pre-treatment were the largest coefficient of multiple determination (R^2) and residual prediction deviation (RPD) (the ratio of standard deviation (S.D.) of the measured data set to root mean square error of prediction (RMSEP) and the smallest RMSEP.

The chemical and spectral data were used to develop calibration models of soil properties under study using partial least squares regression (PLSR), with full-cross validation. The PLS is a bilinear modelling method where information in the original x data is projected onto a small number of underlying ("latent") variables called PLS components. The y data are actively used in estimating the "latent" variables to ensure that the first components are those that are most relevant for predicting the y variables. Interpretation of the relationship between x data and y data is then simplified as this relationship is concentrated on the smallest possible number of components. More detailed information about the PLS can be found in Martens and Naes (1989). The Unscrambler® software Version 7.88 (Camo A/S, Oslo, Norway) was used for spectra pre-treatment and model development.

Table 2. Spectral pre-processing of different soil elements

Soil constitute	Spectral pre-processing
pH	- delete spectra 305-370 and 2151-2200 - reduce by 3 spectra 371-1000 - reduce by 6 spectra 1001-2150 - normalise maximum - s.golay derivative using left hand side 2,2,2 - s.golay smooth using left hand side 2,2,2 - delete spectra 371-421 and 1746-2150
Extractable phosphorus (P $mg.kg^{-1}$)	- delete spectra 305-400 and 1664-2200 - reduce by 3 spectra 401-1000 - reduce by 6 spectra 1001-1663 - normalise maximum - s.golay derivative using left hand side 2,2,2 - s.golay smooth using left hand side 2,3,3
Total nitrogen (TN%)	- noise cut at 305-370 nm and 2151-2200 nm - wavelength reduction by 3 at 371-1000 nm - wavelength reduction by 15 at 1001-2150 nm - Maximum normalisation - S-G derivative using left hand side 2,2,2 - S-G smooth using left hand side 2,2,2 - wavelength cut at 371-419 nm and 1744-2150 nm
Total carbon (TC%)	- noise cut at 305-370 nm and 2151-2200 nm - wavelength reduction by 3 wavelengths at 371-1000 nm - wavelength reduction by 15 wavelength 1001-2150 nm - maximum normalisation - S-G derivative using left hand side 2,2,2 - S-G smooth using left hand side 2,2,2
Moisture content (MC%)	- noise cut at 305-370 nm and 2151-2200 nm - wavelength reduction by 3 at 371-1000 nm - wavelength reduction by 15 at 1001-2150 nm - maximum normalisation - S-G derivative using left hand side 2,2,2 - S-G smooth using left hand side 2,2,2 - spectra cut at 371-422 nm and 1744-2150 nm
Exchangeable potassium (K $cmol+.kg^{-1}$)	- noise cut at 305-400 nm and 1662-2200 nm - wavelength reduction by 3 at 401-1000 nm - wavelength reduction by 6 at 1001-1661 nm - maximum normalisation S-G derivative using left hand side 2,2,2 - S-G smooth using left hand side 2,2,2

Samples used for calibration with cross-validation of pH, P, TN, TC, MC and K were 125, 62, 133, 108, 127 and 43, respectively, whereas samples used for the independent validation were, 48, 23, 22, 24, 45 and 24, respectively. Details of soil elements statistics are listed in Table 3.

Table 3. Sample statistics of calibration dataset for partial least square regression (PLSR) coupled with cross-validation and of independent validation dataset.

	Calibration						Independent validation					
Statistics	pH	P	TN	TC	MC	K	pH	P	TN	TC	MC	K
Sample no	125	62	133	108	127	43	48	23	22	24	45	24
Min	5.39	6.80	0.091	1.207	13.41	0.12	5.16	4.80	0.112	1.304	13.41	0.12
Max	8.21	40.00	0.198	2.351	24.74	0.37	8.17	50.00	0.201	2.455	24.28	0.40
Mean	6.59	17.97	0.141	1.806	17.96	0.21	6.46	22.50	0.151	1.790	18.03	0.23
SD	0.682	8.237	0.020	0.221	2.015	0.073	0.895	15.231	0.024	0.275	2.164	0.078

P: Phosphorus ($mg.kg^{-1}$), TN: Total nitrogen (%), TC: Total carbon (%), MC: Moisture content (%), K: Exchangeable potassium ($cmol+. kg^{-1}$)
SD: Standard deviation

5.2.5 Statistical evaluation of PLS model performance

Root mean square error of calibration (RMSEC) and prediction (RMSEP) are measures of average differences between predicted and measured response values at calibration and validation stages, respectively (Yitagesu et al., 2009). For the evaluation of the model performance, RMSEP was used (Williams and Norris, 2001). The RMSEP can be expressed as follows:

$$\mathrm{RMSEP} = \frac{1}{\mathrm{N}}\sqrt{\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{X_i} - \mathrm{Y_i})^2}$$

where X_i is the predicted value and Y_i is the observed value.

RPD, designated as rate of prediction deviation, which is the ratio of standard deviation (SD) of the measured values to RMSEP, was used to compare different models. The third parameter considered was the coefficients of determination (R^2). In fact, R^2 indicates the percentage of the variance in the Y variable that is accounted for by the X variable. A value

for R^2 between 0.50 and 0.65 indicates that more than 50% of the variance in Y is accounted for variance X, so that discrimination between high and low concentrations can be made. A value for R^2 between 0.66 and 0.81 indicates approximate quantitative predictions, whereas, a value for R^2 between 0.82 and 0.90 reveals good prediction. Calibration models having a value for R^2 above 0.91 are considered to be excellent (Williams, 2003). In the successful analysis of agricultural commodities, it is desirable to have R^2 >0.50, RPD>5. Nevertheless, for samples of complex material, Williams and Norris (2001) classified values as follows: RPD<1.0 indicates very poor model/predictions and their use is not recommended, RPD between 2.4 and 3.0 indicates poor model/predictions where only high and low values are distinguishable, RPD between 3.1 and 4.9 indicates fair model/predictions which may be used for assessment and correlation, RPD values between 5.0 and 6.4 indicate good model/predictions where quantitative predictions are possible, RPD between 6.5 and 8.0 indicates very good, quantitative model/predictions, and RPD>8.1+ indicates excellent model/predictions. However, for complex agricultural material such as soil, another RPD standard was reported by researchers (Saeys et al., 2005; Viscarra Rossel et al., 2006). Viscarra Rossel et al., (2006) classified RPD values as follows: RPD<1.0 indicates very poor model/predictions and their use is not recommended; RPD between 1.0 and 1.4 indicates poor model/predictions where only high and low values are distinguishable; RPD between 1.4 and 1.8 indicates fair model/predictions which may be used for assessment and correlation; RPD values between 1.8 and 2.0 indicates good model/predictions where quantitative predictions are possible; RPD between 2.0 and 2.5 indicates very good, quantitative model/predictions, and RPD>2.5 indicates excellent model/predictions. Chang et al., (2001) classified RPD values into three classes as follows: class A: RPD>2.0; class B: RPD=1.4–2.0; class C: RPD<1.4). The RPD values obtained in this study were classified according to the latter proposed limits, and were used to evaluate the accuracy of PLS models for the prediction of soil properties under study.

5.2.6 Mapping

All maps for studied soil properties were developed by the inverse distance weighting (IDW) method using ArcGIS 10.1 (ESRI, USA) mapping software. Two types of maps were developed for each property, namely full-point and comparison maps (measured, predicted, error). The full point maps consisted of all on-line predicted points of the studied soil properties. The comparison maps compare on-line predicted values and laboratory measured

data using the validation sets. To have useful comparison among the produced maps, the same number of classes (6 classes) was considered for all maps.

With regards to maps produced by geostatistical analysis, there were two types of maps; the first type was for thematic spatial maps of the studied soil properties interpolated using three different geostatistical models (soil, spectral and fusion). Thematic spatial maps were classified into 6 isofrequency classes in order to better disclose spatial patterns. The second type included the maps produced with factor cokriging, in particular the maps of the significant regionalised factors used to delineate management zones (MZs). The MZs maps were classified into 3 main classes of such a size that they can be managed by farmers. All geostatistical analyses were performed with the software package ISATIS (Geovariances, 2013).

5.3 Geostatistical analysis

The complex geostatistical analysis was divided into 2 parts:

1- Part A: aimed at delineation of homogenous zones using separately either soil measured data (soil variables) or estimated data from the spectral variables (Figure 9).

2- Part B: aimed at showing the potential of the fusion between on-line measurements and soil data to improve the delineation of homogenous zones (Figure 10).

Intrinsic stationarity was assumed for all variables. All maps produced using the geostatistical analysis were interpolated on 1m x 1m square grid.

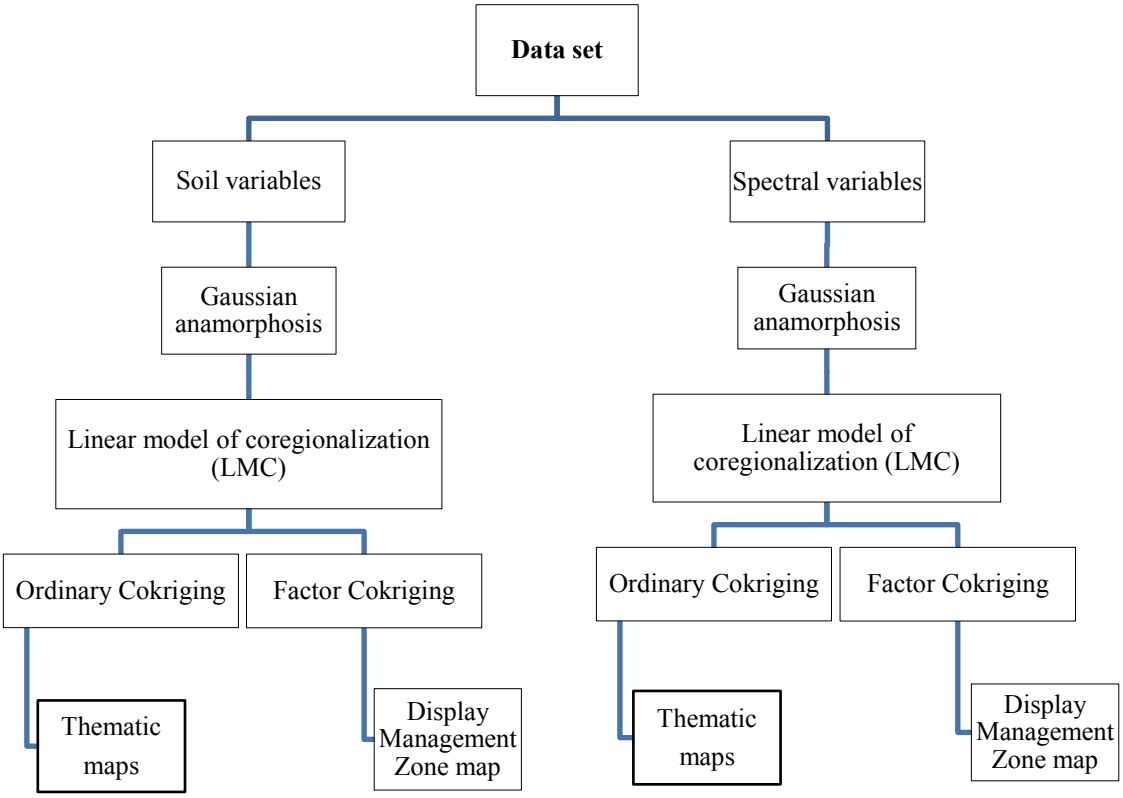

Figure 9. Flow chart of part A represents geostatistical analysis applied on soil and spectral variables data sets

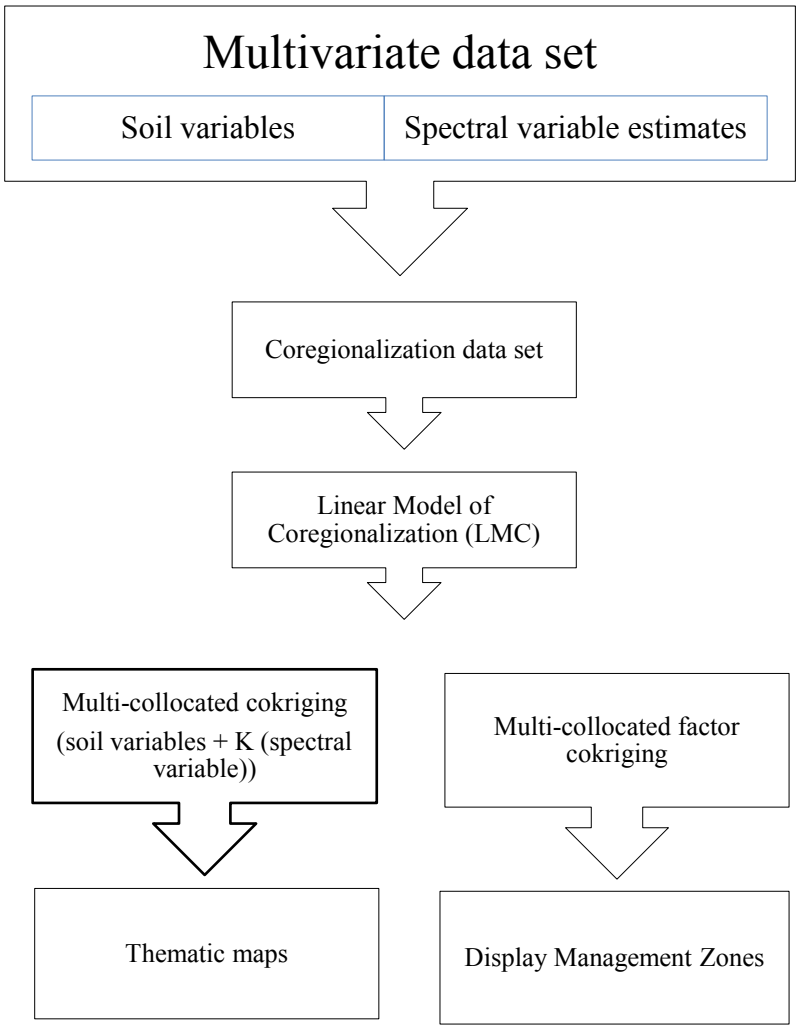

Figure 10. Flow chart of the second part B represents the fusion data set analysis

5.3.1 Data acquisition

Two different types of data were acquired:

1) Spectral variables data (predicted values using the on-line vis-NIR spectra and PLSR models). These spectral variables were soil pH, extractable phosphorus, moisture content and exchangeable potassium. The latter variable was used as secondary attribute in the multi-collocated cokriging analysis.
2) Soil variables data of collected soil samples based on 30 m x 30 m grid. These soil variables were soil pH, moisture content, extractable phosphorus, particle size distribution (Sand, Silt and Clay), exchangeable cations (Ca, Mg, K and Na), cation exchange capacity, total nitrogen and total carbon.

Three Linear models of coregionalization (LMC) were developed: the first is for soil variables, the second is for spectral variables and the third is for the fusion between soil and spectral variables. The multivariate spatial data set was analysed by: multi-collocated co-Kriging to produce the thematic maps of the study variables and Factor cokriging to determine scale-dependent regionalised factors.

Part A:

Geostatistical methods were applied to the whole data sets. The assumption of stationarity was tested. If the variables can be assumed as intrinsically stationary, a multigaussian approach was applied if some variables show departure from normal distribution. It consists of the following steps:

1. Transform the initial variables into Gaussian-shaped variables with zero mean and unit variance through an approach using Hermite polynomials, called Gaussian anamorphosis (Chiles and Delfiner, 1999; Wakernagel, 2003).

Gaussian Anamorphosis Modelling

A difficulty in the practical application of a multivariate approach occurs when the variables are of widely differing sizes. A solution is to standardize the individual variables to give each an average of zero and a variance of unity. Variogram modelling is further complicated by the presence of outliers in highly skewed data distributions. In this case it is better to perform a normalization of data through Gaussian anamorphosis modelling. Gaussian anamorphosis is a mathematical function which transforms a variable Y with a Gaussian standardized distribution in a

new variable Z with any distribution: $Z = \Phi(Y)$. As this function needs to be known for any Gaussian value, a model is required. This is made by fitting a polynomial expansion (Chiles and Delfiner, 1999):

$$\Phi(Y) = \sum \Psi_i H_i(Y)$$

where H_i (Y) are called Hermite Polynomials. In practice the polynomial expansion is truncated at a generally high order (30-100) and tends to be bijective within the interval defined by the minimum and the maximum of the sample values (Wakernagel, 2003). Model fitting then consists of calculating the Ψ_i coefficients of the expansion and, in order to transform the raw variable into a Gaussian one, the anamorphosis function has to be inverted: $Y = \Phi^{-1}(Z)$.

This inversion, outside the interval where the function is not strictly increasing for all the values of Y, is performed using a linear interpolation (Wakernagel, 2003).

2. Adopting a Gaussian model, a LMC was fitted to all experimental variograms, both direct and cross-variograms, of the transformed data, and then ordinary cokriging (Goovaerts, 1997) was applied as conditional expectation estimator. Finally, the estimates were back-transformed to the raw values of the variables through the anamorphosis functions previously calculated.

Linear Model of Coregionalization

The LMC, developed by Journel and Huijbregts (1978), considers all the *n* studied variables as the result of the same independent physical processes, acting over different spatial scales *u*. The *n(n+1)/2* simple and cross semivariograms of the variables are modelled by a linear combination of N_S standardized semivariograms of unit sill, $g^u(h)$. Using the matrix notation, the LMC can be written as:

$$\boldsymbol{\Gamma}(\mathbf{h}) = \sum_{u=1}^{N_S} \mathrm{B}^u g^u(\mathbf{h})$$

where $\boldsymbol{\Gamma}(\mathbf{h})=[\gamma_{ij}(\mathbf{h})]$ is a symmetric matrix of order $n \times n$, whose diagonal and out-of-diagonal elements represent simple and cross semivariograms, respectively; $\mathbf{B}^u=[b^u_{ij}]$ is called coregionalization matrix and it is a symmetric positive semi-definite matrix of order $n \times n$ with real elements b^u_{ij} at a specific spatial scale *u*. The model is authorized if the functions $g^u(\boldsymbol{h})$ are authorized semivariograms models (Castrignanò et al., 2000).

3. Spatial interpolation of Gaussian transformed variables using ordinary (co) kriging.

4. Back transformation to the raw data through a mathematical model calculated in Gaussian anamorphorsis.

PART B:

With geostatistics sparse observations of the primary attribute can be complemented by secondary attributes that are more densely sampled (Castrignanò et al., 2009). The spectral variable of exchangeable potassium was used as secondary attribute to create the fusion model.

To produce the coregionalization (Fusion) data set, the spectral variables were estimated at the sample locations using the previously fitted LMC. The fusion data set of soil and spectral variables was interpolated using multicollocated cokriging.

Multi-collocated cokriging

Ordinary cokriging (OCK) is one of the most basic cokriging methods, which assumes the local mean to be constant but of unknown value. A way of integrating secondary finer-resolution information in primary sparse variable modelling is *collocated cokriging*, where the contribution of a secondary variable to the cokriging estimate relies only on the cross-correlation between the two variables (Goovaerts, 1997). The approach is quite similar to ordinary cokriging with the only difference occurring in the neighbourhood search. As using all secondary exhaustive information contained within the neighbourhood may lead to an intractable solution, due to too much information, the initial solution of collocated cokriging was to use the single secondary value co-located at the target grid node. In ordinary cokriging the weights attached to the secondary variable must add up to zero, therefore, if only one data value is used, its single weight is zero. The original technique was then extended so that the secondary variable is used at the target location and also at all the locations where the primary variable is defined within the neighbourhood. This solution has generally produced more reliable and stable results (Castrignanò et al., 2009; Castrignanò et al., 2012). The modified version, also referred to as "Multi-Collocated Cokriging" (MCCO) in literature (Rivoirard, 2001), is less precise than full cokriging, not using all the auxiliary information contained within the neighbourhood, but it is much less computer time demanding. However, because the co-located secondary datum tends to screen the influence of more distant secondary data, there is actually little loss of information. Differently than the original technique of collocated cokriging, in this new version the influence of the secondary variable on the primary variable

is explicitly taken into account through the estimation of both direct secondary variable variogram and cross-variogram.

Regionalized Principal Component Analysis

Regionalized Principal Component Analysis breaks down each co-regionalization matrix $\mathbf{B}^u$ into the eigenvalues and eigenvector matrices (Wackernagel, 2003):

$$\mathbf{B}^u = \mathrm{Q}^u \Lambda^u \mathrm{Q}^{uT} = \left(\mathrm{Q}^u \sqrt{\Lambda^u}\right)\left(\mathrm{Q}^u \sqrt{\Lambda^u}\right)^T = \mathrm{A}^u \mathrm{A}^{uT}$$

where Q^u is the matrix of eigenvectors, i.e. the regionalised factors $Y_v^u(x)$, and Λ^u is the diagonal matrix of eigenvalues for each spatial scale *u*; $\mathbf{A}^u = \mathbf{Q}^u \sqrt{\Lambda^u}$ is the matrix of order *n x n* of the transformation coefficients a_{iv}^u. The transformation coefficients a_{iv}^u in the matrix A^u correspond to the covariances between the original variables $Z_i(x)$ and the regionalized factors $Y_v^u(x)$.

The approach breaks down the set of original second-order random variables $\{Z_i(x);\ i=1,\ldots,n\}$ into a set of reciprocally orthogonal factors $\{Y_v^u(x);\ v=1,\ldots,n;\ u=1,\ldots,N_S\}$ with transformation coefficients a_{iv}^u. Therefore, combining the spatial with the multivariate decomposition, it results::

$$Z_i(x) = \sum_{u=1}^{N_S} \sum_{v=1}^{n} a_{iv}^u Y_v^u(x)$$

Factor cokriging

Mapping the regionalized factors $Y_v^u(x)$ provides a way to show the behaviour and relationships among variables over different spatial scales. The estimation of the factors is performed by a modified co-kriging system, as described by Wackernagel (1988), (Castrignanò et al., 2000).

It is important to acknowledge that the underlying assumptions of Factor Co-Kriging Analysis (FCKA) are linearity and independence of factors. Moreover, factorial co-kriging depends on variogram modelling, i.e. on a somewhat arbitrary choice of the number/type of nested structures and range of variogram models; hence, when modelling variograms, any physical knowledge of the phenomena acting in the study area should be taken into account.

5.3.2 Prediction comparison:

5.3.2.1 Cross-validation

The performances of the three models for MZ delineation were individually assessed using cross validation (Isaaks and Srivastava, 1989), whereby one observation (z) at a time is temporally removed from the data set and re-estimated (z*) from the remaining data. To assess which method gives the most precise estimation of the soil properties, two statistical criteria were considered: mean error (ME), as an indicator of bias, and mean standardized squared error (MSSE) (scaled by the predicted standard deviation of estimation), as an accuracy measure:

$$ME = \frac{1}{N}\sum_{i=1}^{N}(z_i - z^*)$$

$$MSSE = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{z_i - z^*}{\sigma}\right)^2$$

where N is the number of active observations and σ the cokriging standard deviation. If the estimation is unbiased and accurate, the first statistic should be close to zero, whereas the second one should be close to one, because the latter corresponds to the ratio between an experimental variance and a theoretical one.

The appropriateness of the chosen LMC models was tested by cross-validation. This involves deleting each sample in turn and then kriging it (z*) independently from all other points in the estimation neighbourhood (Castrignanò, 2011). The outcomes of the cross-validation procedure can be shown in a graphic which involves the following plots:

a) The histogram of the standardised estimation errors, which should symmetrical, centred on 0 and without long tails.
b) The scatter diagram of the true data versus the estimated values, which should be close to the first bisector.
c) The scatter diagram of the standardised estimation errors versus the estimated values, which should look like a round cloud, symmetrical to horizontal axis.

5.3.2.2 Validation Test

In geostatistical practice, the usual method of testing is cross-validation. However, its results are actually biased and somewhat too optimistic (Creutin and Obled, 1982), because it retains the same variogram, whereas the variogram should be recomputed and fitted every time that an observation is removed (Laslett et al., 1987). Moreover, cross-validation is not true validation, because the same sample data set is used for both estimation and validation. All these shortcomings can be avoided by using a separate independent set of data for validation.

The validation procedure used an independent data set consisting of 63 soil samples. To check the null hypothesis that the experimental error (prediction - observation) is zero for each surveyed variable, a t-test was performed with unequal variances. This test assumes that the means of the two data sets are normally distributed and the means are equal. By the central limit theorem, mean of samples from a population with finite variance approaches a normal distribution regardless of the distribution of the population. Estimate distributions were also compared with the distributions of the validation data through quantile–quantile plots. This proves to be a good visual tool for comparing two different distributions. Differences between true and estimated values were also shown through boxplots of the experimental errors.

The three (soil, spectral, fusion) models were compared by calculating the following three statistics: mean error (ME), root mean squared error (RMSE) and mean squared error of error standardised by cokriging standard deviation (MSSE). The optimal values should be close to 0 for the first two statistics and 1 for the third one.

5.3.3 Spatial association of yield map with MZs

A yield map of barley (winter) was produced by using inverse distance weighting (IDW) method. The yield map was classified into 3 isofrequency classes (low, medium, high). The degree of similarity between the classes of yield and the ones of MZ delineation was calculated by applying the tool of “extract by mask” of ArcGIS mapping software, for the two, soil and fusion, models. Also, the mean and standard deviation of yield were calculated for each management zones of the delineations produced by the two models.

6. Results and discussion

Results and discussion section of the whole research can be divided into two parts: part A addresses the vis-NIR experiment which includes the on-line measurement, development of calibration models for different soil properties, evaluating the accuracy of the calibration models by using cross-validation and independent validation data sets and finally prediction of the studied soil properties using the spectra collected from the on-line sensor and calibration models. Part B consists of developing three different models, one for soil variables, one for spectral variables and one for the fusion between both types of variables. It also, includes delineation of management zones (MZ) according to the three developed models and finally comparing the estimated MZ map with the yield map of barley (winter) harvested on August 2013 in order to investigate the similarity between the two types of maps and then assess the productivity potential in MZ delineation.

6.1 Part A: vis-NIR experiment

All steps performed in this study regarding vis-NIR experiment are illustrated in the flow diagram shown in Figure 11, from the collection of both soil spectra and soil samples to the development of different maps. This diagram made it clear that the vis–NIR calibration models were developed using soil spectra measured under non-mobile laboratory environment. The applicability of these laboratory calibration models for on-line predictions of soil properties was validated using spectra measured on-line with the vis–NIR on-line soil sensor.

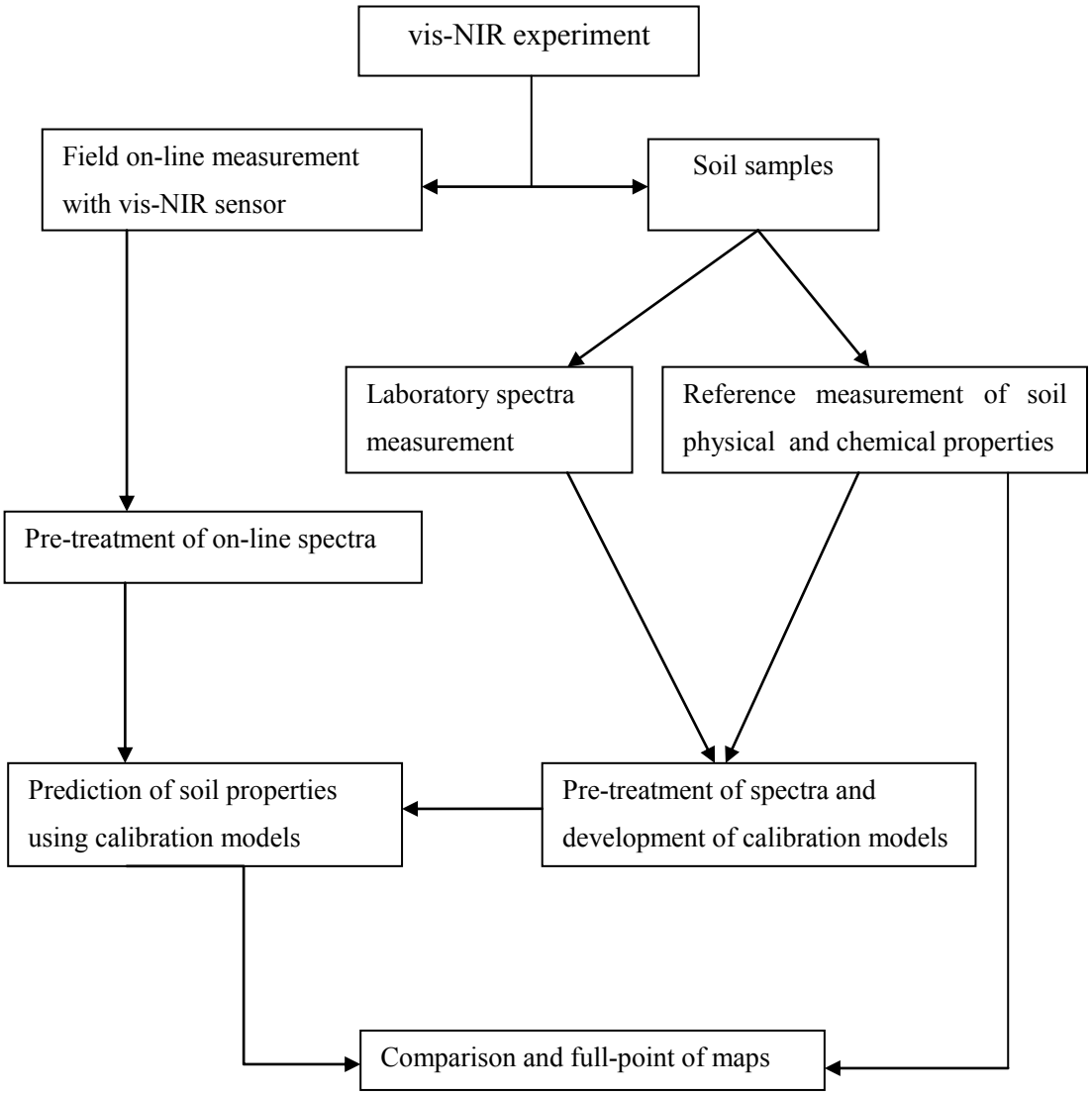

Figure 11. A flow diagram explaining different steps performed from data collection to development of soil maps

6.1.1 Accuracy of calibration models and on-line validation

The PLSR with full cross-validation technique resulted in calibration models for the different soil properties considered in this study. The adequacy of each calibration model was evaluated based on the values of R^2 for predicted versus measured compositions in cross-validation and prediction, and the RPD.

Table 4 shows the calibration models performance in cross-validation. Results illustrate excellent to good relationships between soil properties and soil spectra with R^2 of 0.90, 0.97,

0.74, 0.96, 0.77 and 0.83 for pH, P, TN, TC, MC and K, respectively. In the current study, the results for pH, P and TC are better than those obtained by Kodaira and Shibusawa (2012), whereas the results for TN and MC are worse than those obtained in their study. Root mean square error of calibration (RMSEC) were 0.206, 1.334 (mg.kg^{-1}), 0.010 (%), 0.042 (%), 0.956 (%) and 0.031(cmol+.kg^{-1}) for pH, P, TN, TC, MC and K, respectively. The RPD values of the PLSR cross-validation (Table 4) for pH and P were 1.90 and 2.08, respectively. RMSEP values were 0.358 and 3.96 mg/kg for pH and P, respectively, which is in line with results obtained by others (Shepherd and Walsh, 2002, Cohen et al., 2005, Mouazen et al., 2006). As pH and P do not have direct spectral responses in the NIR range, this successful result can be attributed to covariation with other properties that have direct spectral responses such as N, C and clay (Stenberg et al., 2010).

Table 4. Partial least squares regression (PLSR) – results for calibration datasets.

		Soil properties					
	Statistics	pH	P	TN	TC	MC	K
Calibration	Sample no	125	62	133	108	127	43
	Min	5.39	6.80	0.091	1.207	13.41	0.12
	Max	8.21	40.00	0.198	2.351	24.74	0.37
	Mean	6.59	17.97	0.141	1.806	17.96	0.21
	SD	0.680	8.237	0.020	0.221	2.015	0.073
	RMSEC	0.206	1.334	0.010	0.042	0.956	0.031
	R^2	0.90	0.97	0.74	0.96	0.77	0.83
	LV	7	8	6	13	6	4
Cross-validation	RMSEP	0.358	3.96	0.013	0.101	1.211	0.049
	R^2	0.72	0.77	0.57	0.79	0.64	0.55
	RPD	1.90	2.08	1.53	2.18	1.66	1.48
	Model quality*	B	A	B	A	B	B

P: Phosphorus (mg/kg), TN: Total nitrogen (%), TC: Total carbon (%), MC: Moisture content (%), K: Exchangeable potassium (cmol+. kg^{-1})
SD: Standard deviation
RMSEC: Root mean square error of calibration
RPD: Ratio of prediction to deviation or residual prediction deviation
R^2: Coefficient of determination
LV: Number of latent variables
RMSEP: Root mean square error of prediction
* Model quality is categorized according to Chang et al., 2001.(A: RPD>2.0; B: RPD=1.4–2.0; C: RPD<1.4).

The independent validation of the PLSR calibration models using on-line spectra of the prediction sets (Table 5) provided RPD values of 2.06 and 1.77, and RMSEP values of 0.434 and 8.611 mg/kg for pH and P, respectively, which is better than the result of Mouazen et al., (2009) and Mouazen et al., (2007). RPD values of TN, TC and MC for cross-validation and independent validation were 1.53, 2.18 and 1.66 and 1.77, 1.85 and 1.50 respectively, illustrating good to excellent calibration models. RPD values of K were 1.48 and 1.31 for cross-validation and independent validation, respectively, indicating a poor calibration model

for this property. Literature proves that potassium is the most difficult soil property to be estimated by using vis-NIR spectroscopy (Chang et al., 2001; Mouazen et al., 2006; Cozzolino and Moron, 2003).

Table 5. Validation of partial least squares regression (PLSR) with on-line measured independent validation dataset.

	Independent validation					
Statistics	pH	P	TN	TC	MC	K
Sample no	48	23	22	24	45	24
Min	5.16	4.80	0.112	1.304	13.41	0.12
Max	8.17	50.00	0.201	2.455	24.28	0.40
Mean	6.46	22.50	0.151	1.790	18.03	0.23
SD	0.895	15.231	0.024	0.275	2.164	0.078
RMSEP	0.434	8.611	0.013	0.183	1.449	0.060
R^2	0.73	0.69	0.72	0.57	0.56	0.44
RPD	2.06	1.77	1.85	1.50	1.49	1.31
Model quality*	A	B	B	B	B	C

P: Phosphorus (mg/kg), TN: Total nitrogen (%), TC: Total carbon (%), MC: Moisture content (%), K: Exchangeable potassium (cmol +. kg^{-1})
RMSEP: Root mean square error of prediction
* Model quality is categorized according to Chang et al., 2001.(A: RPD>2.0; B: RPD=1.4–2.0; C: RPD<1.4).

6.1.2 Regression coefficients

Regression coefficients summarize the relationships between all predictors (wavelengths in soil spectra) and a given response (each soil variable). The regression coefficients plots for pH and extractable phosphorus show either positive or negative bands in the vis and NIR ranges (Figures 12-13). Most significant bands are in the visible 400-600 nm, associated with colour and in the NIR range > 1450 nm, associated with water and carbon absorption bands. Although the absorption bands of SOM in the vis–NIR are often weak and not readily apparent to the naked eye, the overall absorption due to SOM in the visible region is broad but clear (Baumgardner et al., 1985; Bowers and Hanks, 1965; Krishnan et al., 1980). Regression coefficients for TC (Figure 15) show that the most significant bands are found in the vis range (480, 579 and 699 nm) and at other wavelengths in the NIR range (747, 792, 984 and 1308 nm). Other studies reported different bands in the vis range (439, 490 and 661 nm) and in the NIR region (1109, 1232, 1414 and 1522 nm) (Mouazen et al., 2007), as compared to the current work. This might be attributed to soil complexity including soil type, parent material and soil physico-chemical characteristics.

The most significant bands (Figure 12) for pH are found in the vis range (444, 498 and 576 nm) and in the NIR range (915, 1105, 1489, 1537 and 1591 nm). Figure 13 shows that the most significant bands for P are found in the vis range (417,459 and 513 nm) and in the NIR range (1045,1225 and 1615 nm). For TN, the most significant bands (Figure 4) are found in the vis range (432 and 555 nm) and in the NIR range (783, 1053 and 1578 nm). Figure 15 shows that the most significant peaks for TC are in the vis range (447, 474, 579 and 696 nm) and in the NIR range at 792, 855, 984, 1158 and 1308 nm. The most significant peaks for MC (Figure 16) are in the vis range (444, 567nm) and in the NIR range (951, 1086 and 1623 nm). For K, regression coefficients (Figure 17) show that the most significant peaks are found in the vis range (414 nm) and in NIR range at 996, 1435, 1609 and 1633 nm.

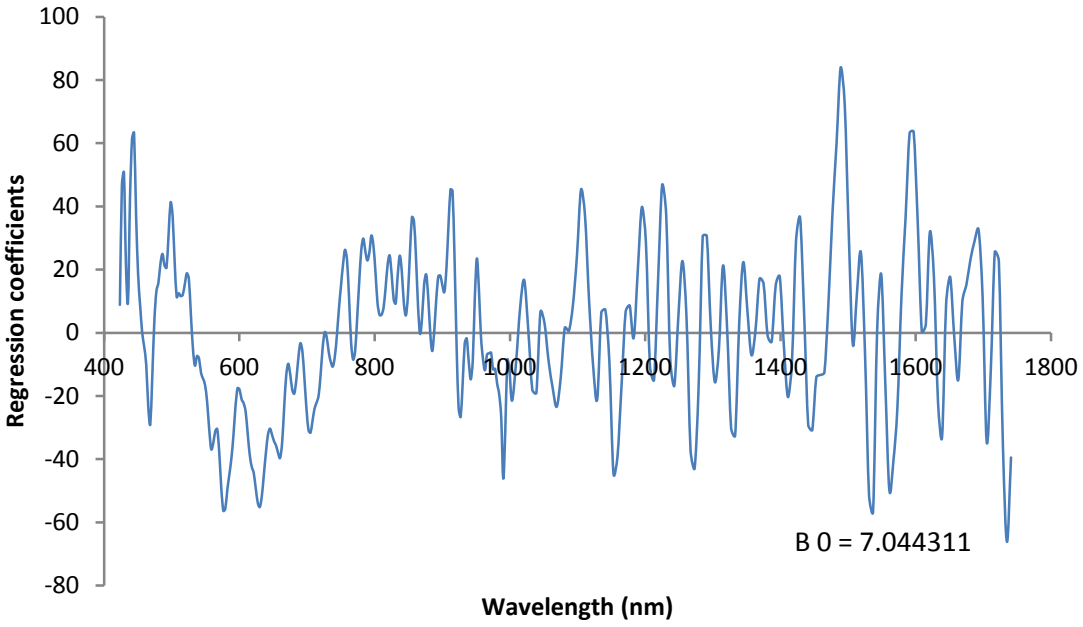

Figure 12. Regression coefficients for soil pH

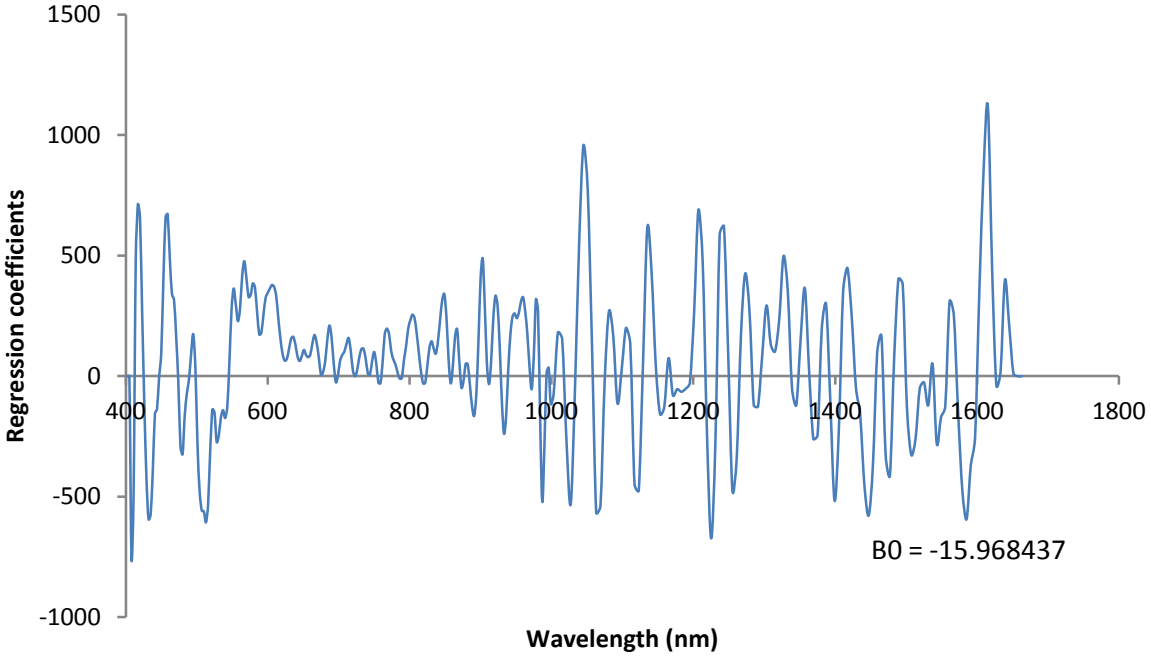

Figure 13. Regression coefficients for extractable phosphorus

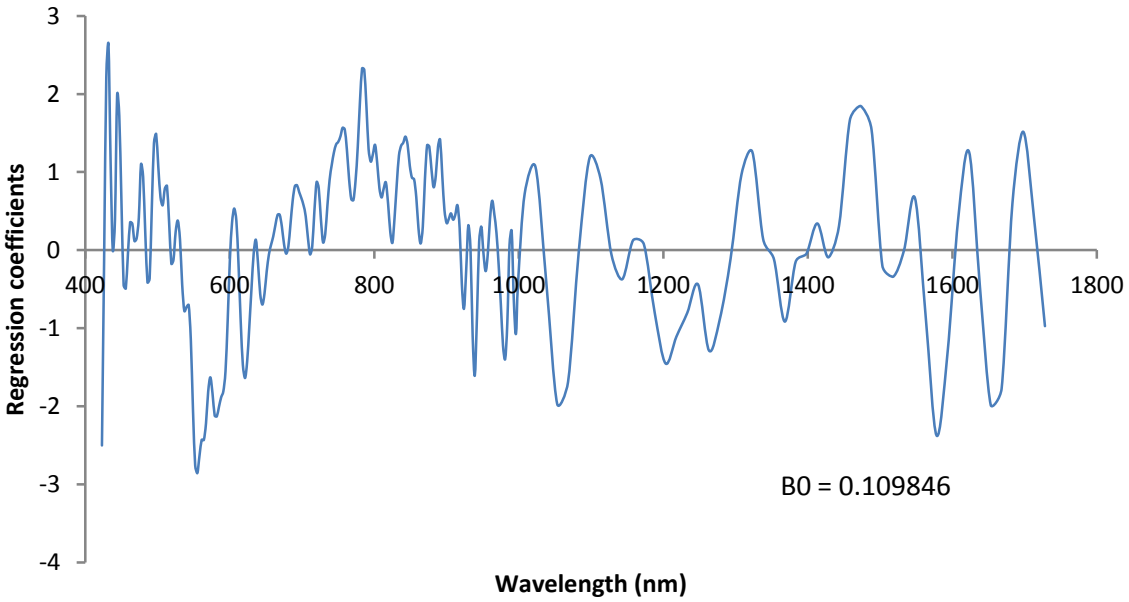

Figure 14. Regression coefficients for total nitrogen

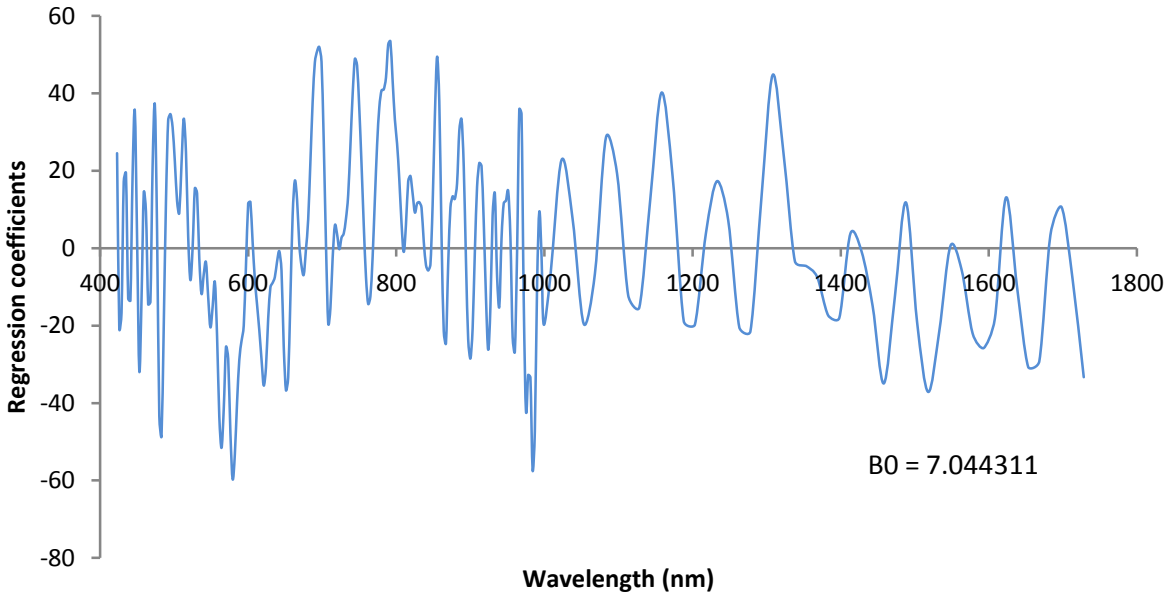

Figure 15. Regression coefficients for total carbon

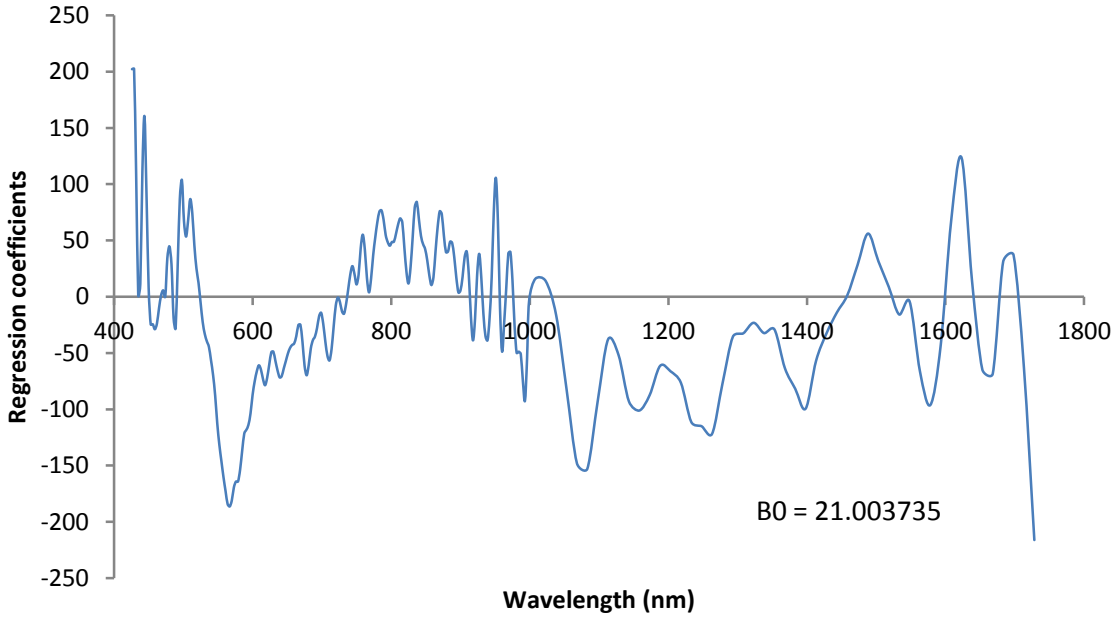

Figure 16. Regression coefficients for moisture content

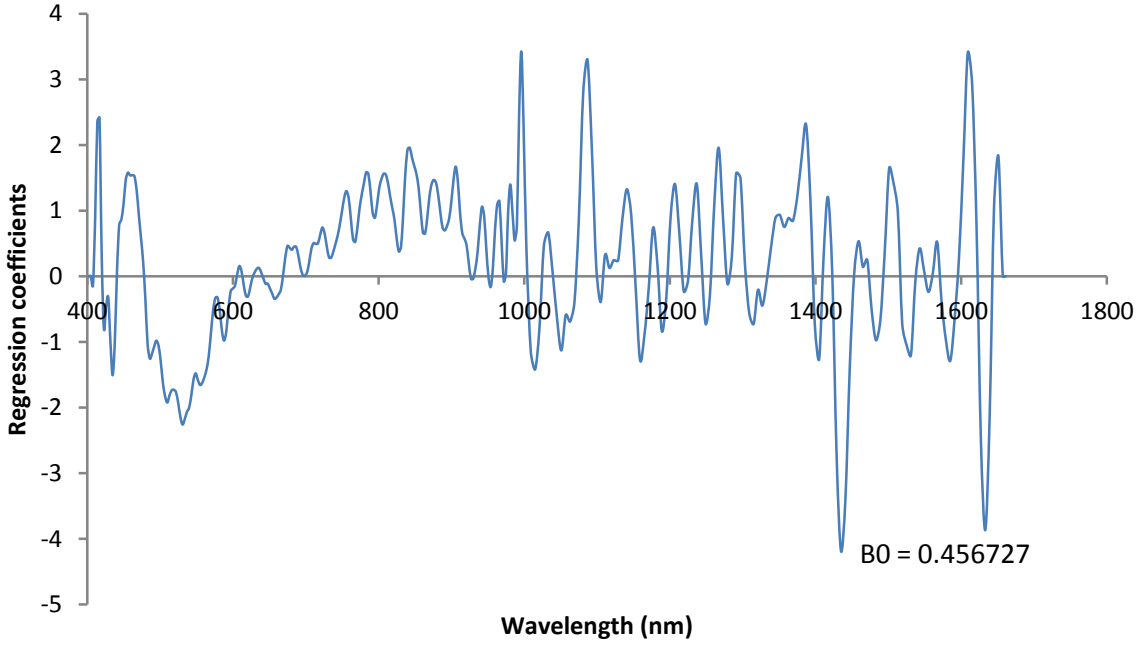

Figure 17. Regression coefficients for exchangeable potassium

6.1.3 Measured versus predicted values in cross-validation and in independent validation data sets

Figures 18-23 show the linear correlation between measured and cross-validation predicted values of each one of the six soil properties. The data points were closely grouped in the cross-validation data set with the exception of exchangeable potassium and slightly more scattered in the independent validation data set (Figures 24-29). Soil samples used in cross-validation were 125, 62, 133,108, 127 and 43 for pH, P, TN, TC, MC, K respectively, whereas samples used in the independent validation data set were 48, 23, 22, 24, 45 and 24 respectively for the same properties motioned above.

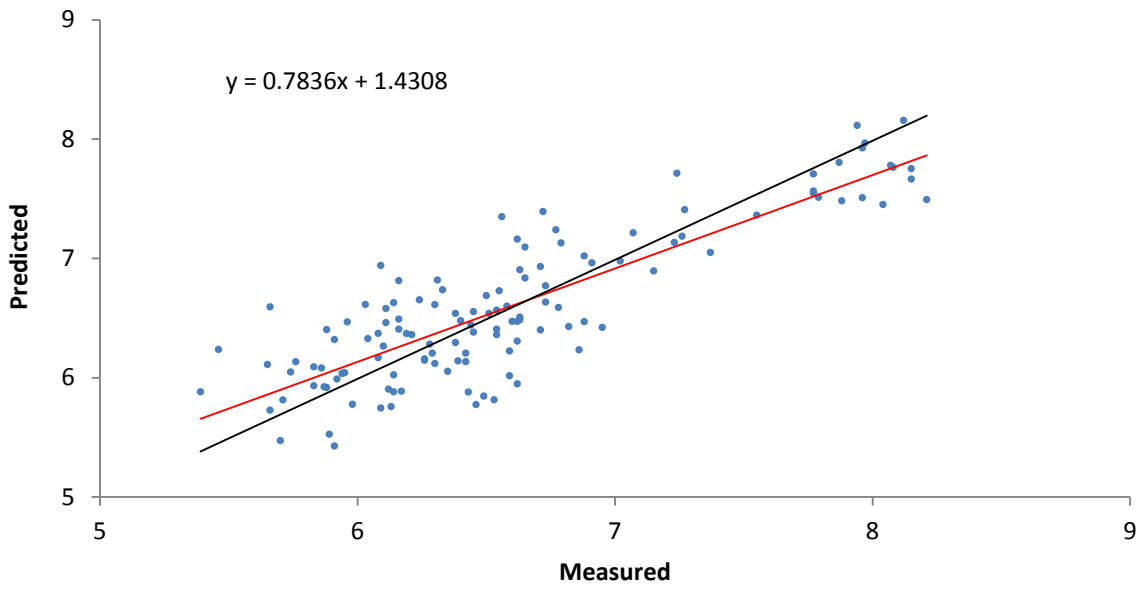

Figure 18. Scatter plot of measured values versus cross-validation predicted pH

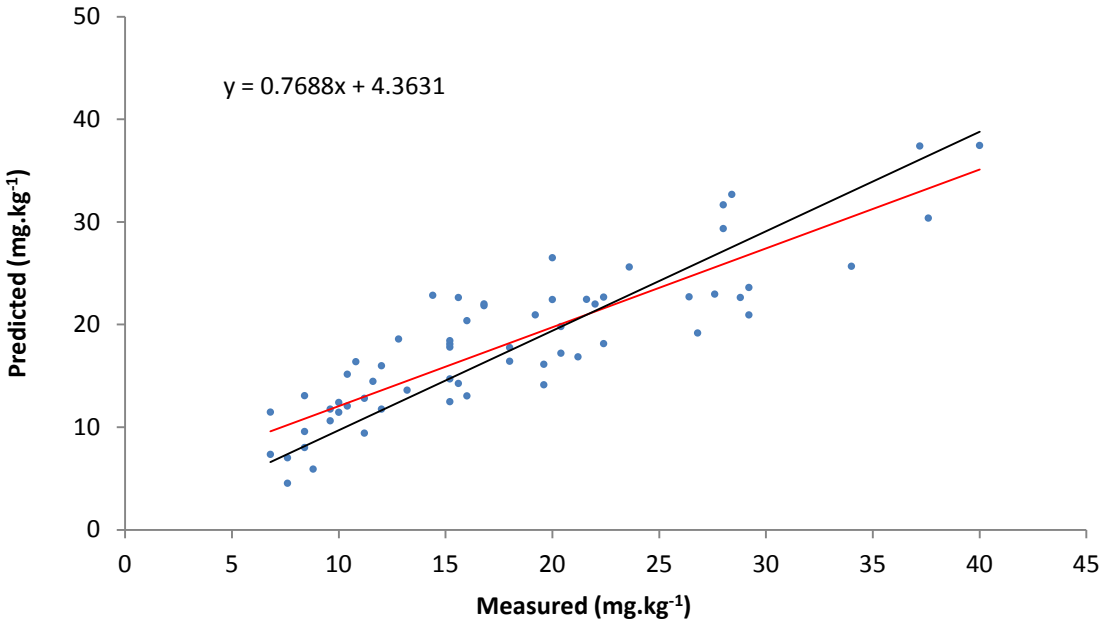

Figure 19. Scatter plot of measured values versus cross-validation predicted extractable phosphorus

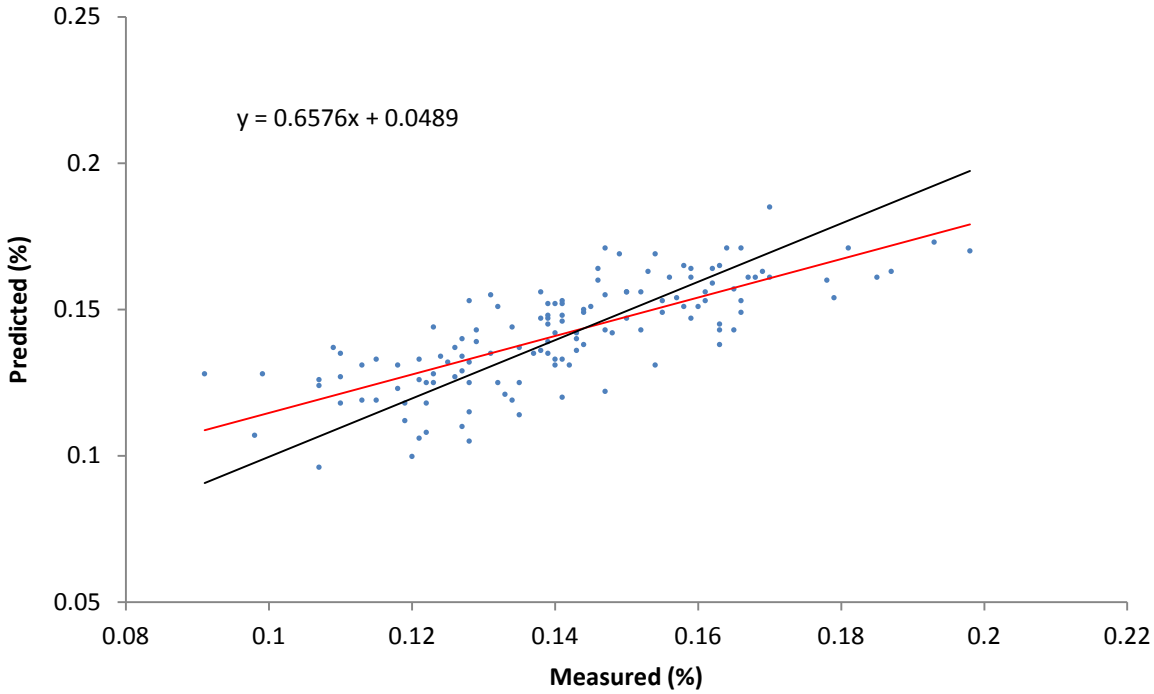

Figure 20. Scatter plot of measured values versus cross-validation predicted total nitrogen

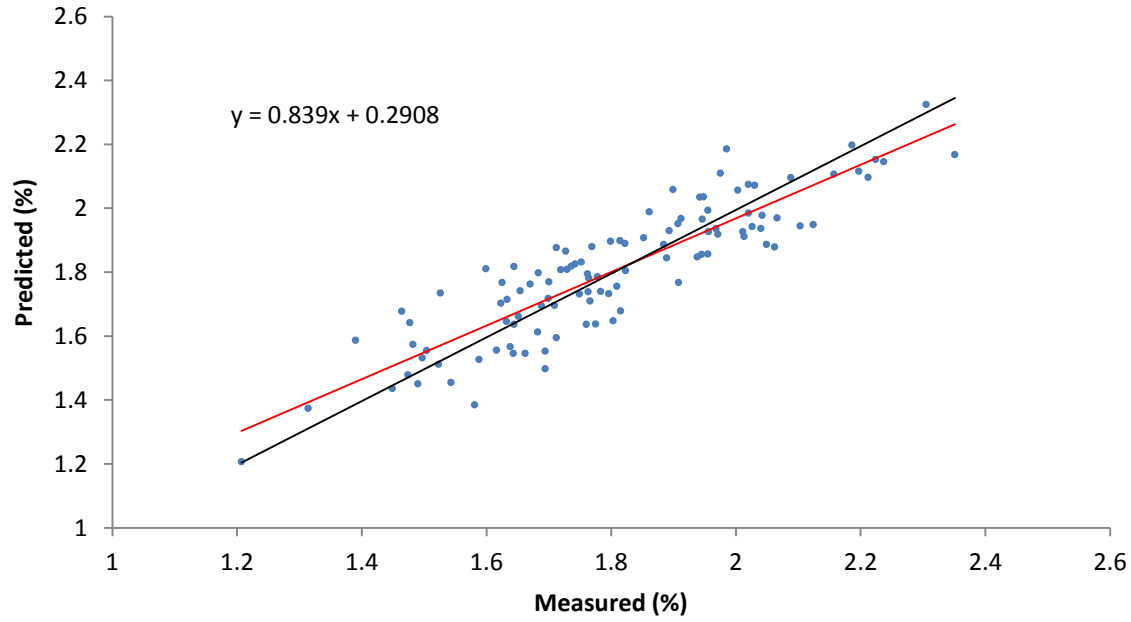

Figure 21. Scatter plot of measured values versus cross-validation predicted total carbon

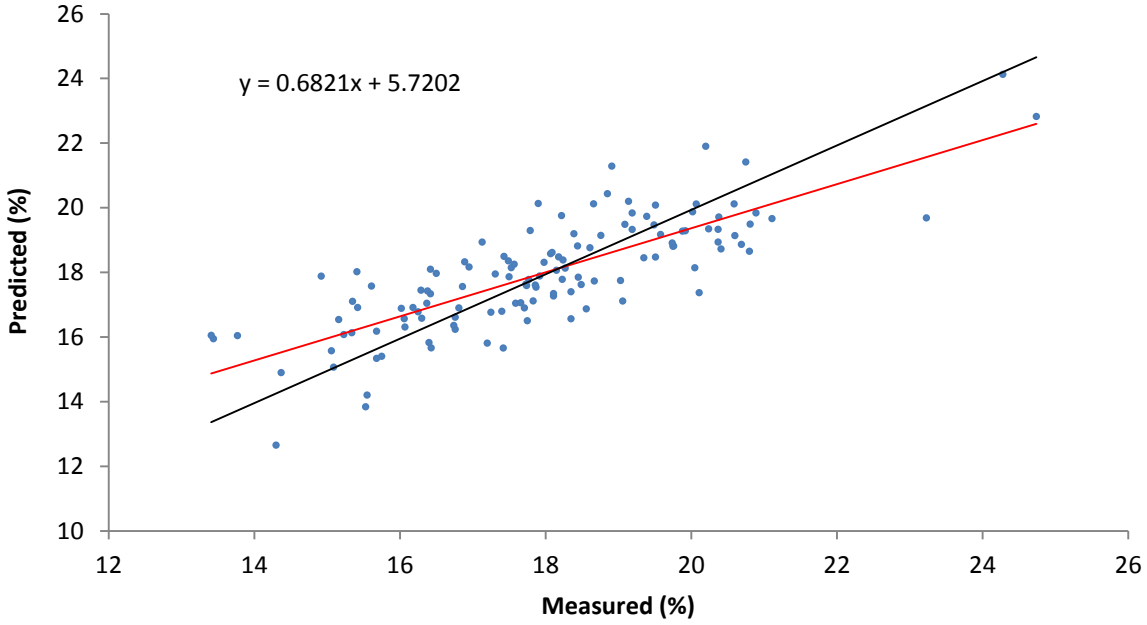

Figure 22. Scatter plot of measured values versus cross-validation predicted moisture content

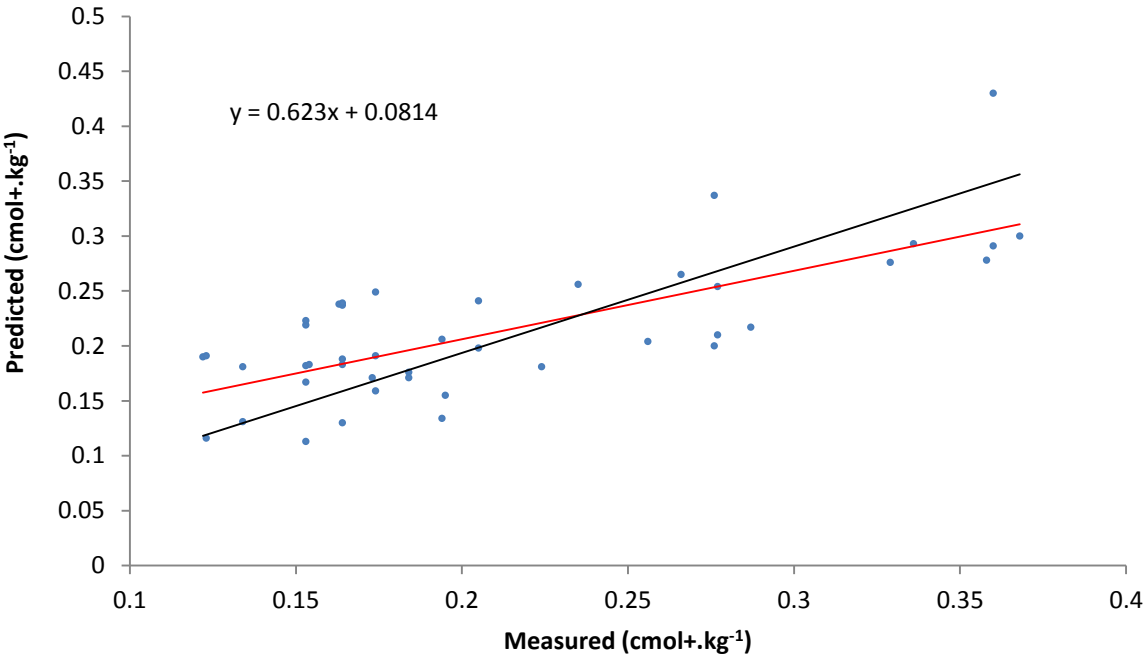

Figure 23. Scatter plot of measured values versus cross-validation predicted exchangeable potassium

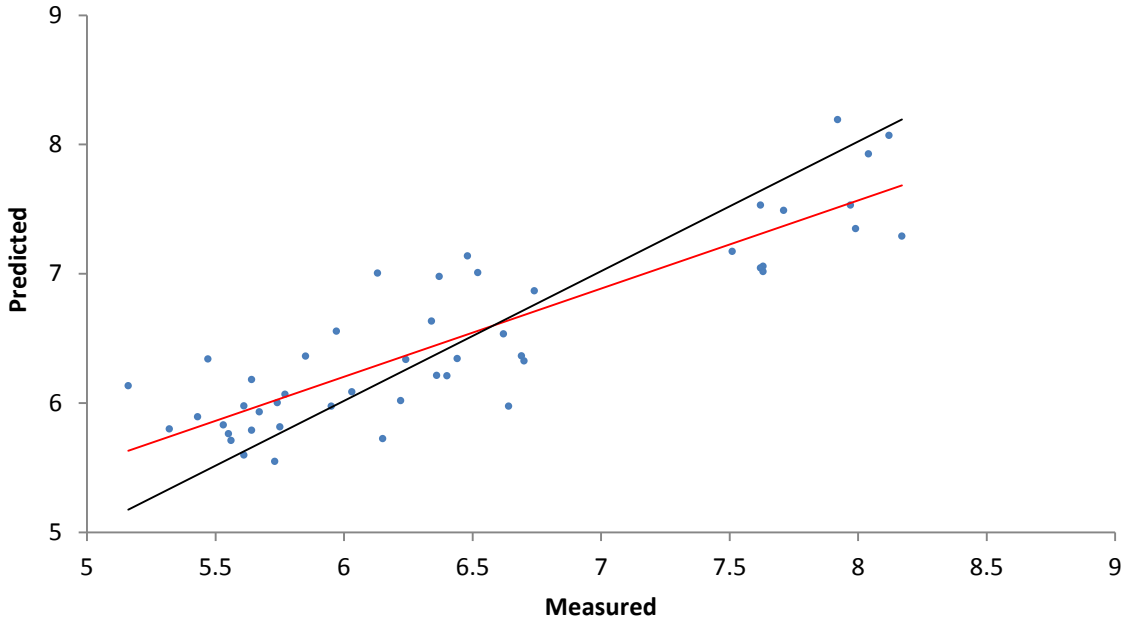

Figure 24. Scatter plot of measured versus predicted pH using the independent validation dataset

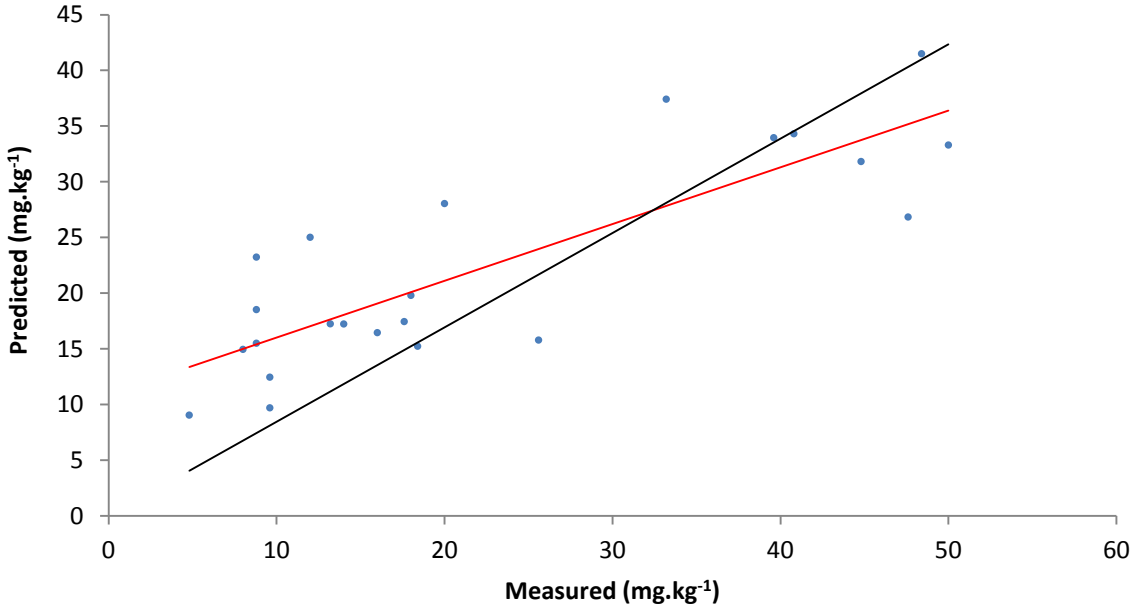

Figure 25. Scatter plot of measured versus predicted extractable phosphorus using the independent validation dataset

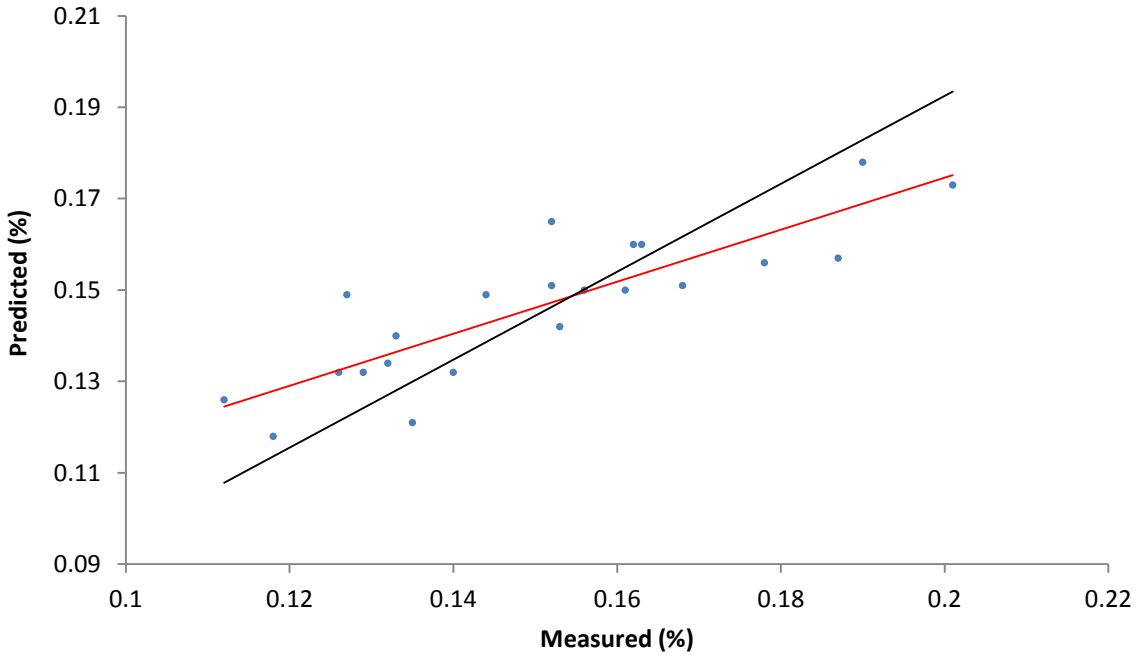

Figure 26. Scatter plot of measured versus predicted total nitrogen using the independent validation dataset

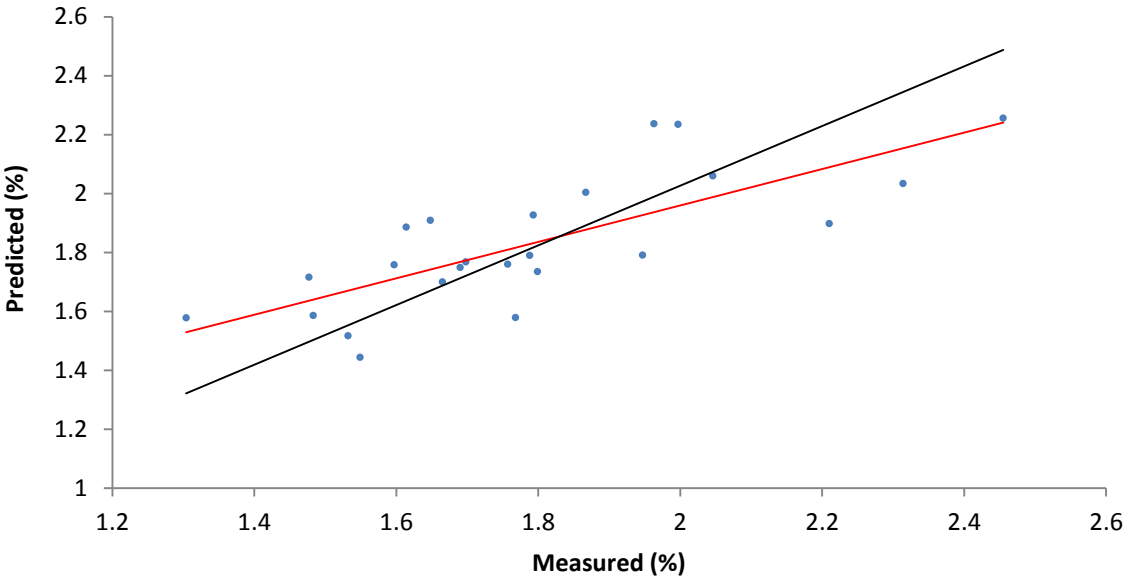

Figure 27. Scatter plot of measured versus predicted total carbon using an independent validation dataset

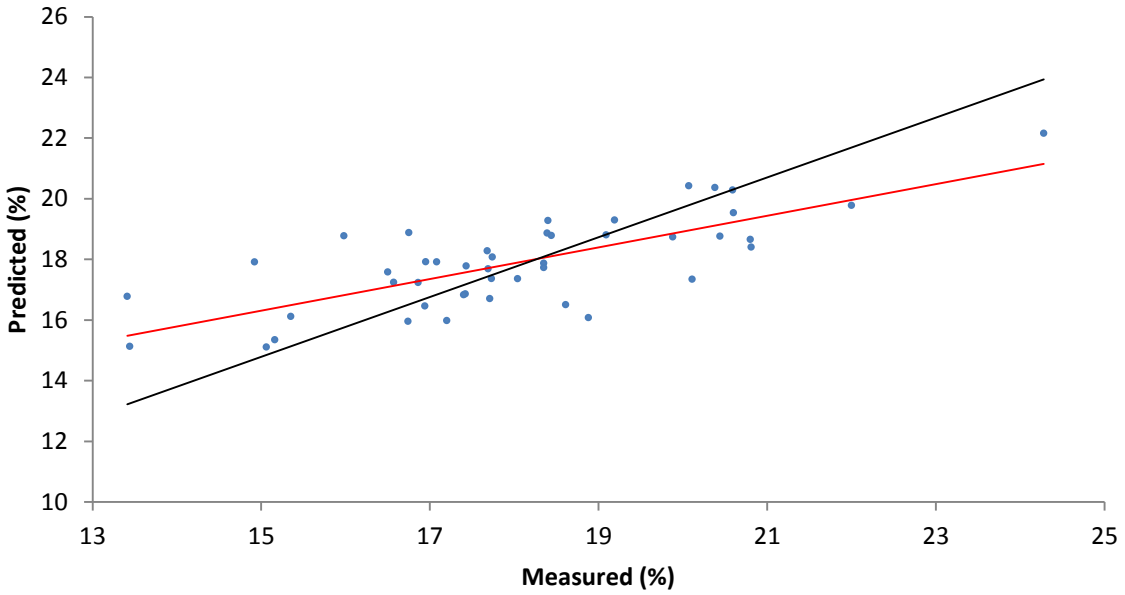

Figure 28. Scatter plot of measured versus predicted moisture content using the independent validation dataset

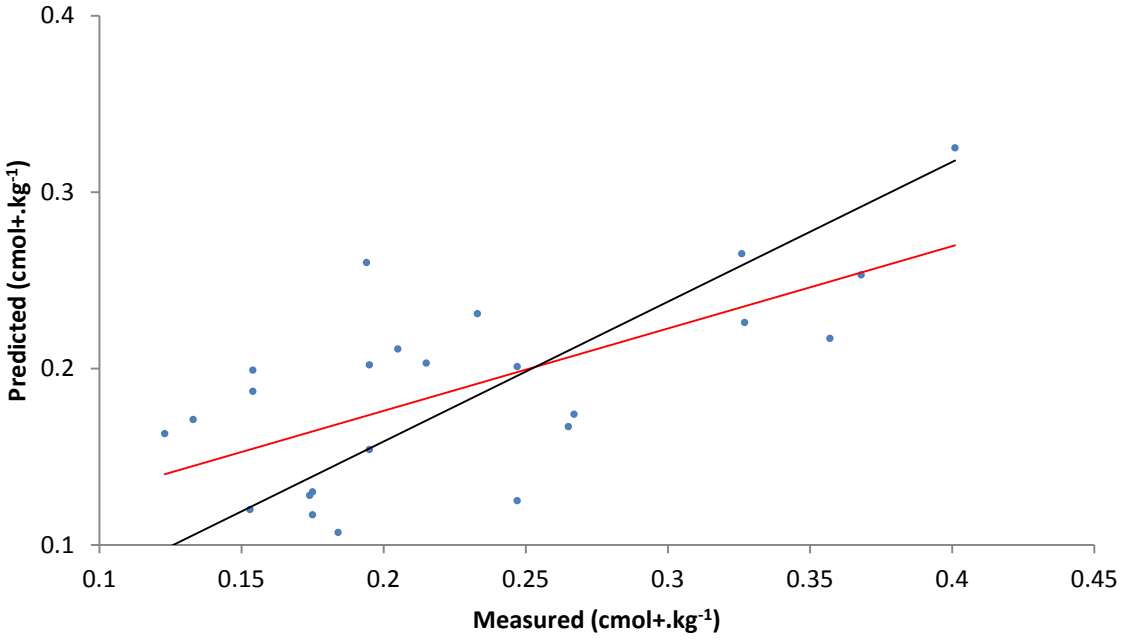

Figure 29. Scatter plot of measured versus predicted exchangeable potassium using the independent validation dataset

6.1.4 Residual variance

The total residual and explained variances show how well the model fits to the data. Models with small total residual variance (close to 0) or large total explained variance (close to 100%) explain most of the variation in the data. Figures 30-35 show small residual variances of the studied properties. Those figures show also the number of latent variables that minimize the residual variances of each of soil properties. Latent variables can be ordered as follow: K (4LV), MC and TN (6LV), pH (7LV), P (8LV) and TC (13LV). The different number of latent variables illustrates the degree of complexity of variation of each soil property.

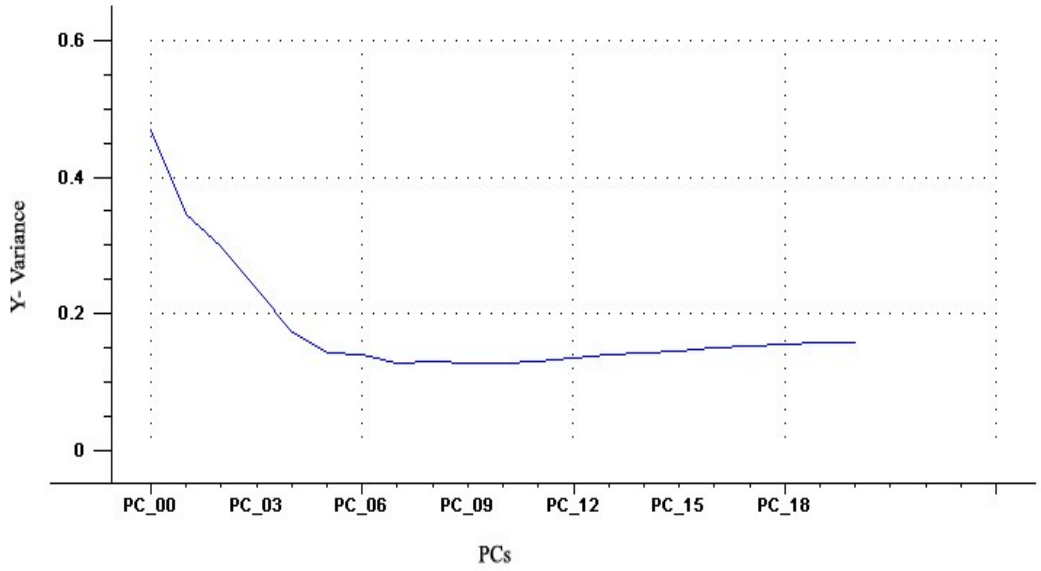

Figure 30. Residual variance for pH in cross-validation

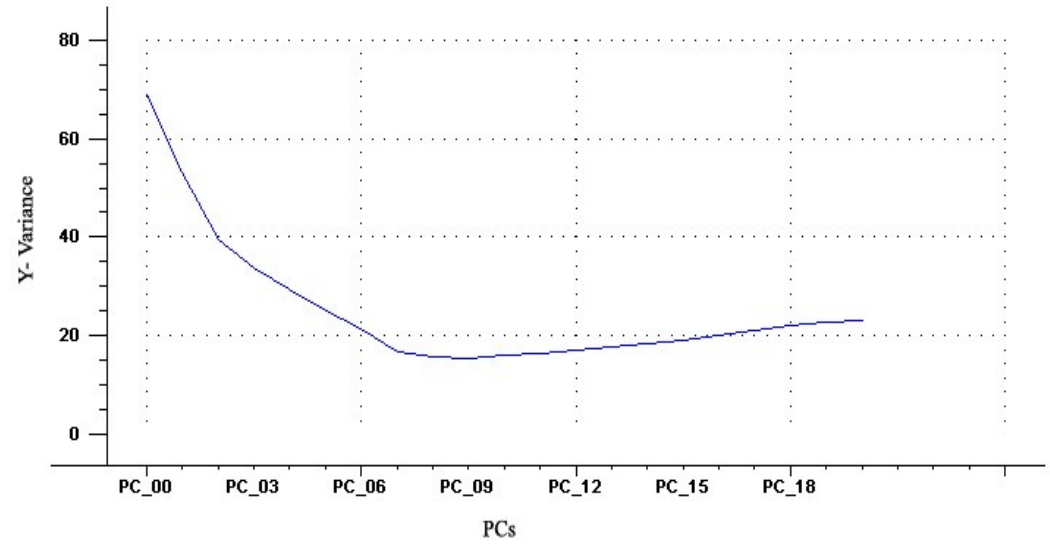

Figure 31. Residual variance for extractable phosphorus in cross-validation

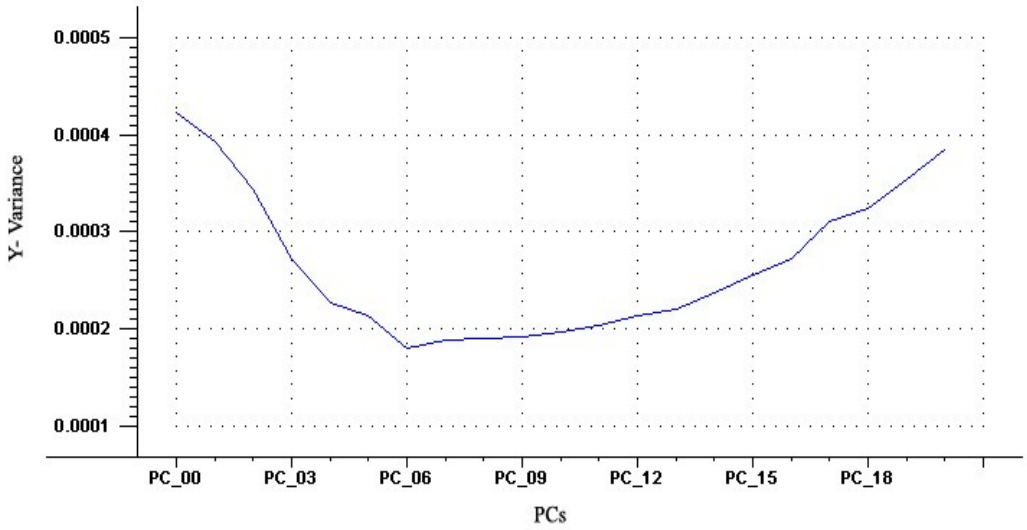

Figure 32. Residual variance for total nitrogen in cross-validation

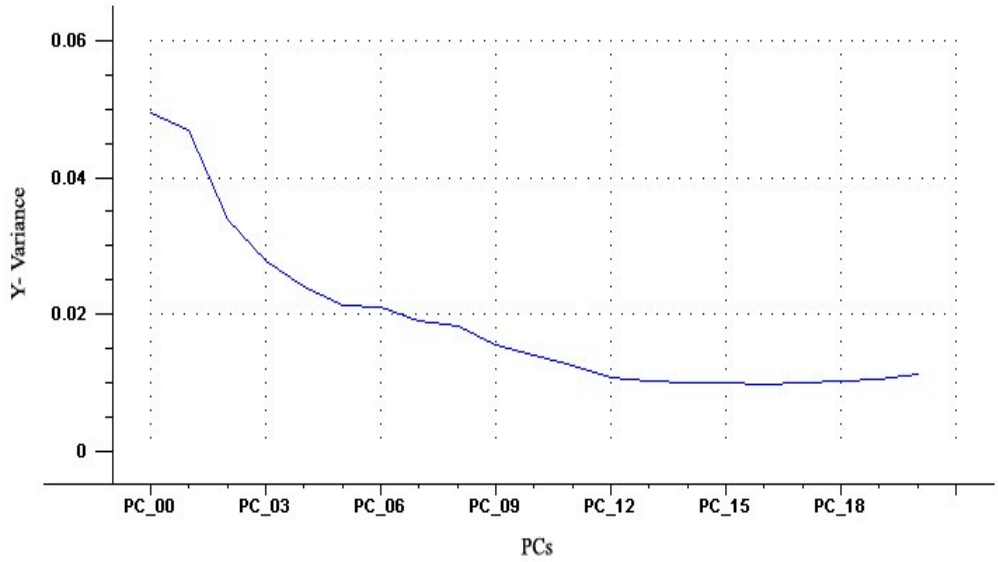

Figure 33. Residual variance for total carbon in cross-validation

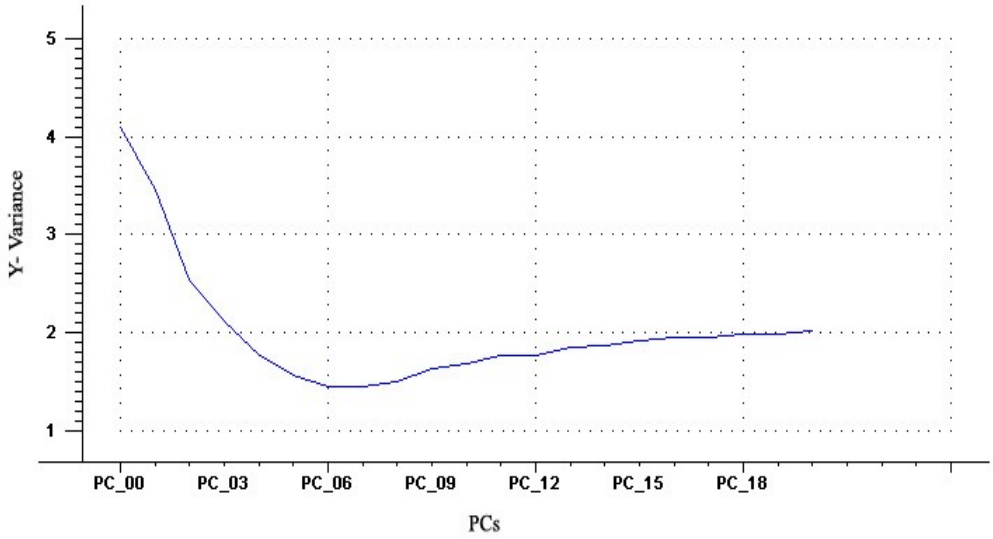

Figure 34. Residual variance for moisture content in cross-validation

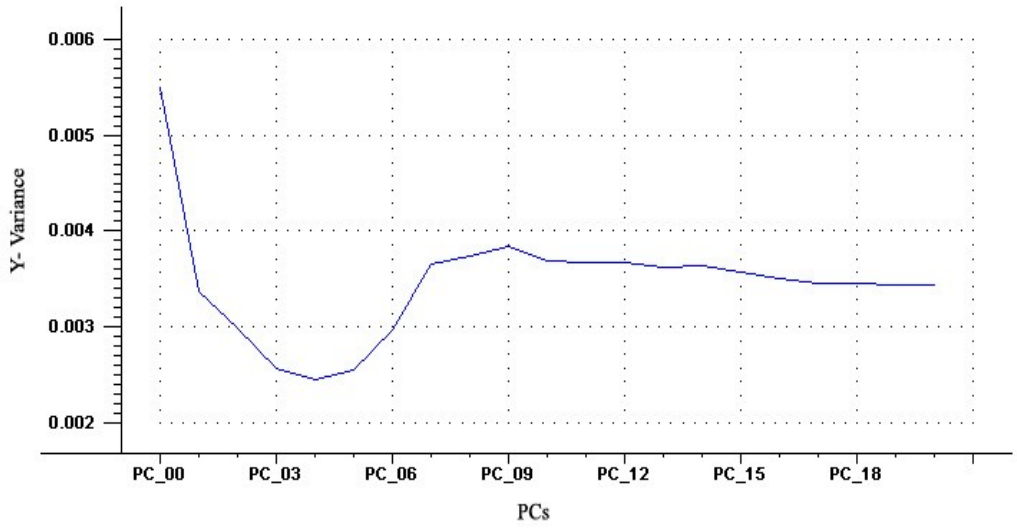

Figure 35. Residual variance for exchangeable potassium in cross-validation

6.1.5 Mapping

The PLS calibration models of pH, extractable phosphorus, total nitrogen, total carbon, moisture content and exchangeable potassium were utilised to transfer the on-line measured soil spectra into quantitative values.

Maps for pH, extractable phosphorus, total nitrogen, total carbon, moisture content and exchangeable potassium developed using the on-line measured VIS-NIR spectra were compared to the corresponding maps developed using the laboratory reference methods. Figure 36 shows a moderate spatial similarity between the measured and predicted values of pH at most areas over the field. Visually the on-line map of soil pH showed a partial similarity compared to the map of reference values. This is due to the large number of points (>11000 points) used to create the on-line map, whereas only few points were used to develop the latter map. Error map of pH shows that the maximum error occurred in few areas in the studied field. The mean error between the on-line and reference measurement methods was 3 %. This confirms a good prediction of soil pH that can help the application of site-specific fertiliser management. Thus, the on-line measurement system can be used to provide detailed data about the spatial variation of soil pH that can be useful information to adjust soil acidity by lime.

Maps for extractable phosphorus (Figure 37) show a good spatial similarity between measured and predicted values at most areas over the field. The spatial similarity between the on-line and measured maps is poor in this case. The average calculated error between the two measurement methods of P was quite large (38 %), which can be attributed to the small number (62) of samples used to establish the calibration model of P (Table 4). Therefore, a larger sample set might improve the accuracy of the on-line measurement of P. Mouazen et al., (2007) used 175 samples to develop the calibration model and obtained an average error of 27.5 % between on-line and reference measurement methods.

Map of measured values of total nitrogen shows a good spatial similarity, as compared to the corresponding map of predicted values (Figure 38). The spatial distribution of total nitrogen measured by the on-line and by reference methods shows partial similarity. The mean error between the on-line and reference measurement method was very small (8 %), which is a good indicator that the on-line sensor performance provides appreciable accuracy.

For total carbon, there is a spatial similarity between reference and predicted values (Figure 39). Also, there is a partial similarity between the on-line and reference measurement

methods. The mean error between the two measurement methods was small (9 %). Mouazen et al., (2007), reported a smaller average error of 6 % between on-line and reference measurement methods.

Maps for soil moisture content (Figure 40) show a partial spatial similarity between reference and predicted values. The calculated mean error between the on-line and reference was 23 %, which is very high. The spatial similarity between the two measurement methods is poor, which might be attributed to the small range of soil moisture content (13 % - 24 %) used to create the calibration model. An average error of 0.37 % was obtained by Mouazen et al., (2007) with a wider range of soil moisture content (2.2 % - 36 %).

Figure 41 shows a moderate spatial similarity between the reference and predicted maps of exchangeable potassium. The mean error between the on-line and reference measurement methods was small (8 %). The spatial similarity between the on-line and reference methods is good.

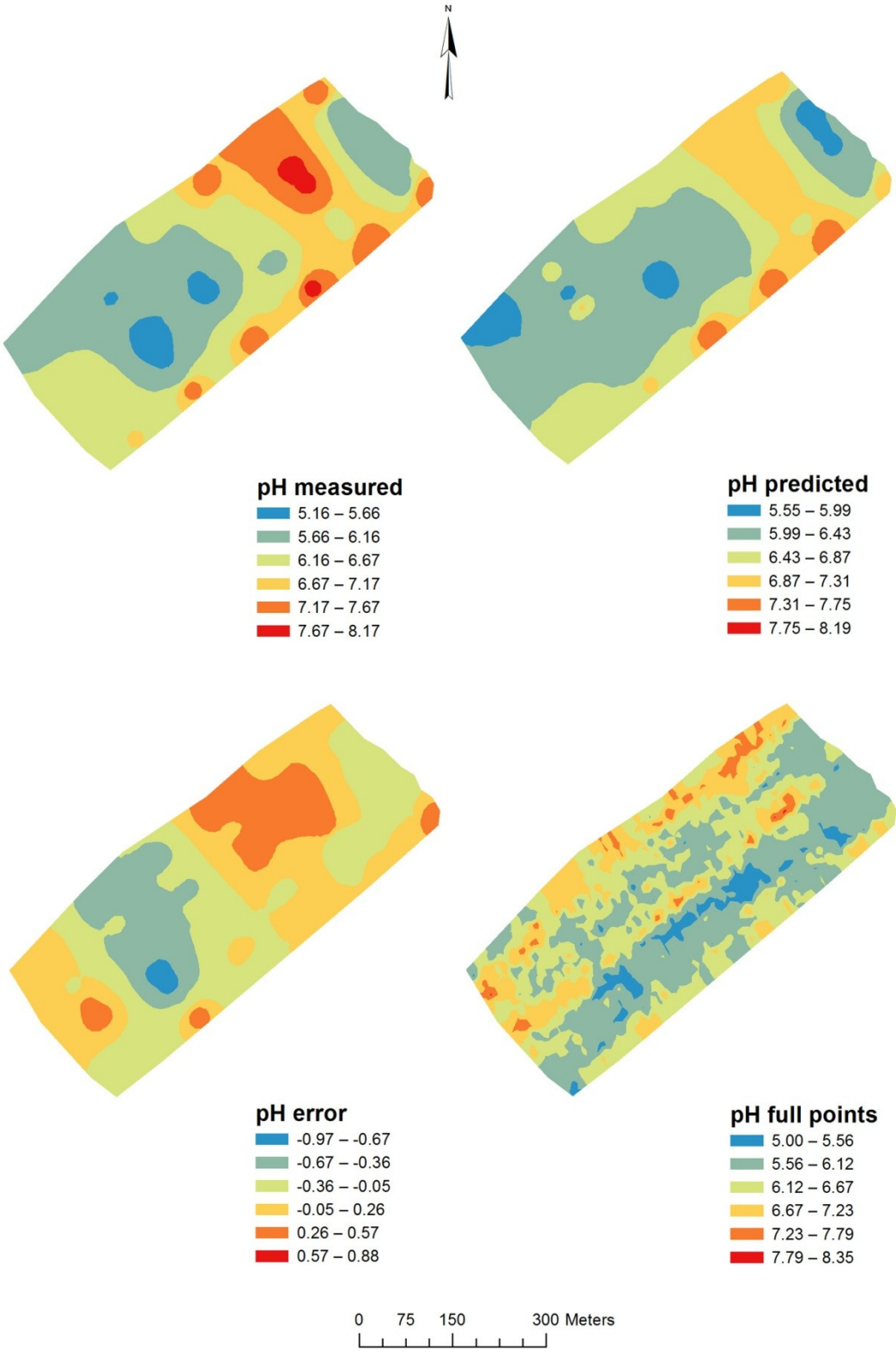

Figure 36. pH measured, predicted, error and full points maps

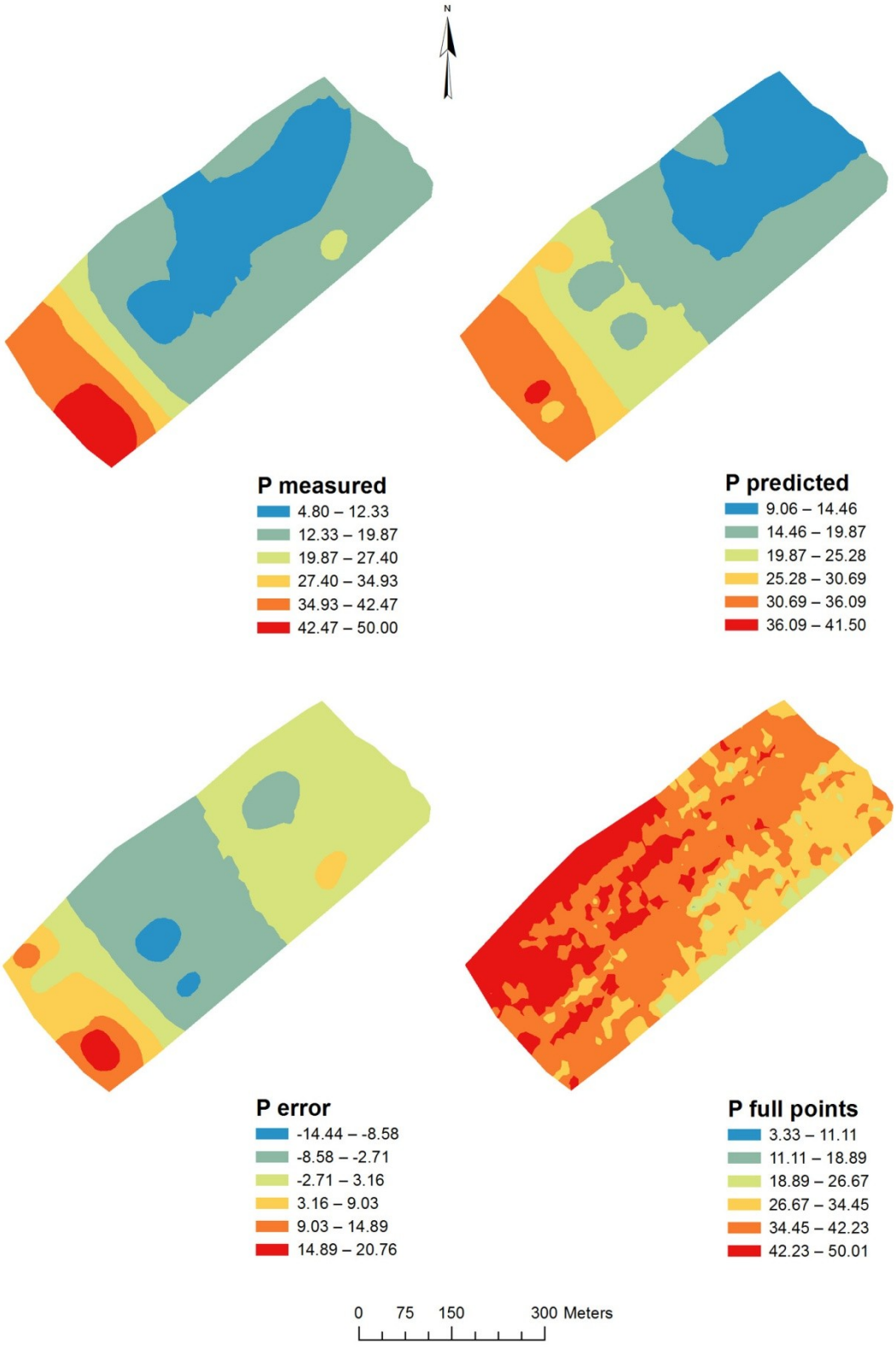

Figure 37. Extractable phosphorus measured, predicted, error and full points maps

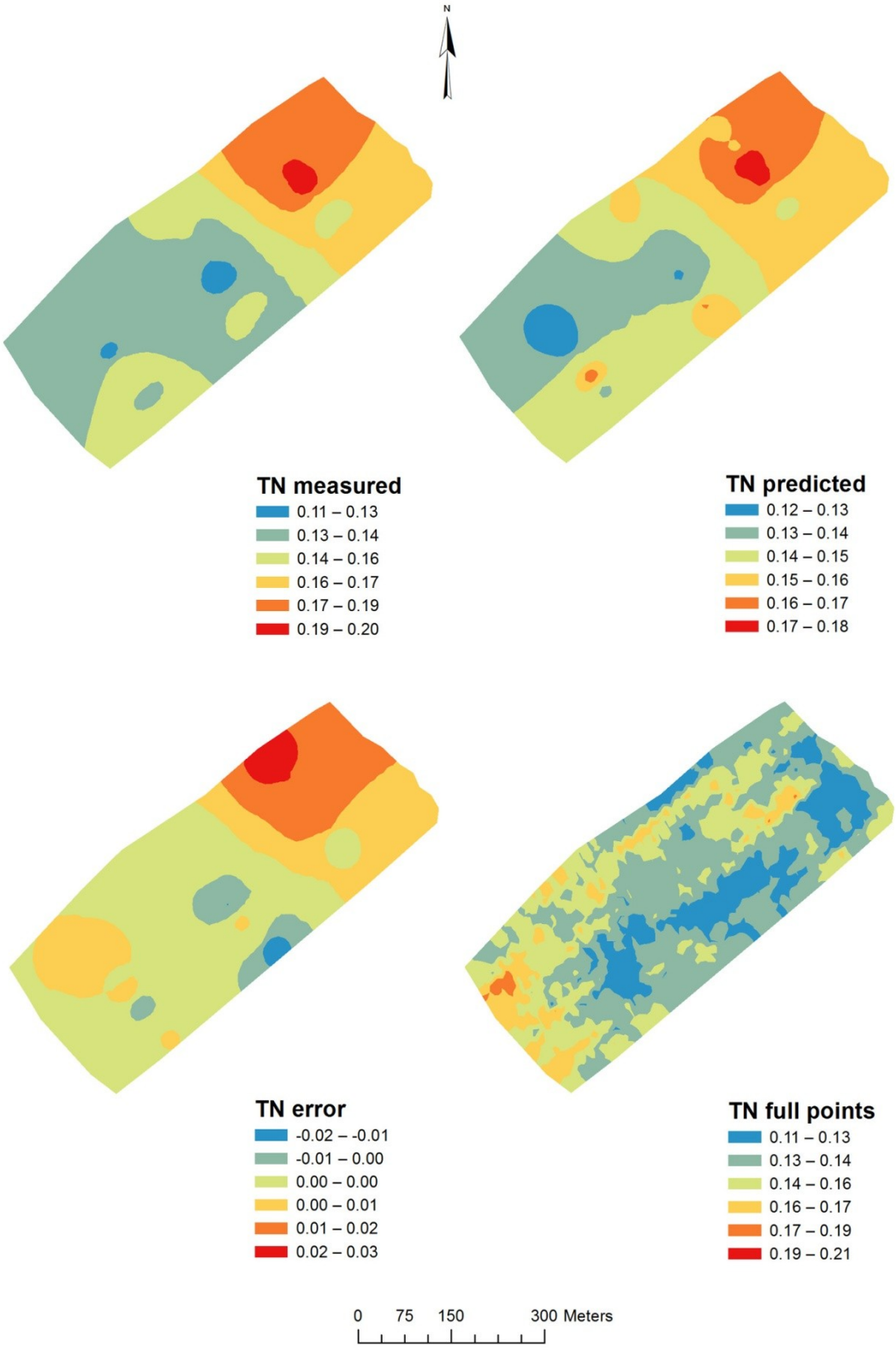

Figure 38. Total nitrogen measured, predicted, error and full points maps

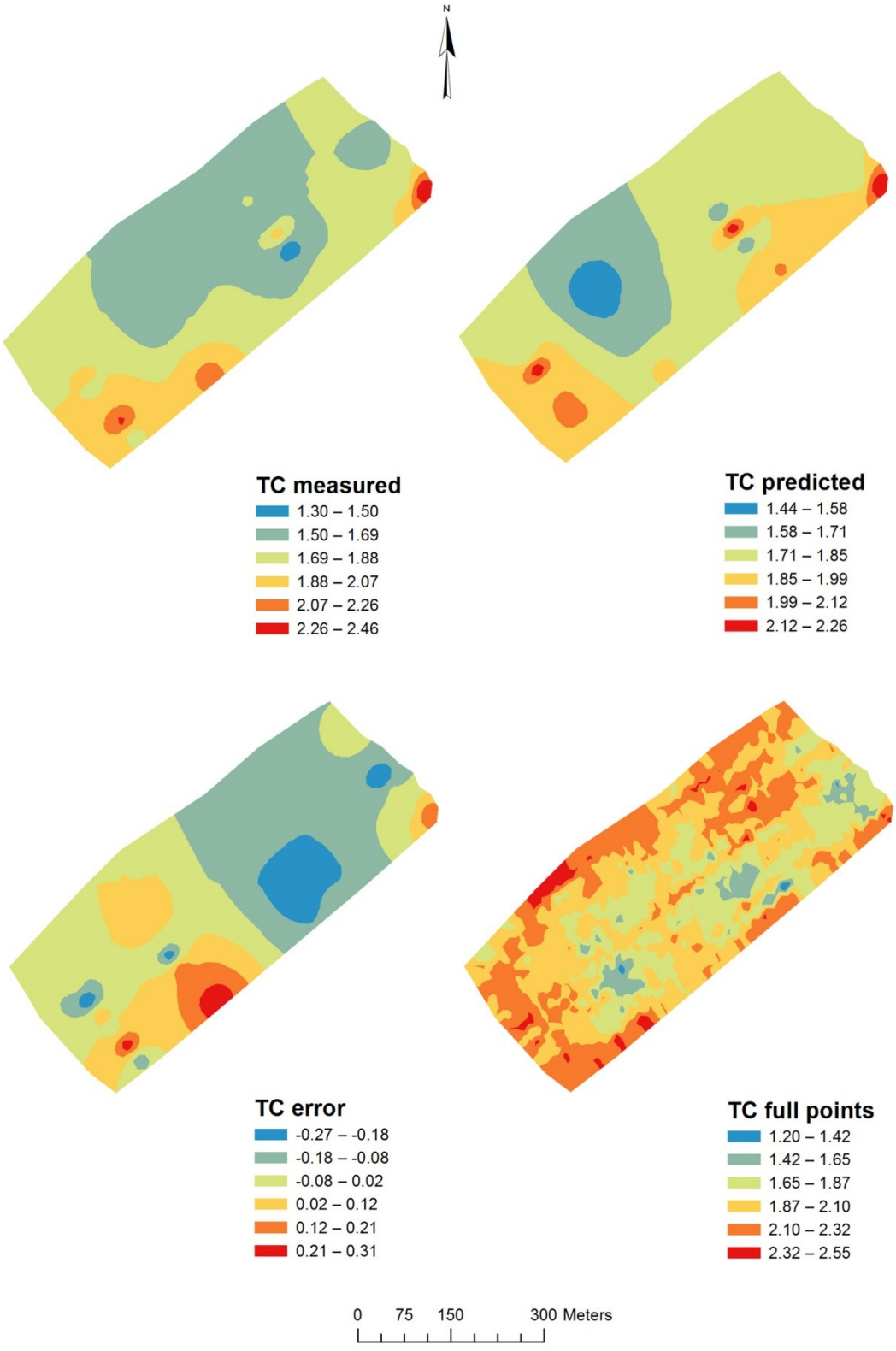

Figure 39. Total carbon measured, predicted, error and full points maps

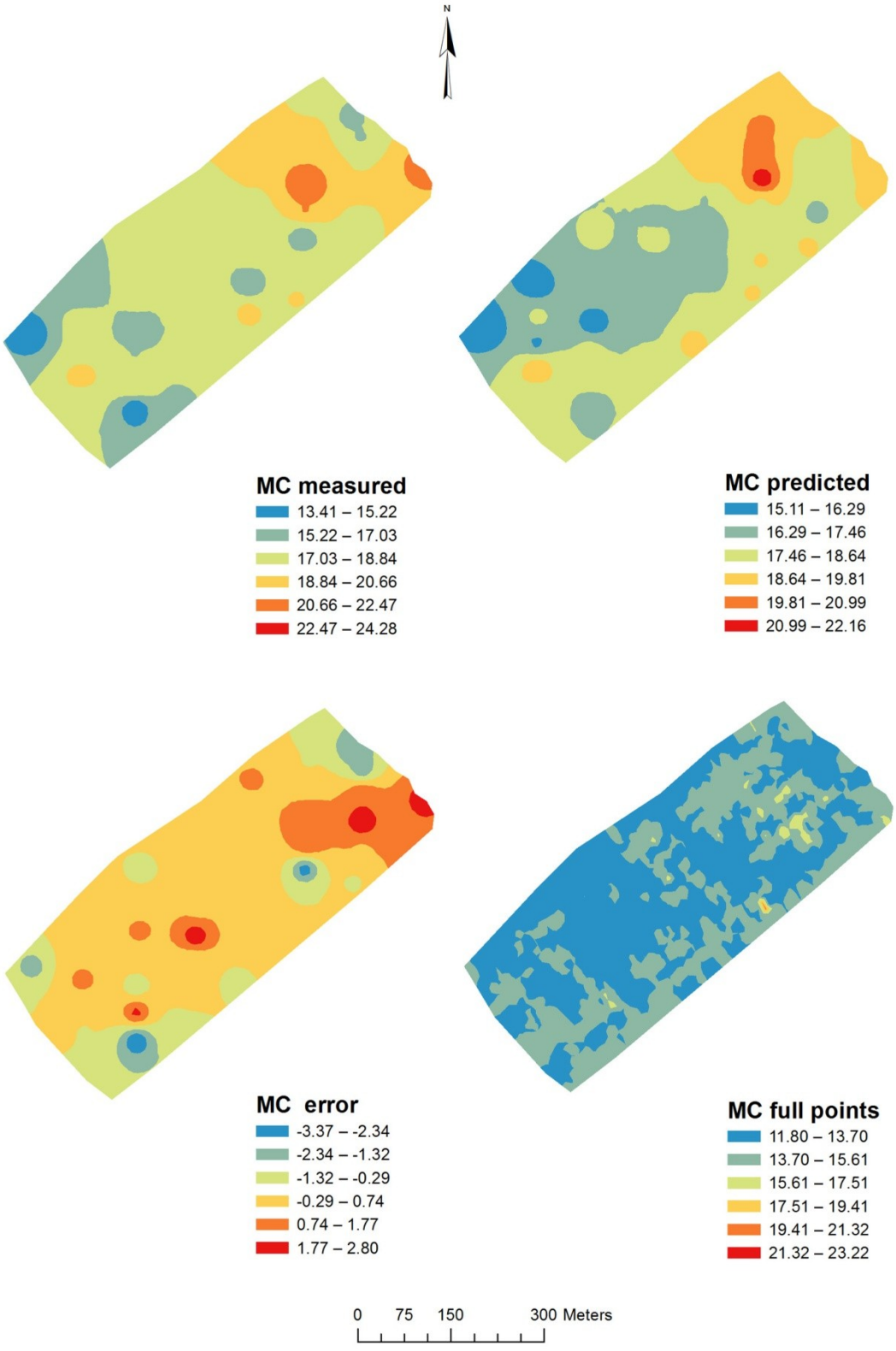

Figure 40. Moisture content measured, predicted, error and full points maps

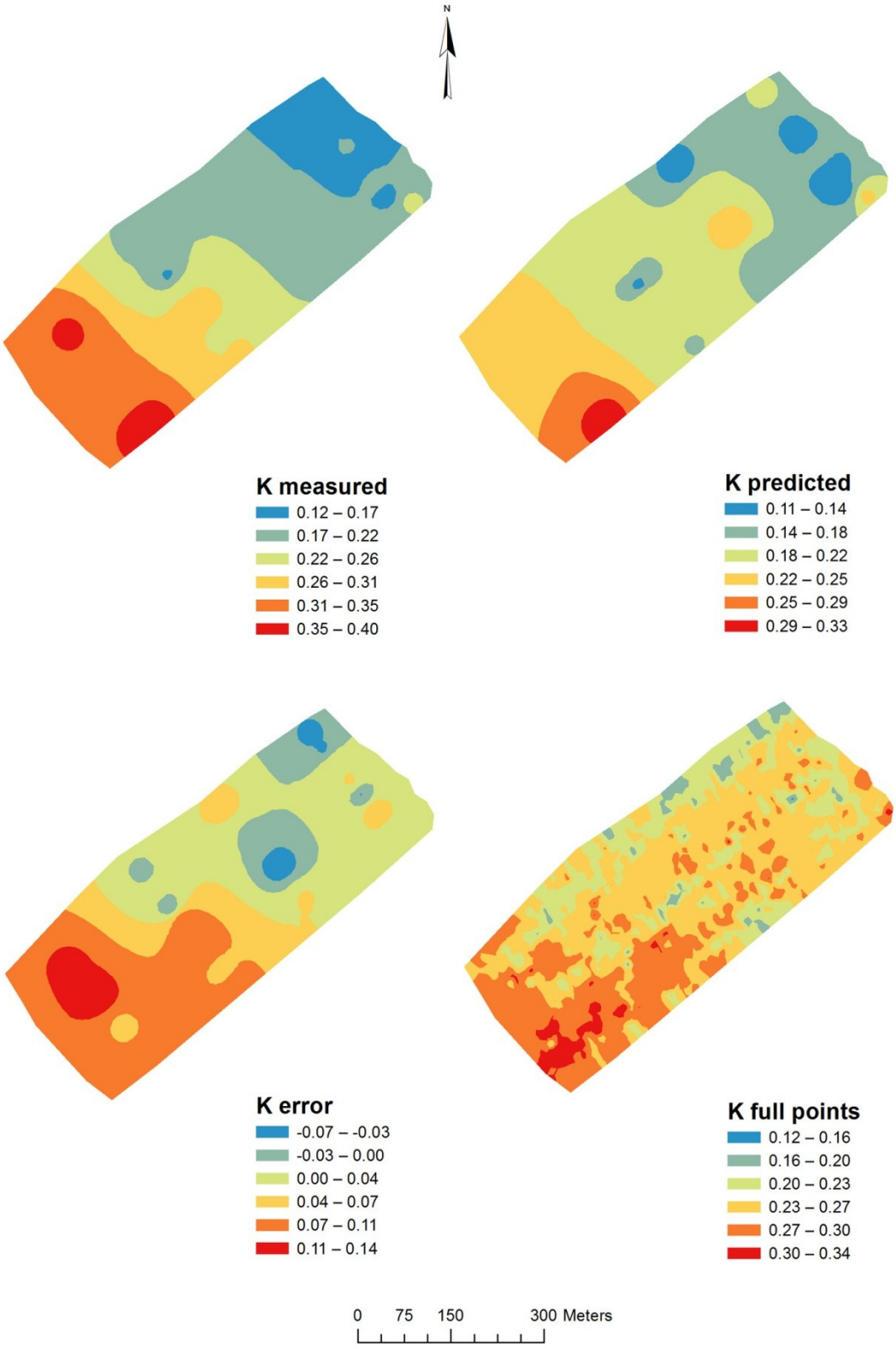

Figure 41. Exchangeable potassium measured, predicted, error and full points maps

6.2 Part B: Geostatistical analysis

The descriptive statistics of soil calibration data set are summarized in Table 6. Most soil properties are markedly skewed except silt, total carbon and sand. For this reason data transformation by Gaussian anamorphosis was carried out.

Table 6. Statistics of raw data of soil calibration data set

VARIABLE	Count	Minimum	Maximum	Mean	Std. Dev	Skewness	Kurtosis
Sand (%)	45	42.93	61.25	52.93	4.52	-0.50	2.41
Silt (%)	45	24.02	34.25	29.91	2.42	-0.10	2.29
TC (%)	120	1.08	2.50	1.78	0.23	0.39	4.07
TN (%)	120	0.11	0.24	0.15	0.02	0.89	4.45
CEC (cmol+.kg^{-1})	61	7.12	11.80	8.94	1.04	0.67	3.17
K (cmol+.kg^{-1})	61	0.05	0.27	0.13	0.06	0.78	2.58
P (mg.kg^{-1})	61	6.80	65.60	23.38	14.10	1.07	3.40
pH	120	5.12	8.15	6.49	0.73	0.72	2.82

Table 7 lists the Pearson correlation coefficients of all soil properties used in calibration. Sand was negatively correlated with silt, CEC and positively with P and K. TC was positively correlated with TN, pH and CEC. P was positively correlated with K and negatively with silt. It is commonly known that soil organic matter and clay are colloidal materials with large surface area that increase the exchange capacity of the soil and this explains the positive correlation between TC and CEC and the negative correlation between sand and CEC.

Table 7. Correlation matrix of soil calibration data set

Variables	CEC	pH	K	P	TN	TC	Sand	Silt
CEC	**1.00**	0.17	-0.16	-0.23	**0.40**	**0.46**	**-0.74**	**0.41**
pH		**1.00**	**-0.33**	-0.12	**0.46**	**0.55**	-0.06	-0.28
K			**1.00**	**0.70**	-0.05	0.16	**0.36**	**-0.38**
P				**1.00**	0.21	0.19	**0.44**	**-0.56**
TN					**1.00**	**0.70**	-0.15	-0.18
TC						**1.00**	-0.09	**-0.29**
Sand							**1.00**	**-0.83**
Silt								**1.00**

Values in bold are different from 0 with a significance level alpha=0.05

Table 8 shows the correlation matrix of the spectral variables. The correlations were generally quite low or not significant. The significant correlations of MC with the other spectral variables might be due to the fact that the absorption peaks of water (e.g. at 1950 nm) can mask the wavebands associated with the absorption by other soil attributes and thus deteriorate model performance and accuracy of prediction of these properties (e.g. OC, clay, etc) (Mouazen et al., 2005). This negative effect of MC explains why the calibration of the spectral model did not work properly.

Table 8. Experimental correlation matrix of spectral model

Variables	K spec	MC spec	P spec	pH spec
K spec	**1.00**	**-0.04**	0.02	**-0.17**
MC spec		**1.00**	**-0.34**	**0.04**
P spec			**1.00**	**0.12**
pH spec				**1.00**

Values in bold are different from 0 with a significance level alpha=0.05

Table 9 shows the correlation matrix of the Gaussian transformed soil and spectral variables used to develop the fusion model. Exchangeable K (spectral variable) was positively correlated with sand content, silt content and soil K, and to a less extent with P (soil variable) and pH (spectral variable). Moisture content was positively correlated with TC, TN and pH (soil variable). P (spectral variable) was negatively correlated with TC, TN and pH (soil variable). pH (spectral variable) was positively correlated with pH (soil variable), whereas negatively with K (soil variable).

It is clear from the previous results, that K (spectral variable) is the most correlated spectral variable with soil variables and that is why it was used as auxiliary variable in the multi-collocated cokriging analysis.

Table 9. Experimental correlation matrix of fusion model

Variables	[1]gKspec	gMC spec	gP spec	gpH spec	gP	gTC	gTN	gCEC	gK	gpH	gSand	gSilt
gK spec	**1.00**	**0.22**	**0.05**	**0.38**	**0.37**	0.17	0.08	0.19	**0.51**	0.02	**0.60**	**0.53**
gMC spec		**1.00**	**-0.21**	**0.23**	0.01	**0.54**	**0.48**	**0.32**	-0.01	**0.44**	-0.06	-0.24
gP spec			**1.00**	**0.28**	0.07	**-0.41**	**-0.18**	-0.17	0.12	**-0.34**	0.17	0.10
gpH spec				**1.00**	-0.02	**0.18**	**0.27**	0.11	**-0.26**	**0.50**	-0.14	-0.10
gP					**1.00**	0.14	**0.24**	**-0.29**	**0.58**	-0.14	**0.53**	**-0.58**
gTC						**1.00**	**0.70**	**0.53**	0.10	**0.56**	-0.10	**-0.32**
gTN							**1.00**	**0.39**	-0.07	**0.40**	-0.11	-0.22
gCEC								**1.00**	-0.14	0.19	**-0.62**	**0.34**
gK									**1.00**	**-0.34**	**0.36**	**-0.34**
gpH										**1.00**	-0.08	**-0.34**
gSand											**1.00**	**-0.80**
gSilt												**1.00**

Values in bold are different from 0 with a significance level alpha=0.05
[1] *g means gaussian*

Table 10 shows the LMC of soil model. LMC of soil model includes two basic spatial structures: a nugget effect; an isotropic spherical model with range of 343.96 m.

Table 10. Linear model of coregionalization of the transformed Gaussian soil variables. There are reported the coregionalization matrices (sills of variograms)

Nugget effect								
	gP	gTC	gTN	gCEC	gK	gpH	gSand	gSilt
gP	0.3452	0.1294	0.1899	0.0214	0.1278	-0.1062	-0.113	0.0245
gTC	–	0.4162	0.363	0.1602	0.0789	0.1697	-0.1097	-0.019
gTN	–	–	0.6381	0.146	0.0124	0.1534	-0.1324	-0.0898
gCEC	–	–	–	0.1416	0.0421	0.1297	-0.034	-0.0048
gK	–	–	–	–	0.2547	0.0081	-0.1186	0.0888
gpH	–	–	–	–	–	0.3655	-0.0853	-0.0394
gSand	–	–	–	–	–	–	0.1163	-0.0088
gSilt	–	–	–	–	–	–	–	0.057
Spherical model - Range = 343.96 m								
gP	0.5064	0.1251	0.1872	-0.0271	0.2293	0.0545	0.4009	-0.4407
gTC	–	0.8258	0.4643	0.5272	0.0425	0.5825	0.0251	-0.5208
gTN	–	–	0.5583	0.4076	-0.1614	0.3715	-0.0458	-0.1676
gCEC	–	–	–	0.7918	-0.0084	0.1558	-0.3041	0.0048
gK	–	–	–	–	0.6983	-0.4314	0.2573	-0.2148
gpH	–	–	–	–	–	0.9678	-0.0254	-0.4339
gSand	–	–	–	–	–	–	0.5858	-0.5773
gSilt	–	–	–	–	–	–	–	0.9204

The LMC of spectral model (Table 11) involves three spatial structures: a nugget effect and a spherical model at shorter range (50 m); a spherical model at longer range (210 m). The coregionalization matrices are reported (sills of variograms)

Table 11. Linear model of coregionalization of the transformed Gaussian spectral variables

Nugget effect	gK spec	gMC spec	gP spec	gpH spec
gK spec[1]	0.7778	-0.0509	-0.0074	-0.0261
gMC spec	–	0.7864	-0.2707	-0.0014
gP spec	–	–	0.4997	0.0064
gpH spec	–	–	–	0.5947
Spherical model - Range = 50.00 m				
gK spec	0.1212	0.0117	0.0475	-0.1084
gMC spec	–	0.2017	-0.0079	0.0521
gP spec	–	–	0.1509	-0.0238
gpH spec	–	–	–	0.3194
Spherical model - Range = 210.00 m				
gK spec	0.0654	0.0246	-0.059	-0.0273
gMC spec	–	0.0127	-0.0296	-0.0008
gP spec	–	–	0.3274	0.1388
gpH spec	–	–	–	0.1066

[1] spec means spectral variable

The LMC of fusion model (Table 12) involves two spatial structures: a nugget effect; an isotropic spherical model with range of 181.70 m. The coregionalization matrices are reported (sills of variograms)

Table 12. Linear model of coregionalization of the transformed Gaussian fusion variables (soil and spectral variables). The coregionalization matrices are reported (sills of variograms)

Nugget effect	gK spec	gMC spec	gP spec	gpH spec	gP	gTC	gTN	gCEC	gK	gpH	gSand	gSilt
gK spec[1]	0.0389	0.0106	-0.0121	-0.0266	-0.0564	-0.0163	0.0121	-0.0038	-0.0437	0.0359	0.0243	0.0209
gMC spec	–	0.0758	0.0418	0.0369	-0.0009	0.0238	0.0358	0.0307	0.0445	0.0257	-0.011	0.0039
gP spec	–	–	0.1024	0.0463	0.0111	0.0911	0.0629	0.0817	0.0583	0.0638	-0.0092	-0.027
gpH spec	–	–	–	0.1134	-0.0394	0.058	0.0173	0.0141	0.0263	0.0602	-0.0316	-0.0478
gP	–	–	–	–	0.2296	0.0478	0.0926	0.0096	0.1144	-0.1602	-0.0323	-0.0195
gTC	–	–	–	–	–	0.2877	0.221	0.0651	0.0573	0.0268	-0.0138	-0.0525
gTN	–	–	–	–	–	–	0.4729	0.0513	0.0475	0.0063	0.0022	-0.1093
gCEC	–	–	–	–	–	–	–	0.107	0.0723	0.0209	0.0106	0.003
gK	–	–	–	–	–	–	–	–	0.1993	-0.0511	-0.065	0.0051
gpH	–	–	–	–	–	–	–	–	–	0.1979	-0.0005	-0.0243
gSand	–	–	–	–	–	–	–	–	–	–	0.0458	0.0125
gSilt	–	–	–	–	–	–	–	–	–	–	–	0.0674
Spherical model - Range = 181.70m												
gK spec	0.1145	0.0335	-0.0139	-0.0662	0.09	0.1008	0.0055	0.0267	0.1204	-0.0312	0.0948	-0.1108
gMC spec	–	0.1133	-0.0922	0.0325	0.0182	0.1843	0.1695	0.0789	-0.0802	0.1861	-0.0184	-0.0871
gP spec	–	–	0.2973	0.0298	-0.1138	-0.3308	-0.2043	-0.1013	-0.0259	-0.2746	0.0107	0.164
gpH spec	–	–	–	0.1913	0.0129	0.0886	0.1739	0.1264	-0.1461	0.2551	-0.1055	-0.0081
gP	–	–	–	–	0.4941	0.1936	0.2559	0.0054	0.1546	0.098	0.0892	-0.1663
gTC	–	–	–	–	–	0.7226	0.5025	0.4624	0.0556	0.5897	-0.1425	-0.2624
gTN	–	–	–	–	–	–	0.6306	0.3778	-0.165	0.4778	-0.2619	-0.0498
gCEC	–	–	–	–	–	–	–	0.5393	-0.047	0.2944	-0.2371	-0.0381
gK	–	–	–	–	–	–	–	–	0.4791	-0.1639	0.0383	0.0356
gpH	–	–	–	–	–	–	–	–	–	0.8627	-0.163	-0.2733
gSand	–	–	–	–	–	–	–	–	–	–	0.4362	-0.3482
gSilt	–	–	–	–	–	–	–	–	–	–	–	0.5309

[1] spec means spectral variable

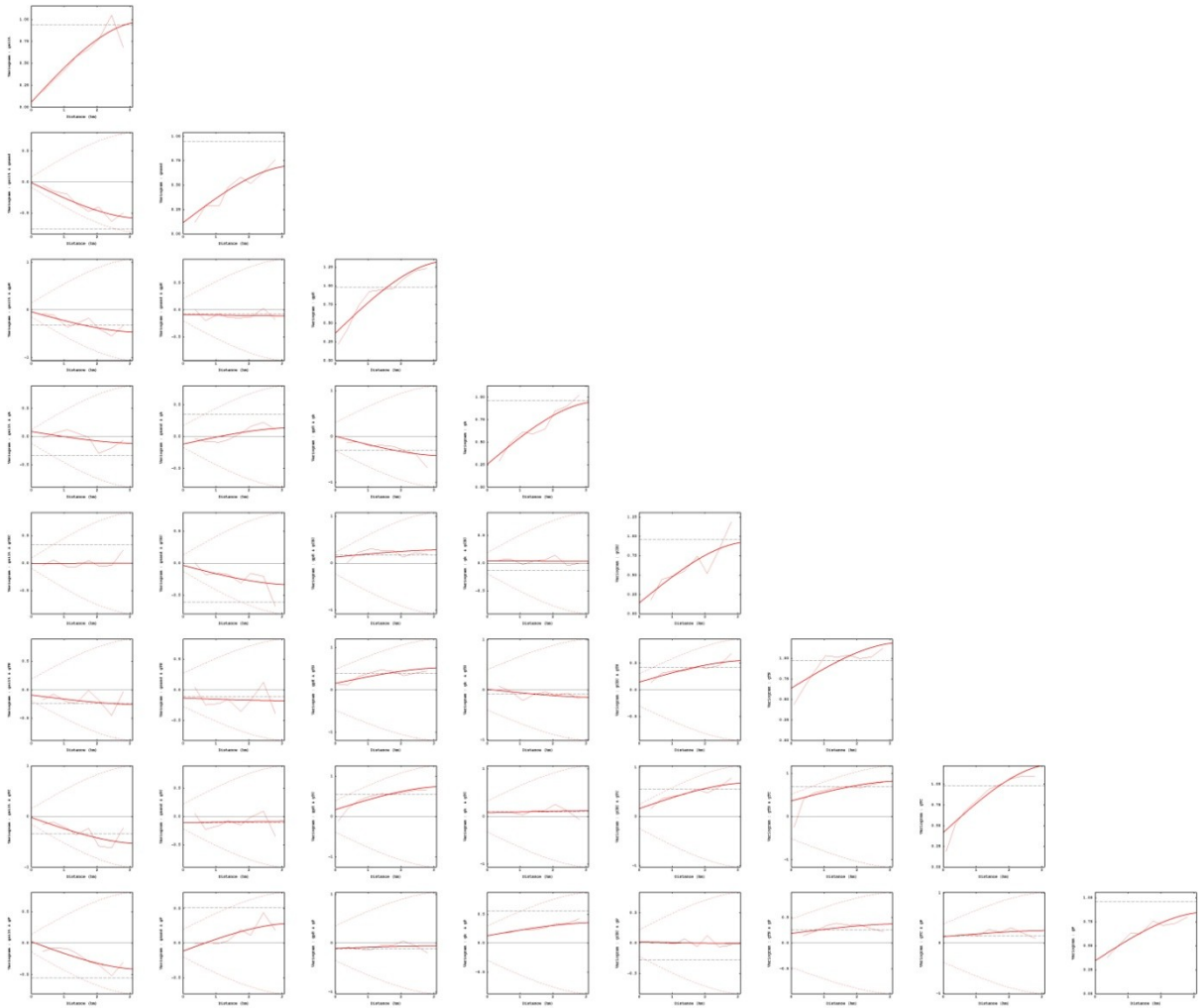

Figure 42. Matrix of the experimental (fine-line) and soil model (bold-line) variograms of the variables (P, TC, TN, CEC, K, pH, Sand and Silt), with the hull of perfect correlation (dash-line)

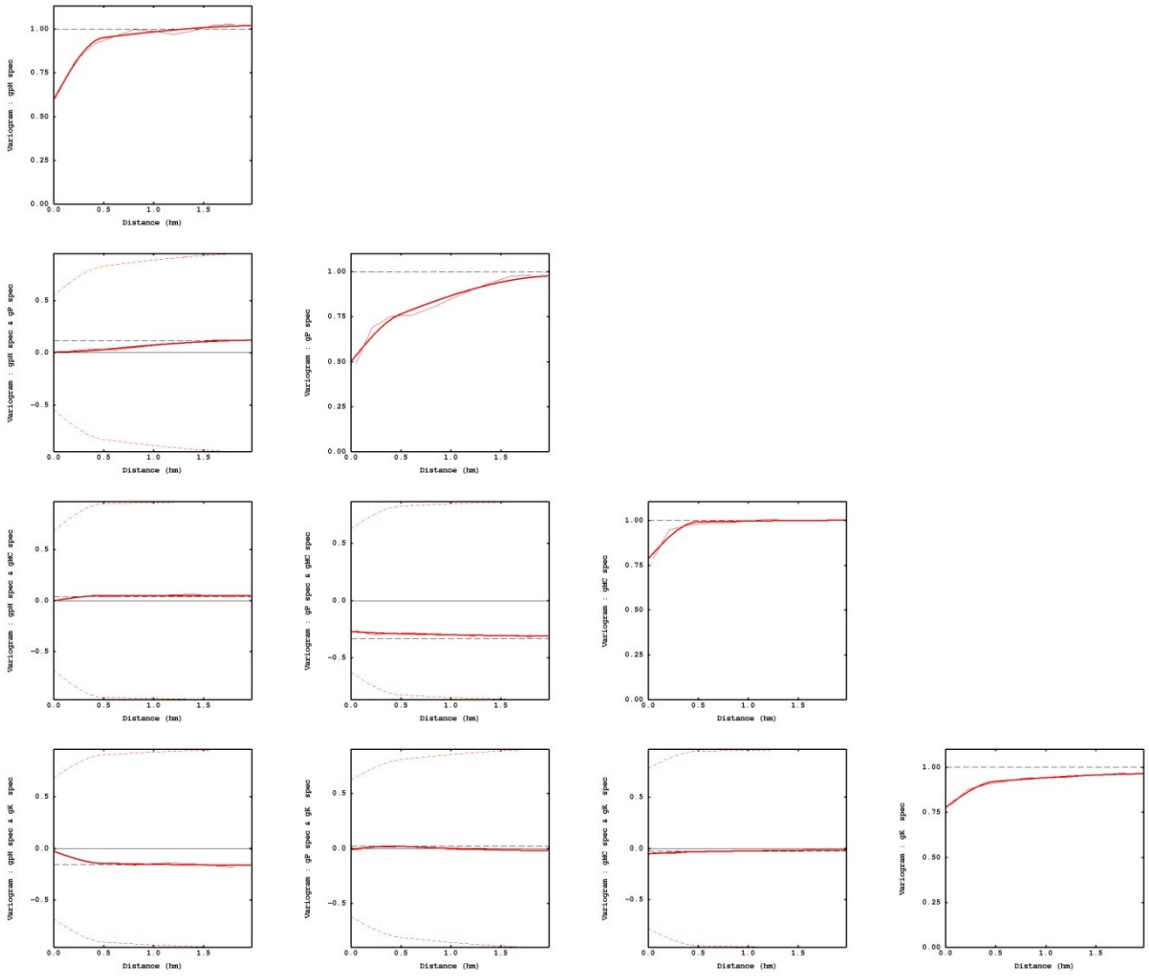

Figure 43. Matrix of the experimental (fine-line) and spectral model (bold-line) variograms of the variables (P spec, MC spec, K spec and pH spec), with the hull of perfect correlation (dash-line)

Most of variograms of soil properties look bounded (Figure 42) with the exception of TC, TN and pH, which probably show variability at a scale greater than the size of the field. TN has a nugget effect greater than pH, K, TC, and P, and those have nugget effects greater than silt, sand and CEC. The spatial correlations of P with CEC and pH are null as well as for TC with K and sand. Also, the spatial correlations of CEC with K and silt are null as well as between TN and K, and between pH and sand. Silt was negatively correlated with P, pH, TN, TC and sand whereas P was positively correlated with TN, TC and K. Positive correlations occurred between TC and TN, CEC and pH as well as between TN and CEC and pH.

The graph of the direct and cross variograms (Figure 43) for the spectral variables reveals these variables to be highly noisy and not significantly correlated between them.

All variograms of the fusion model look bounded (Figure 44) with low to moderate nugget effects except for TN which has high nugget effect. With regards to the spatial correlations among the variables, there are null spatial correlations of P with MC spec, pH spec and CEC as well as for K with TN, TC and silt. Null spatial correlations are found also between P spec and sand, and between CEC and silt.

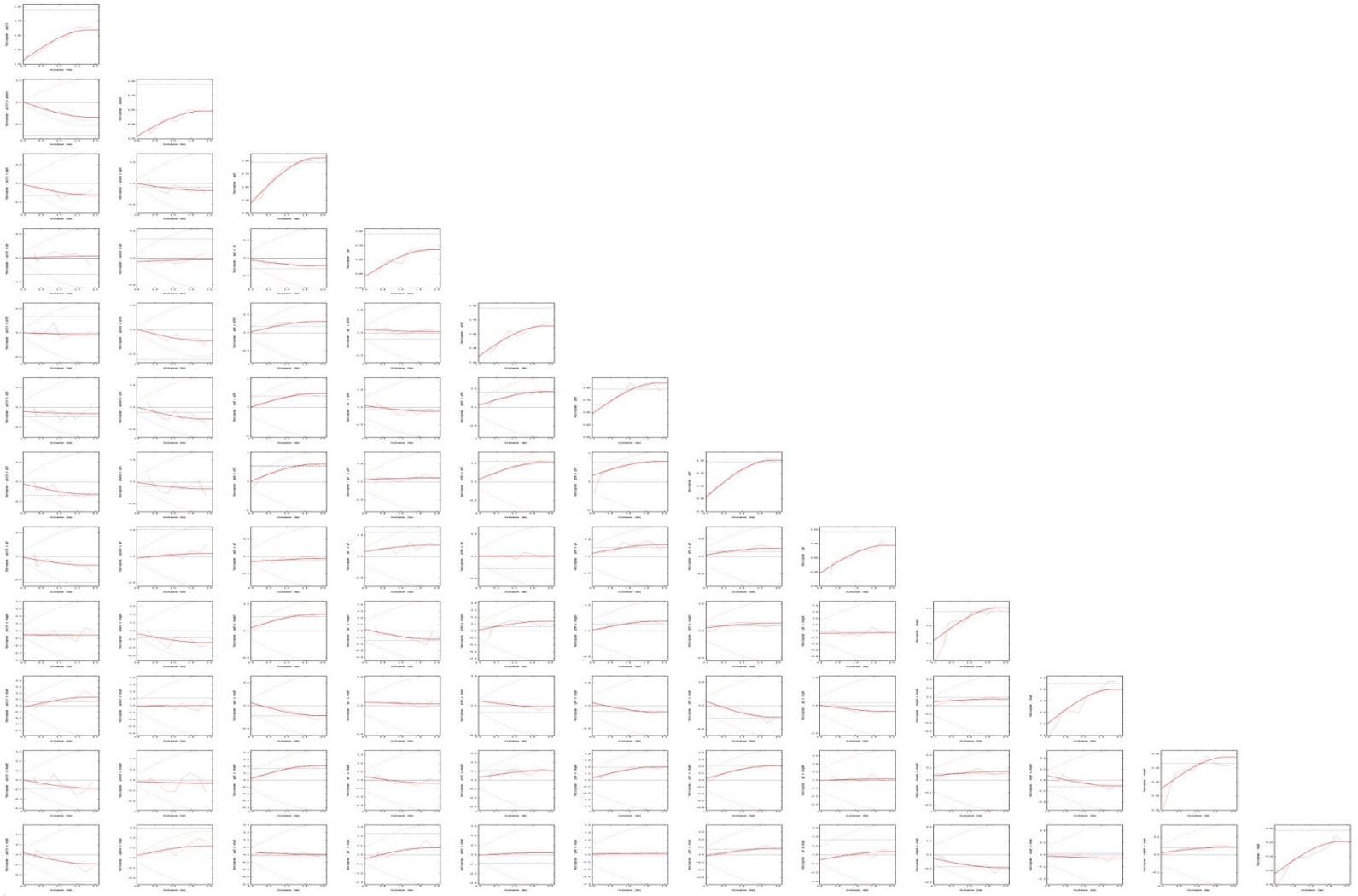

Figure 44. Matrix of the experimental (fine-line) and fusion model (bold-line) variograms of the variables (P, TC, TN, CEC, K, pH, Sand, Silt, P spec, MC spec, K spec and pH spec), with the hull of perfect correlation (dash-line)

6.2.1 Cross-validation of geostatistical models

The appropriateness of the chosen LMC models was tested by cross-validation. This involves deleting each sample in turn and then kriging it (z*) independently from all other points in the estimation neighbourhood (Castrignanò, 2011). The outcomes of the cross-validation procedure can be shown in a graphic which involves the following plots:

a) The histogram of the standardised estimation errors, which should symmetrical, centred on 0 and without long tails (Lower left plot)
b) The scatter diagram of the true data versus the estimated values, which should be close to the first bisector (Upper left plot)
c) The scatter diagram of the standardised estimation errors versus the estimated values, which should look like a round cloud, symmetrical to horizontal axis (Lower right plot).

6.2.1.1 Soil model

The cross-validation results for CEC (Figure 45) show strong correlation between the true and estimated values with correlation coefficient (r) of 0.881. The scatter diagram of the standardised estimation errors versus the estimated values looks like a round cloud with r value of 0.145. The histogram of the standardised estimation errors is symmetrical but with some points found outside the range (-2.5 – 2.5) , these points are known as *outliers*. The mean standardized error is close to 0 (0.023) which means that the model is unbiased and the standard deviation of standardized error is close to 1 (1.17) which means that the estimation resulting from the model is accurate.

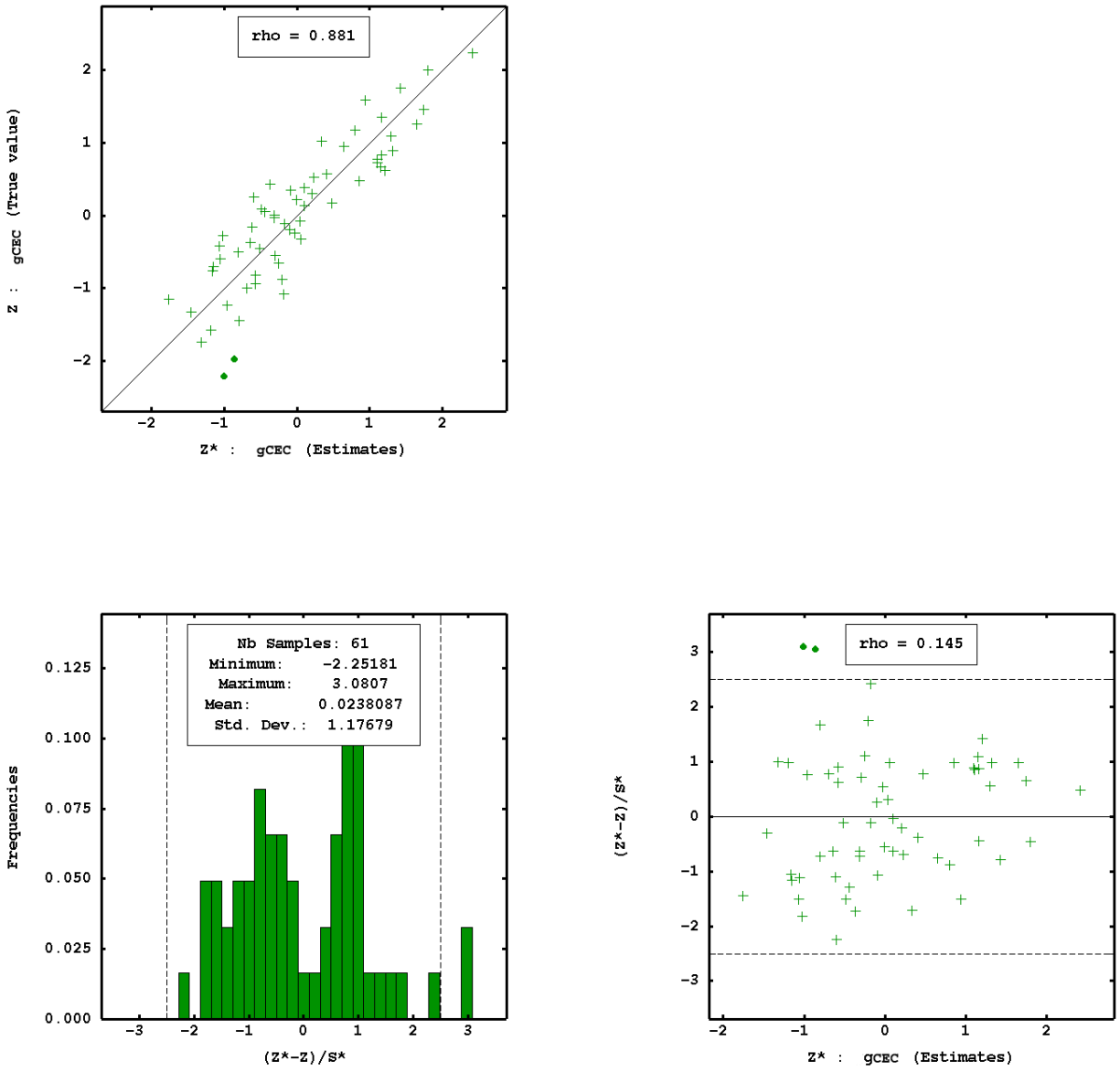

Figure 45. Outcomes of the cross-validation of CEC interpolated by soil model.

Cross-validation results of exchangeable potassium (Figure 46) shows strong correlation between true and estimated values with correlation coefficient (r) of 0.814. The scatter diagram of the standardised estimation errors versus the estimated values looks like a round cloud with r value of 0.130 and symmetrical to the horizontal axis. The histogram of the standardised estimation errors is symmetrical but with few points found outside the range (-2.5 – 2.5), these points have values smaller than –2.5. The mean standardized error is close to 0 (-0.003) which means that the model is unbiased and the standard deviation of standardized error is close to 1 (1.11) which means that the model estimation is accurate.

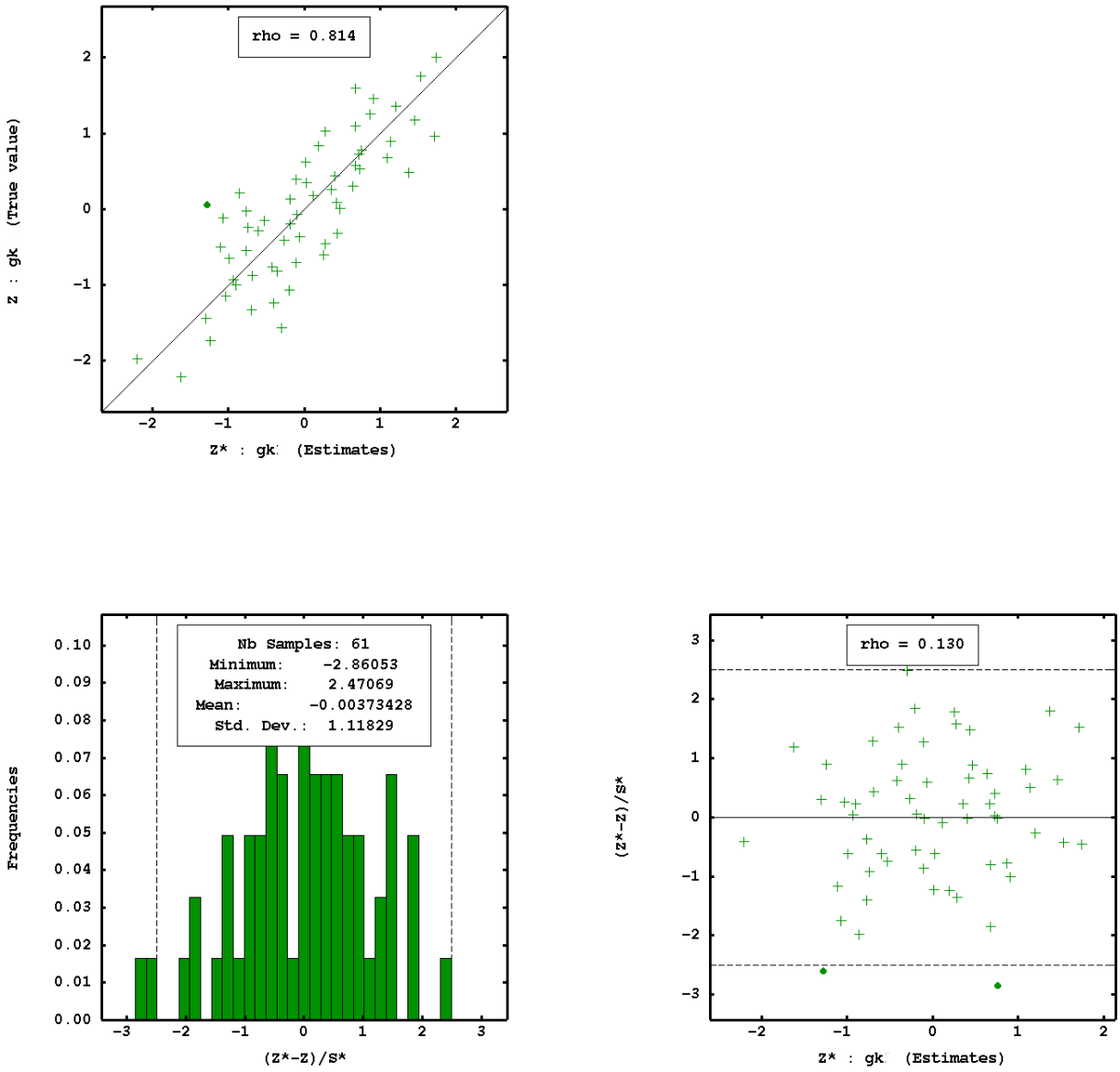

Figure 46. Outcomes of the cross-validation of K interpolated by soil model.

For extractable phosphorus, there is a strong correlation (Figure 47) between the true and estimated values of P with an r value of 0.820. The scatter diagram of the standardised estimation errors versus the estimated values looks like a round cloud with r value of 0.165 and symmetrical to the horizontal axis. The histogram of the standardised estimation errors is symmetrical but with few points found outside the range (-2.5 – 2.5) , these points have values less than –2.5 and greater than 2.5. The mean standardized error is close to 0 (-0.005) which means that the model is unbiased and the standard deviation of standardized error is close to 1 (1.14) which means that the model is good for theoretical estimates.

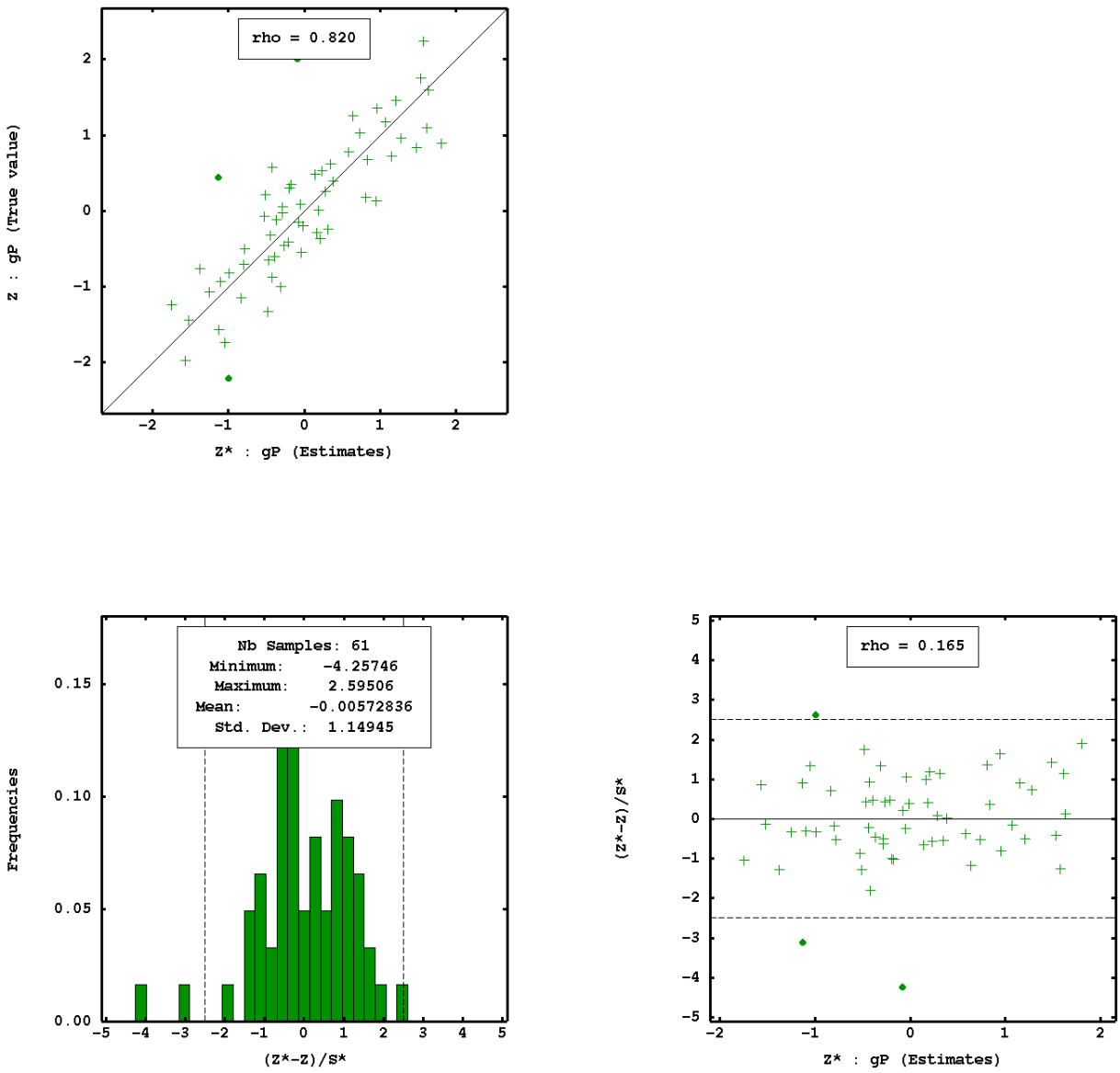

Figure 47. Outcomes of the cross-validation of P interpolated by soil model.

Figure 48 shows a strong correlation between the true and estimated values of pH with r value of 0.764. The scatter diagram of the standardised estimation errors versus the estimated values looks like a round cloud with r value of 0.070 and symmetrical to the horizontal axis. The histogram of the standardised estimation errors is symmetrical with few points found outside the range (-2.5 – 2.5), these points have values greater than 2.5. The mean standardized error is close to 0 (0.007) which means that the model is unbiased and the standard deviation of standardized error is close to 1 (1.09) which means that the model estimation is accurate.

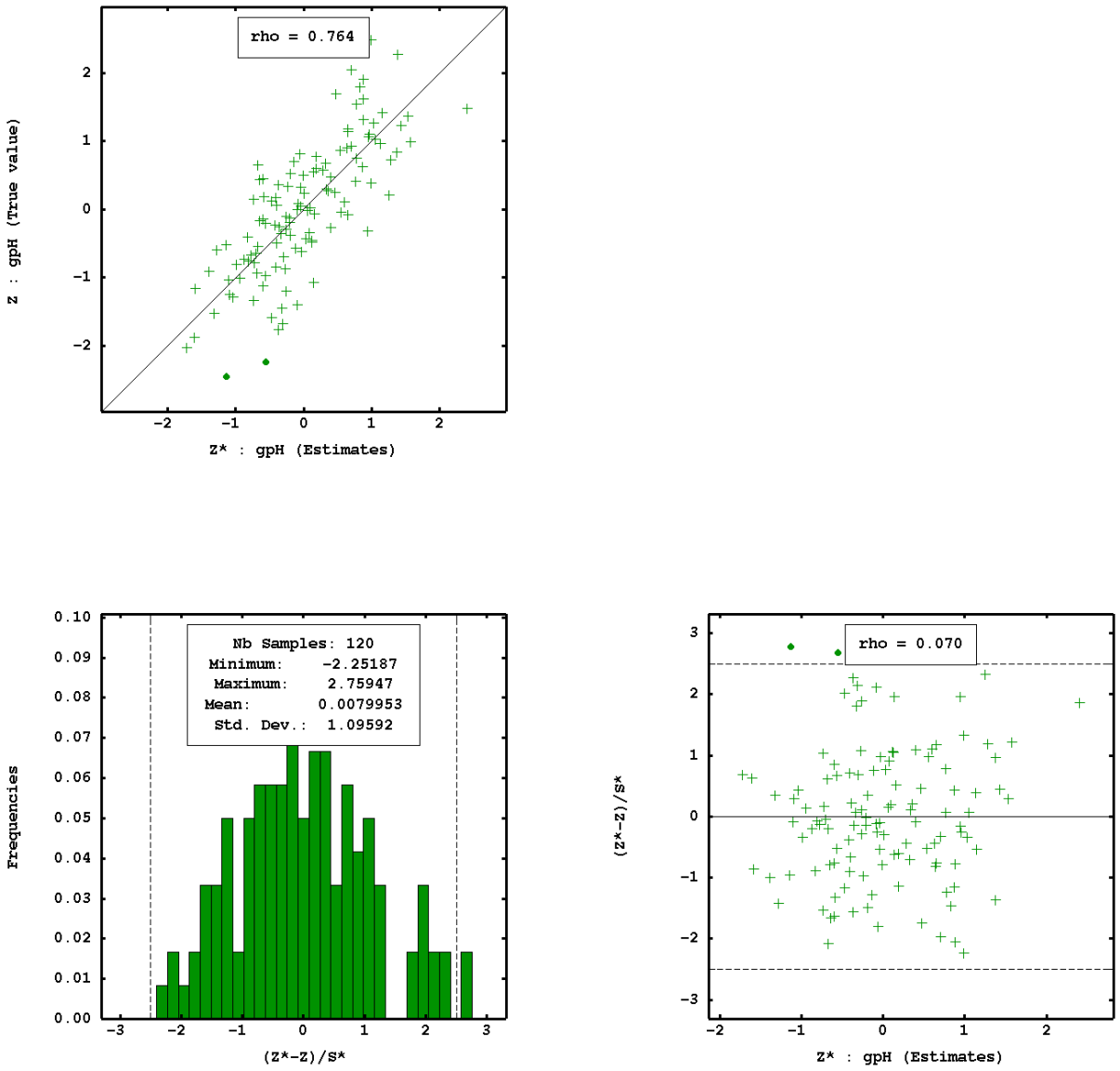

Figure 48. Outcomes of the cross-validation of pH interpolated by soil model.

Cross-validation results (Figure 49) shows a very strong correlation between true and estimated values of sand with r value of 0.929. The scatter diagram of the standardised estimation errors versus the estimated values looks like a round cloud with r value of 0.180 and symmetrical to the horizontal axis. The histogram of the standardised estimation errors is not symmetrical, as it results from the mean value of standardised error different from 0 (0.11), and with some points found outside the range (-2.5 – 2.5), these points have values less than –2.5 and greater than 2.5. The mean standardized error is somewhat close to 0 (0.10) which means that the model can be assumed unbiased and the standard deviation of standardized error is somewhat far from 1 (1.32) but within the tolerance interval ($\pm 3\sqrt{2}/N$, N is number of observations) (Chiles and Delfiner, 1999), which means that the model is accurate

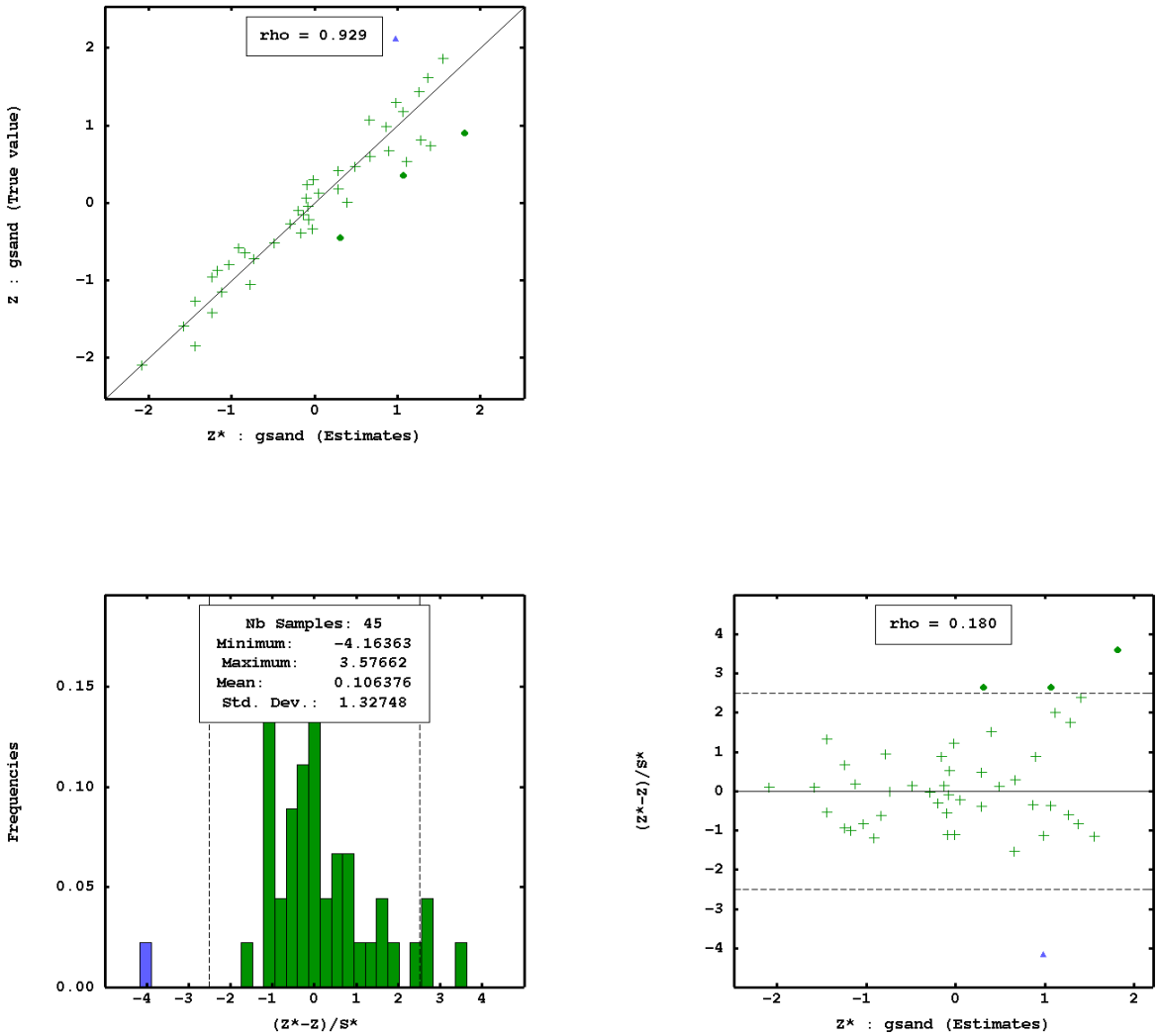

Figure 49. Outcomes of the cross-validation of sand interpolated by soil model.

A very strong correlation between the true and estimated values of silt with r value of 0.923 is shown in Figure 50. The scatter diagram of the standardised estimation errors versus the estimated values looks like a round cloud with r value of 0.141 and symmetrical to the horizontal axis. The mean value of the standardised error is a little different from 0 (=0.10) and some points fall outside the range (-2.5 – 2.5); these points have values less than –2.5 and greater than 2.5. The mean standardized error is somewhat close to 0 (0.10) which means that the model is unbiased and the standard deviation of standardized error is (1.39), within the tolerance interval, which means that the model estimation is accurate.

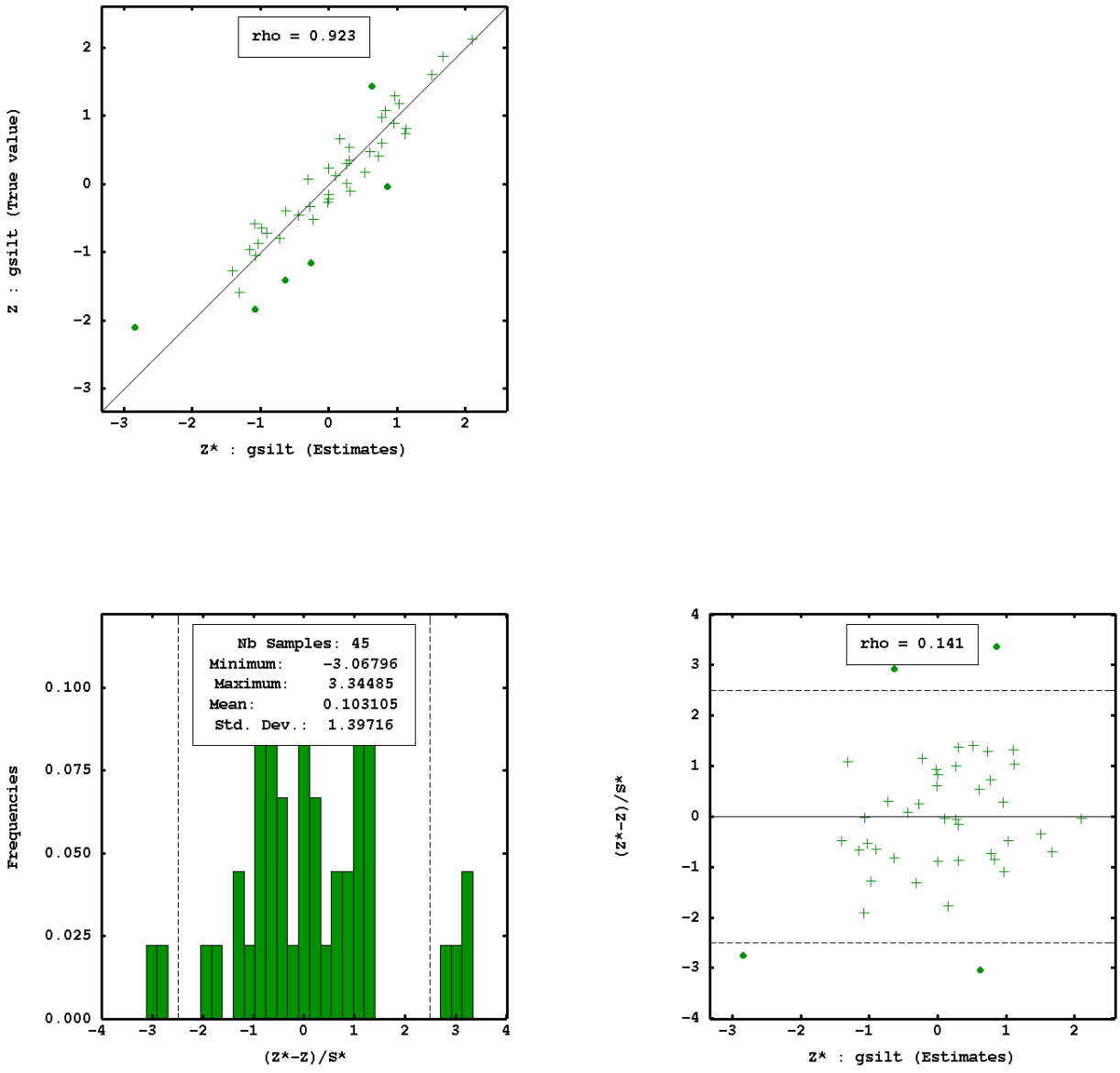

Figure 50. Outcomes of the cross-validation of silt interpolated by soil model.

Figure 51 shows a strong correlation between true and estimated values of TC with r value of 0.865. The scatter diagram of the standardised estimation errors versus the estimated values looks like a round cloud with r value of 0.048 and symmetrical to the horizontal axis. The histogram of the standardised estimation errors is symmetrical with some points found outside the range (-2.5 – 2.5) , these points have values less than –2.5 and greater than 2.5. The mean standardized error is close to 0 (0.024) which means that the model is unbiased and the standard deviation of standardized error is close to 1 (1.08) which means that the model estimation is accurate.

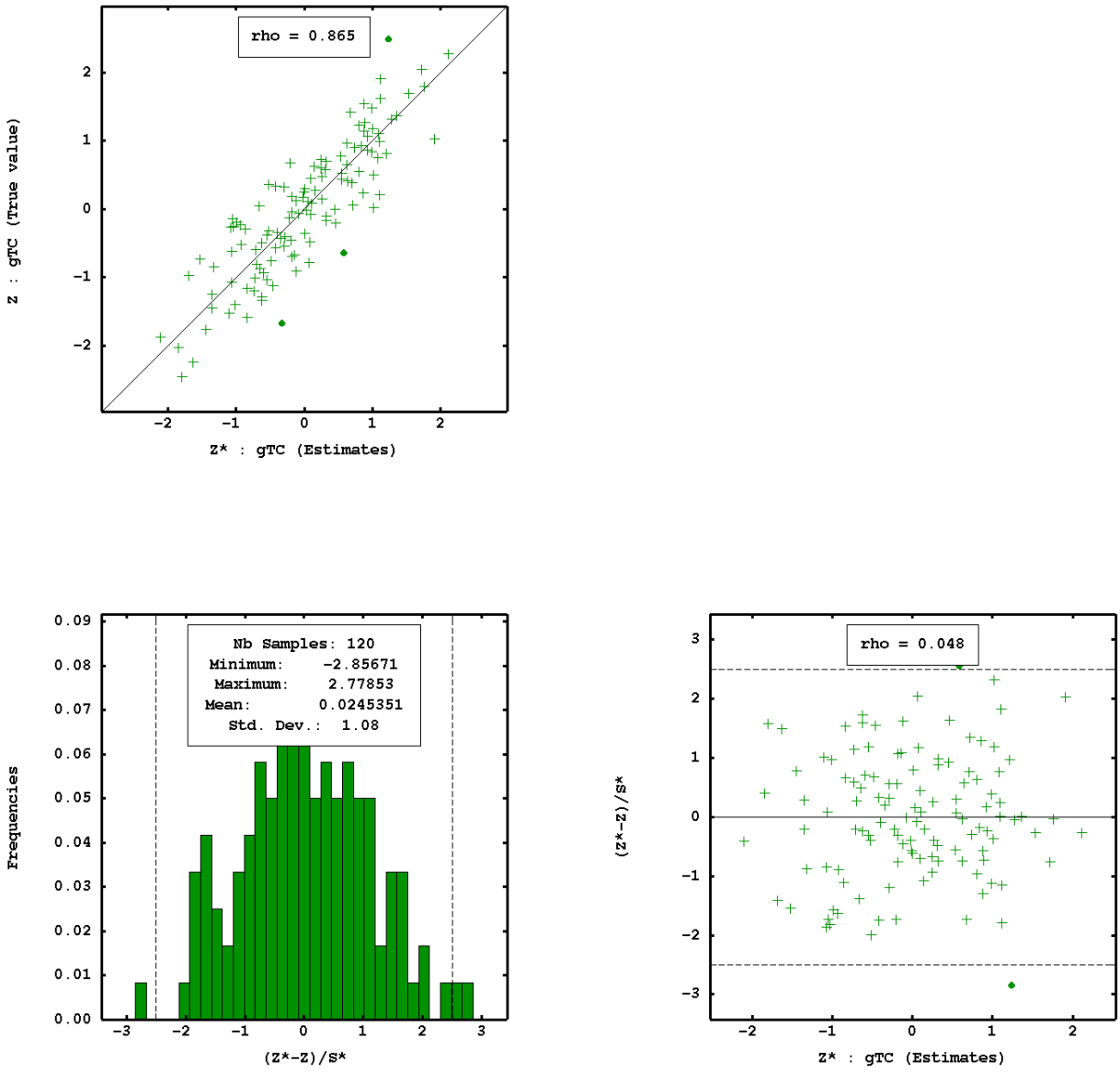

Figure 51. Outcomes of the cross-validation of TC interpolated by soil model.

A strong correlation between true and estimated values of TN is shown in Figure 52 with r value of 0.761. The scatter diagram of the standardised estimation errors versus the estimated values looks like a round cloud with r value of 0.041 and symmetrical to the horizontal axis. The histogram of the standardised estimation errors is symmetrical with some points found outside the range (-2.5 – 2.5) , these points have values less than –2.5 and greater than 2.5. The mean standardized error is close to 0 (-0.019) which means that the model is unbiased and the standard deviation of standardized error is close to 1 (1.12) which means that the model estimation is accurate.

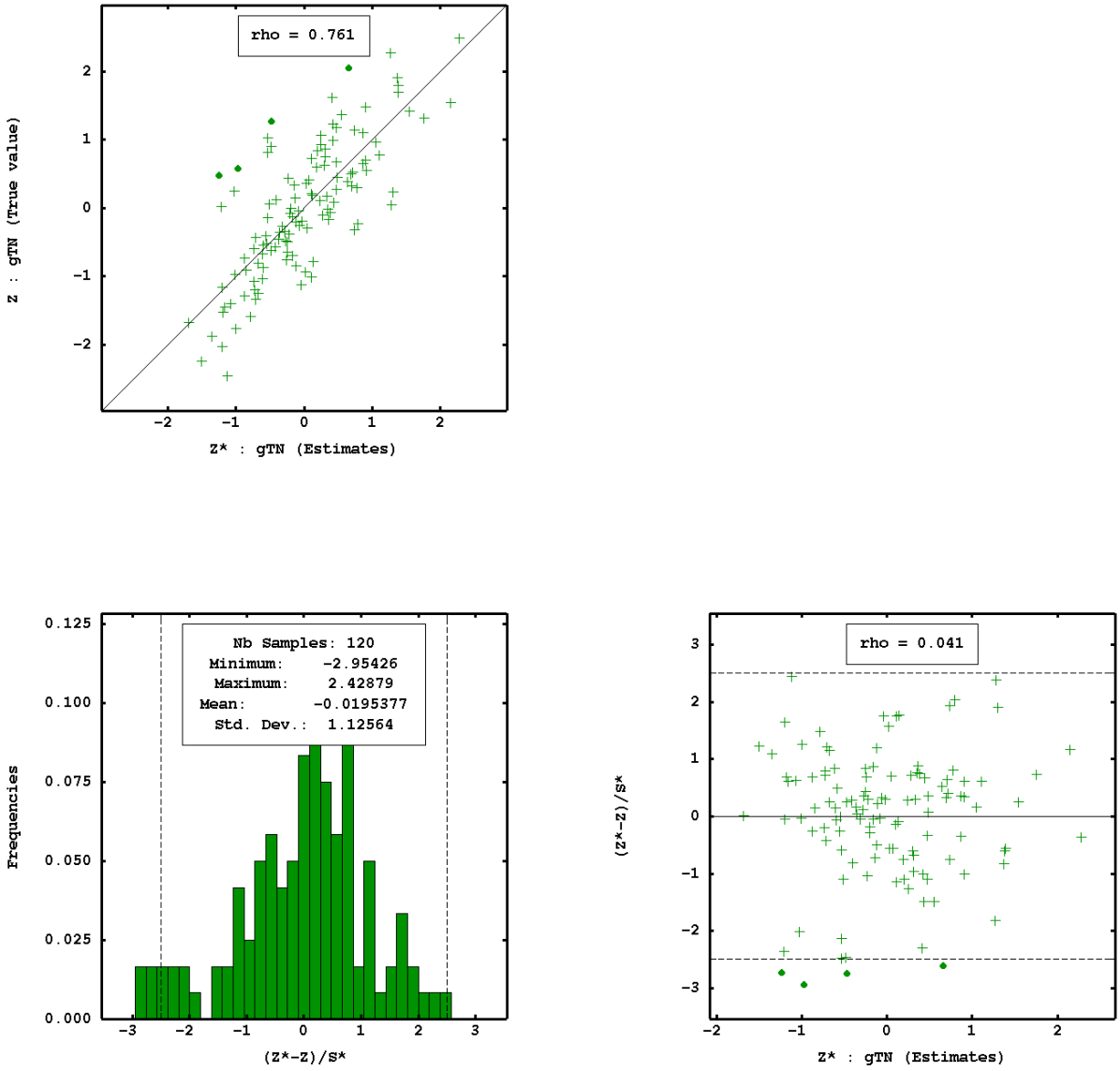

Figure 52. Outcomes of the cross-validation of TN interpolated by soil model.

Summing up the results of cross-validation, the goodness of fitting of the soil model can be considered as satisfactory.

6.2.1.2 Spectral model

A weak correlation between the true and estimated values of exchangeable potassium (Figure 53) using the spectral model is found with a low r value of 0.448, due to the strong smoothing of the model. The histogram of the standardised estimation errors is symmetrical. The scatter diagram of the standardised estimation errors versus the estimated values looks like a round cloud with r value of -0.024 and symmetrical to the horizontal axe. The mean standardized error is quite close to 0 (-0.002) which means that the model is unbiased and the standard deviation of standardized error is close to 1 (0.99) which means that the model estimation is accurate.

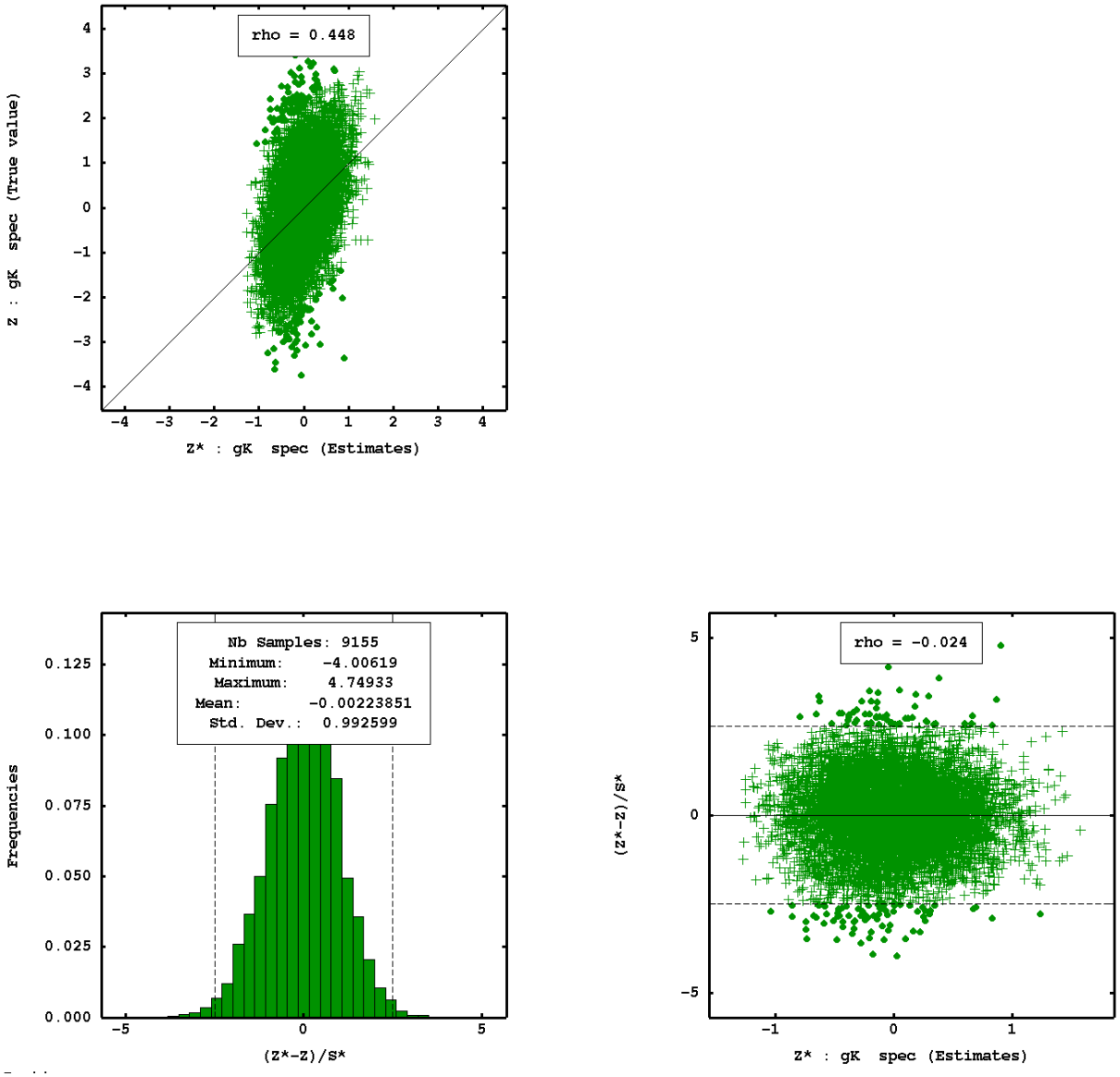

Figure 53. Outcomes of the cross-validation of exchangeable potassium interpolated by spectral model.

Figure 54 shows intermediate slight higher correlation between true and estimated values of soil moisture content with r value of 0.585. The histogram of the standardised estimation errors is symmetrical with some points outside the range. The scatter diagram of the standardised estimation errors versus the estimated values looks like a round cloud with r value of -0.038 and symmetrical to the horizontal axis. The mean standardized error is close to 0 (0.0007) which means that the model is unbiased and the standard deviation of standardized error is close to 1 (0.97) which means that the model estimation is accurate.

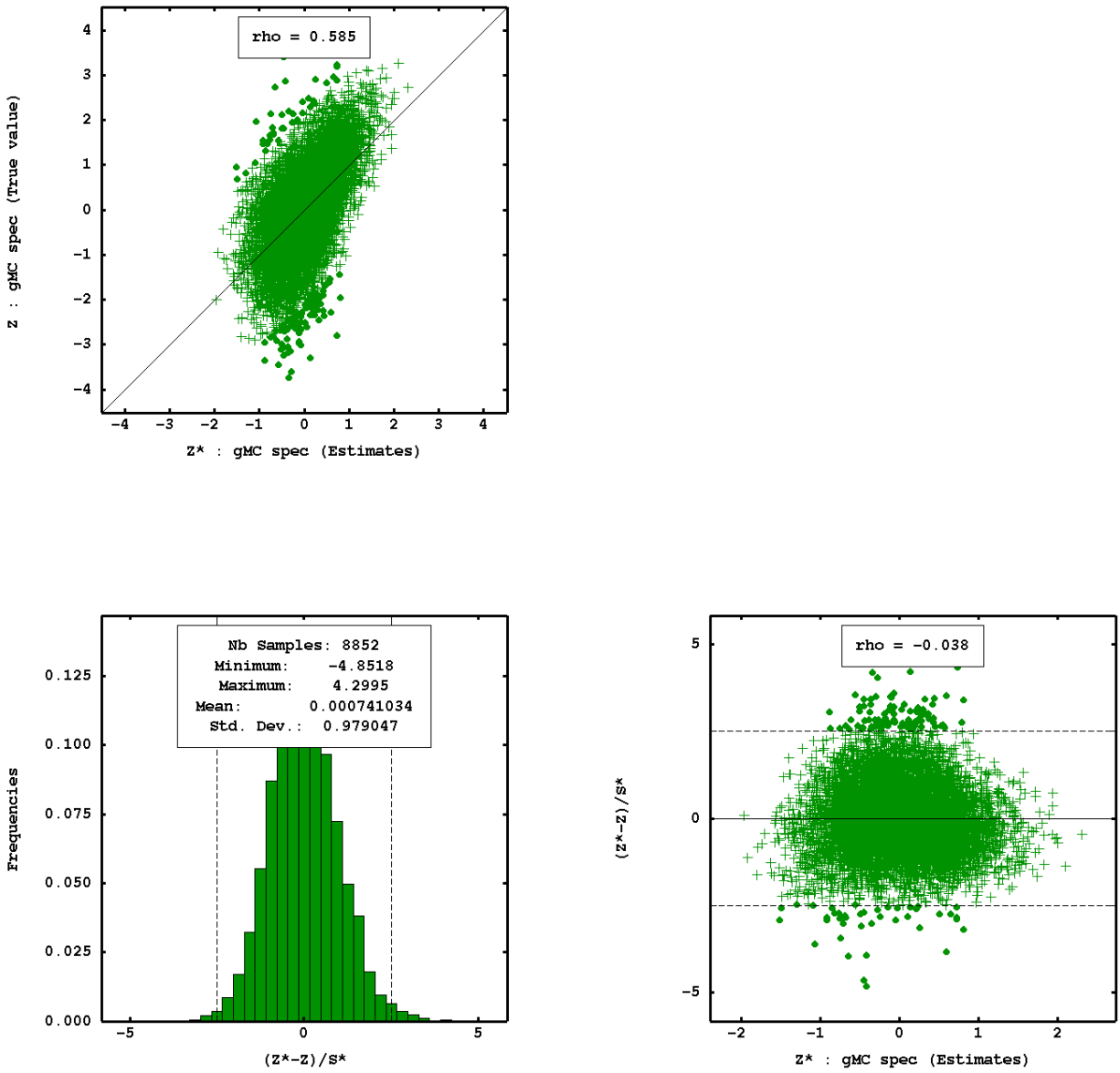

Figure 54. Outcomes of the cross-validation of moisture content interpolated by spectral model.

Unlike with the previous two soil properties Figure 55 shows a strong correlation between true and estimated values of extractable phosphorus with r value of 0.750. The histogram of the standardised estimation errors is symmetrical with some points outside the range. The scatter diagram of the standardised estimation errors versus the estimated values looks like a round cloud with r value of -0.027 and symmetrical to the horizontal axe. The mean standardized error is close to 0 (-0.004) which means that the model is unbiased and the standard deviation of standardized error is close to 1 (0.98) which means that the model estimation is accurate.

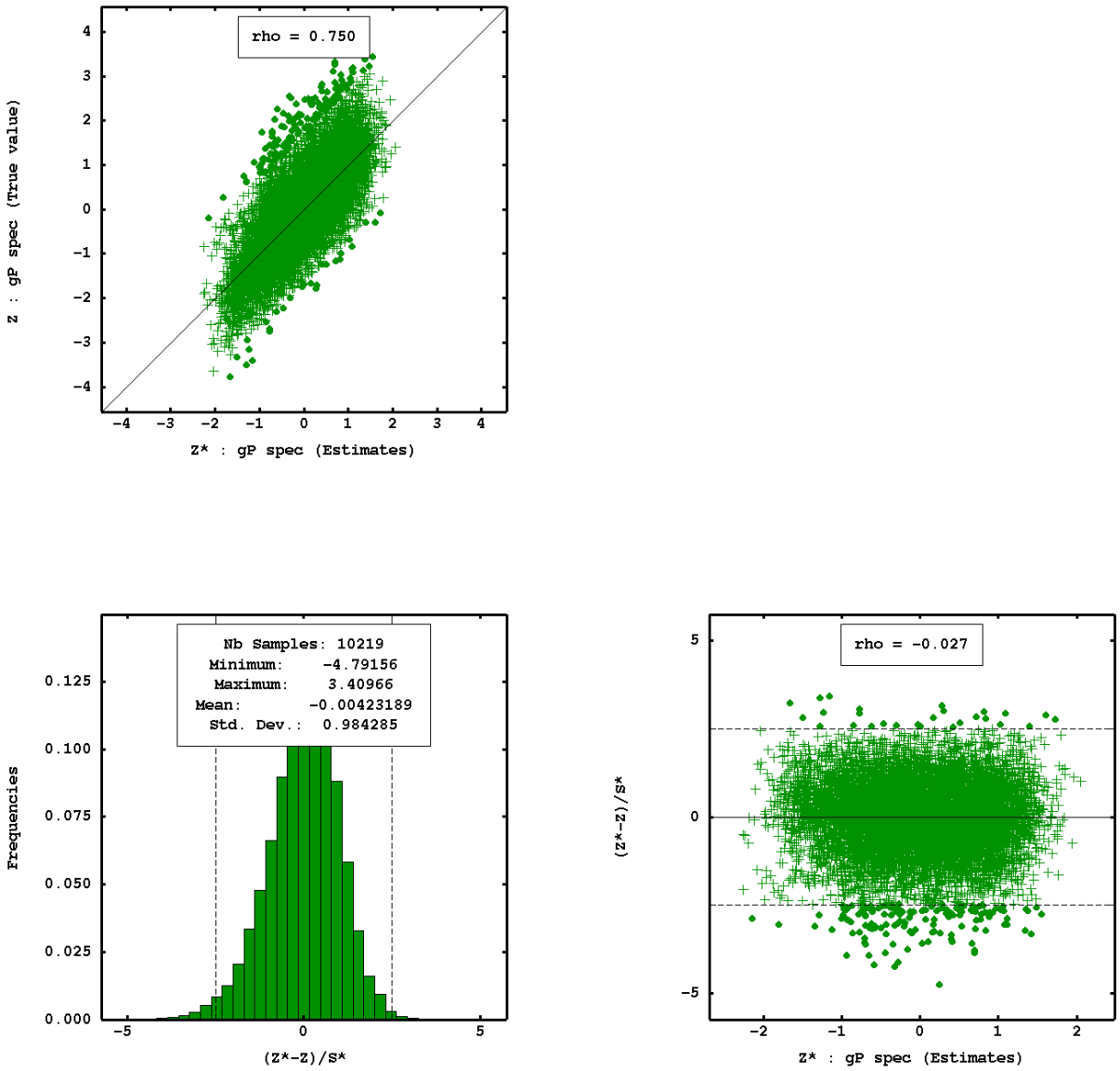

Figure 55. Outcomes of the cross-validation of extractable phosphorus interpolated by spectral model.

The correlation coefficient between true and estimated values of soil pH (0.582) shows some smoothing of the model (Figure 56). The histogram of the standardised estimation errors is symmetrical with some points outside the range. The scatter diagram of the standardised estimation errors versus the estimated values looks like a round cloud with r value of -0.004 and symmetrical to the horizontal axe. The mean standardized error is close to 0 (-0.006) which means that the model is unbiased and the standard deviation of standardized error is close to 1 (1.01) which means that the model estimation is accurate.

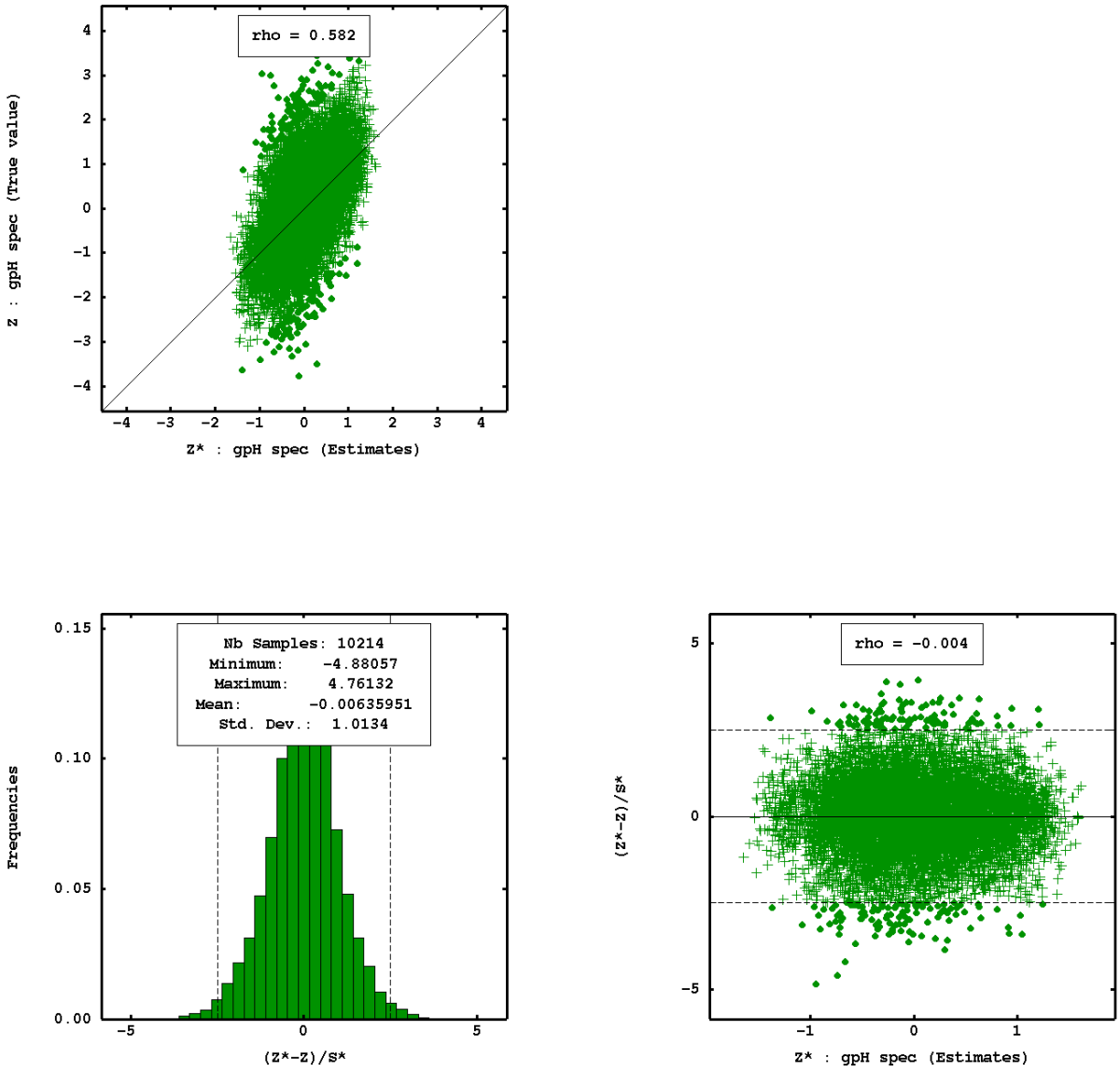

Figure 56. Outcomes of the cross-validation of soil pH interpolated by spectral model.

Summing up the results, the spectral model was unbiased and accurate but the range of variability of the estimates was much shorter than the one of observations, revealing a strong effect of smoothing.

6.2.1.3 Fusion model

A stronger correlation between true and estimated values of CEC using the fusion model (Figure 57) is found with r value of 0.911 as compared with r value of CEC (0.881) estimated by soil model. The scatter diagram of the standardised estimation errors versus the estimated values looks like a round cloud with r value of 0.042 and symmetrical to the horizontal axis. The histogram of the standardised estimation errors is symmetrical with mean error quite close to 0 and some points outside the range (-2.5 – 2.5) , these points have values greater than 2.5. The mean standardized error is close to 0 (0.006) which means that the model is unbiased and the standard deviation of standardized error is close to 1 (0.98) which means that the model estimation is accurate.

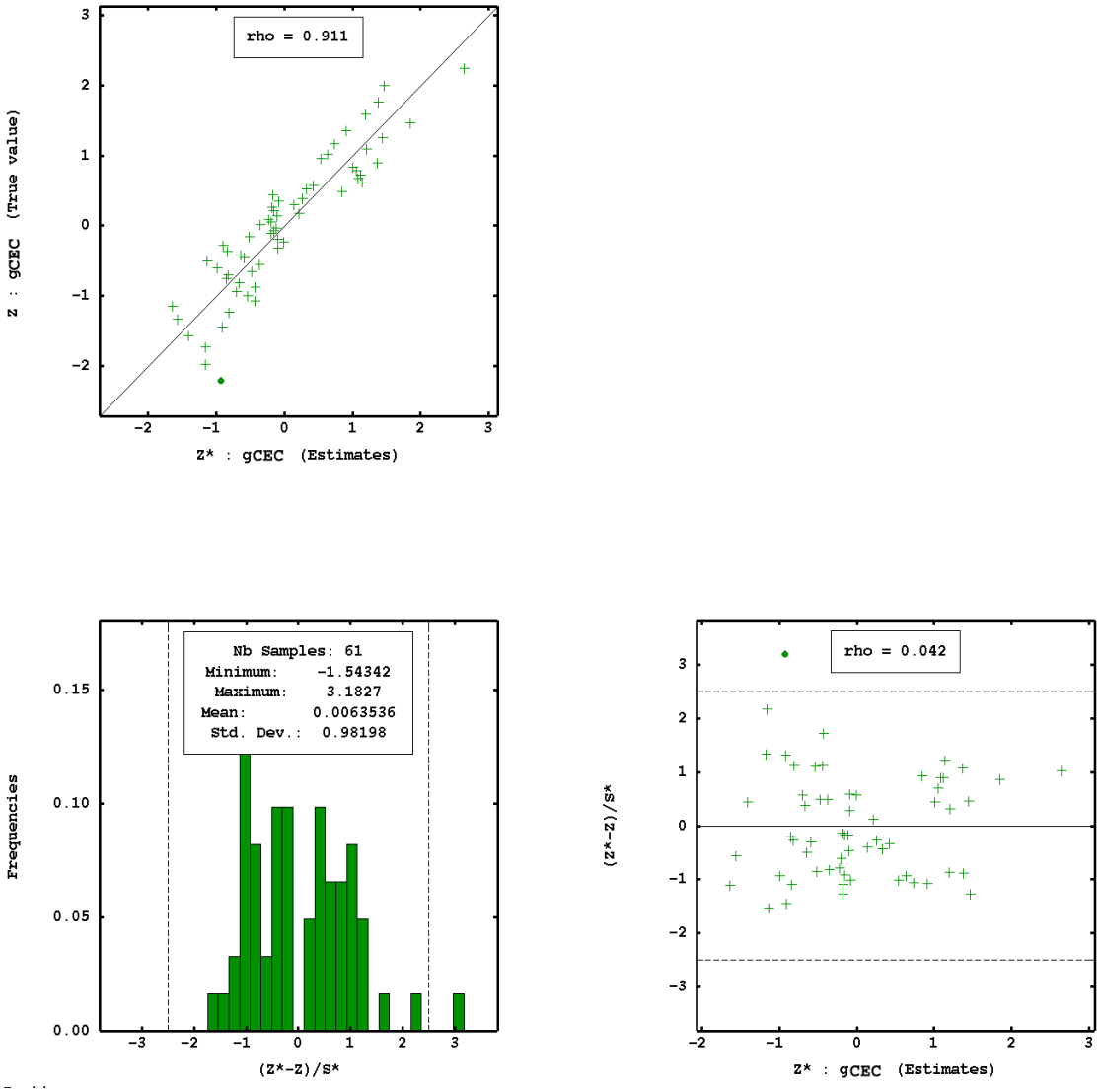

Figure 57. Outcomes of the cross-validation of CEC interpolated by fusion model.

A weaker correlation between true and estimated values of K using the fusion model (Figure 58) is found with r value of 0.793 as compared with r value of K (0.814) estimated by soil model. The scatter diagram of the standardised estimation errors versus the estimated values looks like a round cloud with r value of 0.271 and symmetrical to the horizontal axis. The histogram of the standardised estimation errors is symmetrical with some points found outside the range (-2.5 – 2.5) , these points have values less than –2.5 and greater than 2.5. The mean standardized error is close to 0 (-0.025) which means that the model is unbiased and the standard deviation of standardized error is somewhat close to 1 (1.40), however within the tolerance interval, therefore the model estimation can be assumed as accurate.

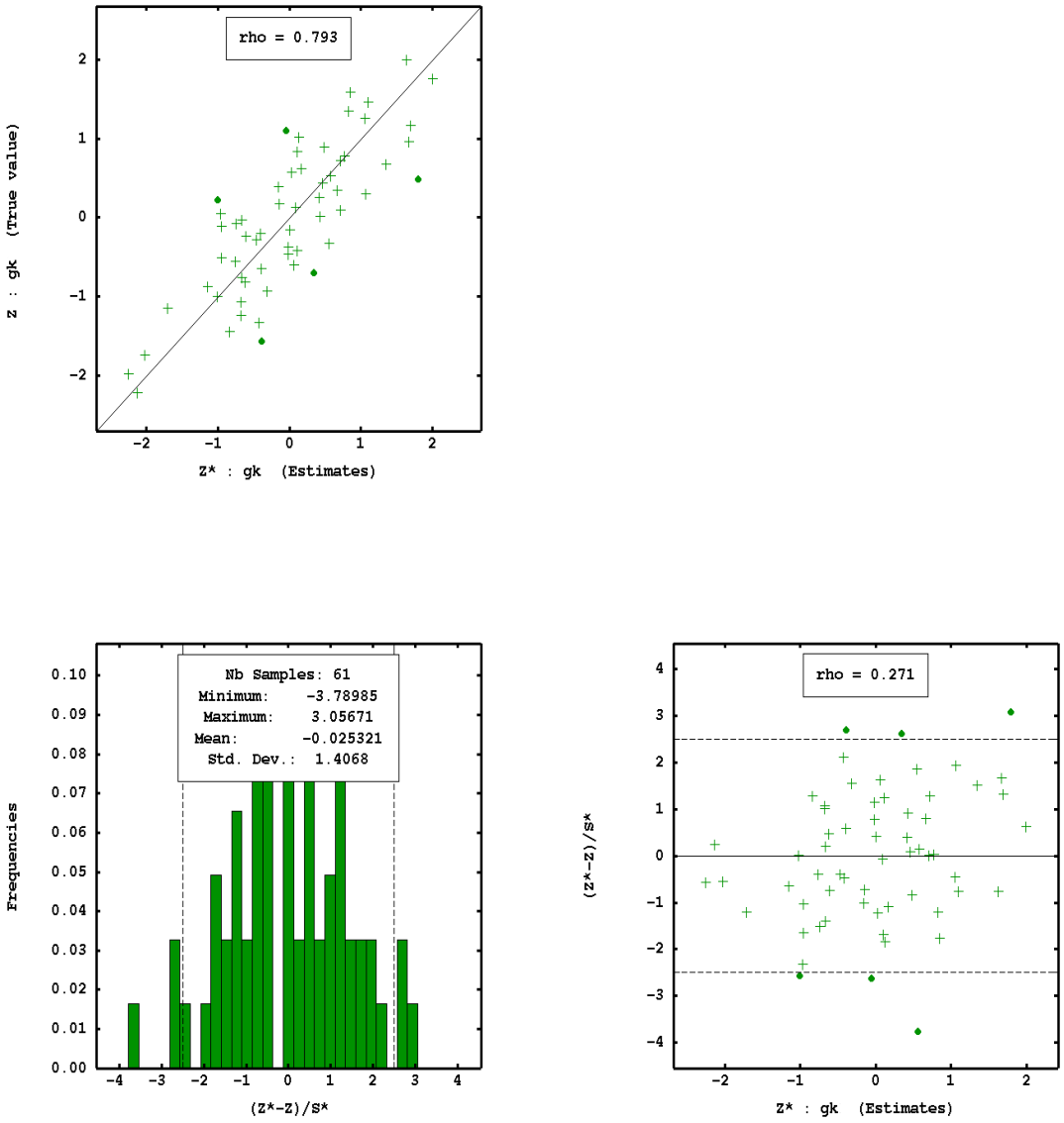

Figure 58. Outcomes of the cross-validation of exchangeable potassium interpolated by fusion model.

A stronger correlation between true and estimated values of extractable phosphorus using the fusion model (Figure 59) is found with r value of 0.834 as compared with r value of P (0.820) estimated by soil model. The scatter diagram of the standardised estimation errors versus the estimated values looks like a round cloud with r value of 0.165 and symmetrical to the horizontal axis. The histogram of the standardised estimation errors is symmetrical but with two large outliers less than -2.5. The mean standardized error is close to 0 (-0.037) which means that the model is unbiased and the standard deviation of standardized error is close to 1 (1.24) which means that the model estimation is accurate.

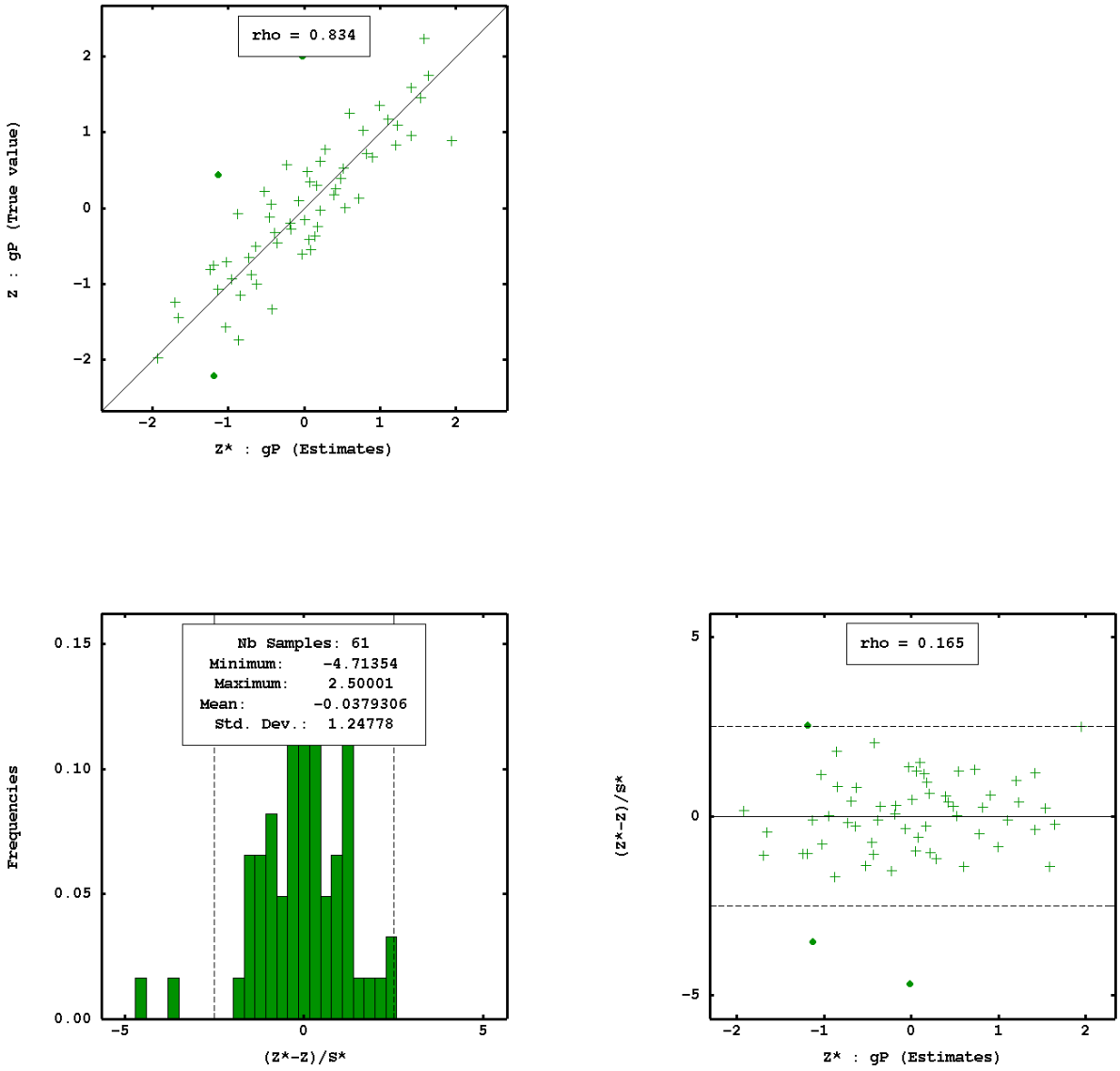

Figure 59. Outcomes of the cross-validation of extractable phosphorus interpolated by fusion model.

A stronger correlation between true and estimated values of soil pH using the fusion model (Figure 60) is found with r value of 0.837 as compared with r value of pH (0.764) estimated by soil model. The scatter diagram of the standardised estimation errors versus the estimated values looks like a round cloud with r value of 0.087 and symmetrical to the horizontal axis. The histogram of the standardised estimation errors is symmetrical with three outliers less than -2.5. The mean standardized error is close to 0 (-0.021) which means that the model is unbiased and the standard deviation of standardized error is close to 1 (1.17) which means that the model estimation is accurate.

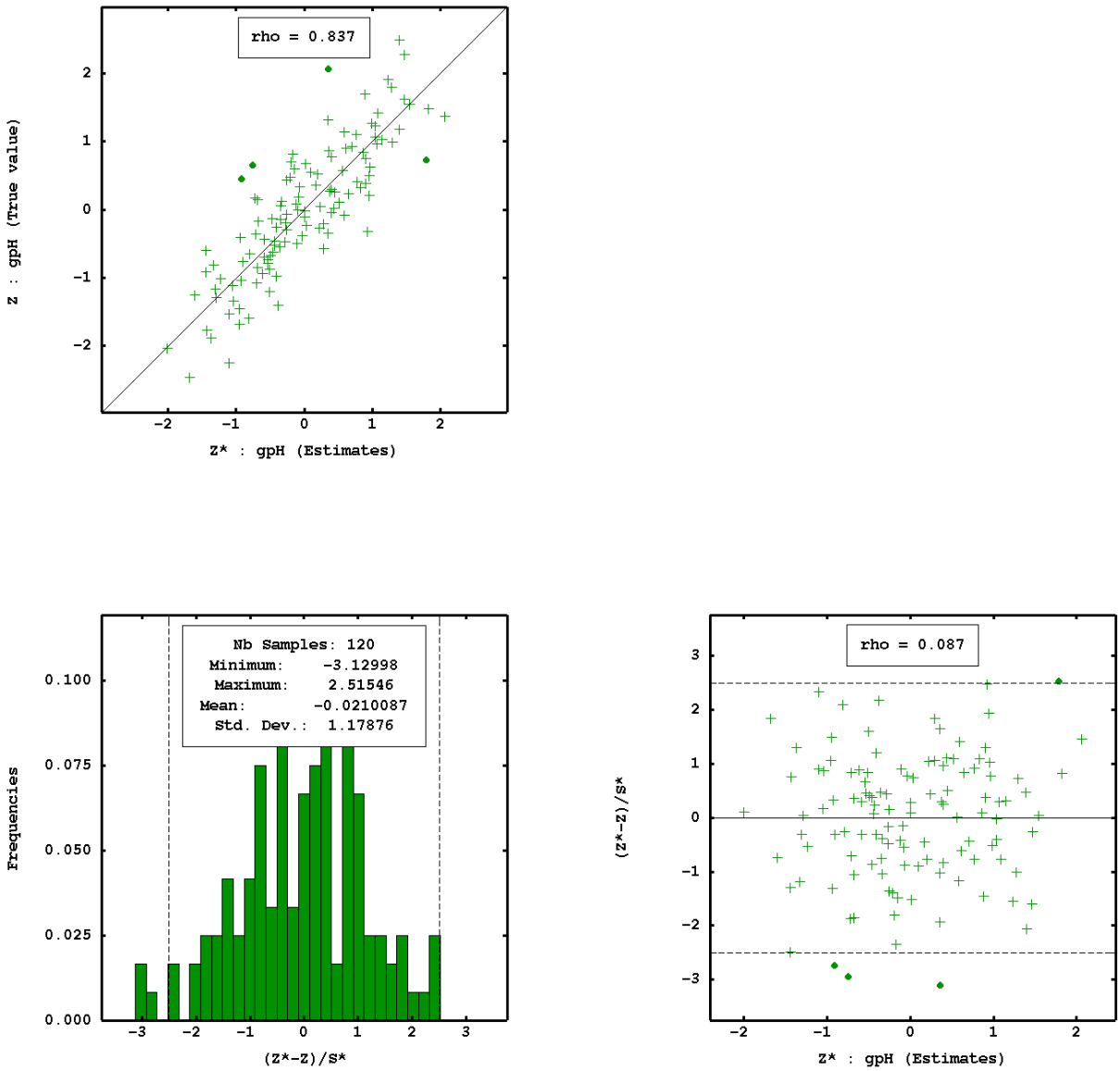

Figure 60. Outcomes of the cross-validation of soil pH interpolated by fusion model.

A very high correlation between the true and estimated values of sand was found using the fusion model (Figure 61) with r value of 0.918. The scatter diagram of the standardised estimation errors versus the estimated values looks like a round cloud with r value of 0.291 and symmetrical to the horizontal axis. The histogram of the standardised estimation errors is symmetrical however there are several large outliers, which causes the standard deviation of standardised errors to differ significantly from 1 (1.90) (outside the tolerance interval). The mean standardized error is close to 0 (0.024) which means that the model is unbiased, but the sample variance was sensibly underestimated by the model.

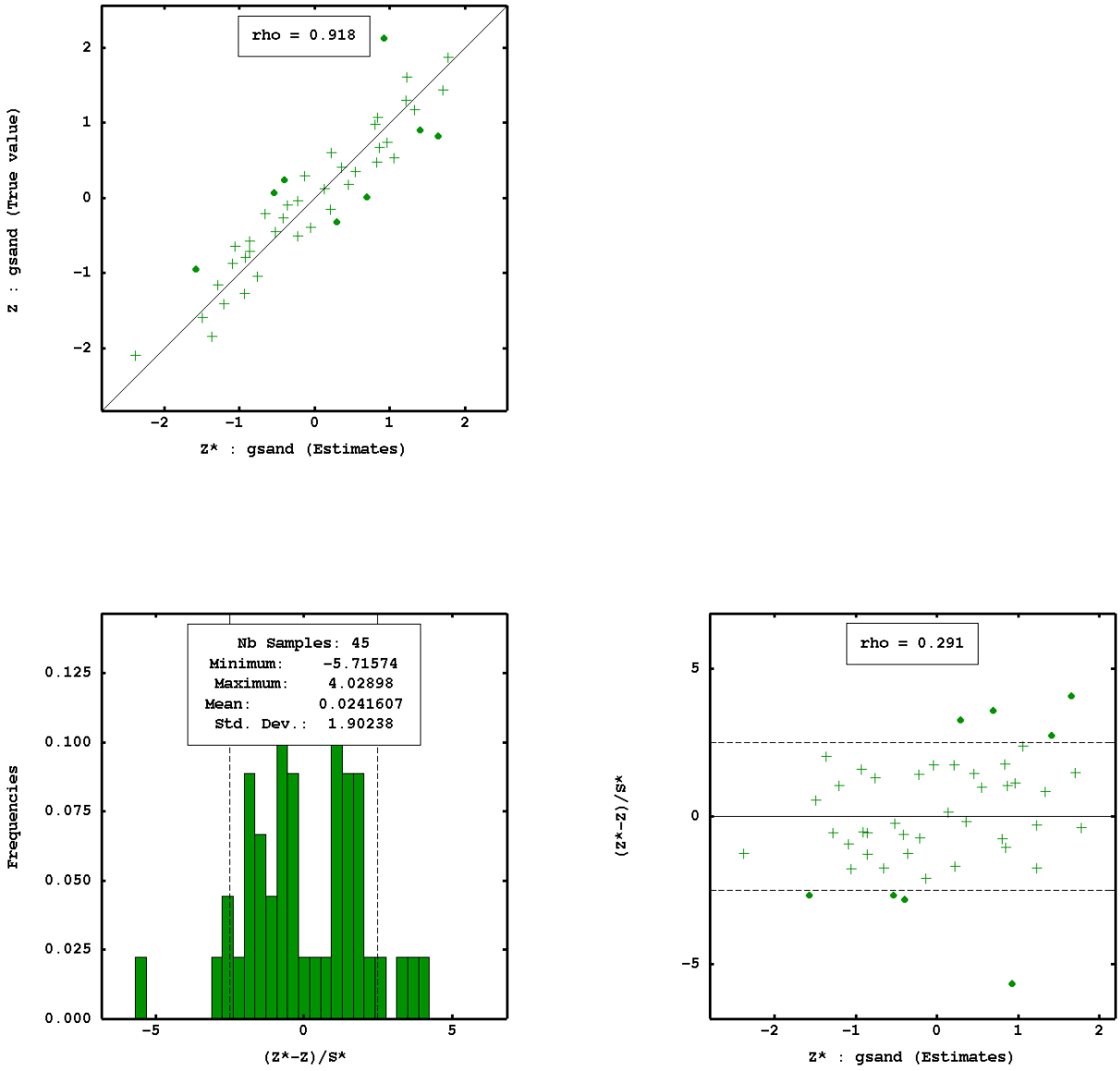

Figure 61. Outcomes of the cross-validation of sand content interpolated by fusion model.

A slightly weaker correlation between true and estimated values of silt using the fusion model (Figure 62) is found with r value of 0.898 as compared with r value of silt (0.923) estimated by soil model. The scatter diagram of the standardised estimation errors versus the estimated values looks like a round cloud with r value of 0.299 and symmetrical to the horizontal axis. The histogram of the standardised estimation errors is symmetrical but with some points found outside the range (-2.5 – 2.5) , which causes standard deviation to significantly differ from 1 (1.74). The mean standardized error is close to 0 (0.048) which means that the model is unbiased, but sample variance was underestimated.

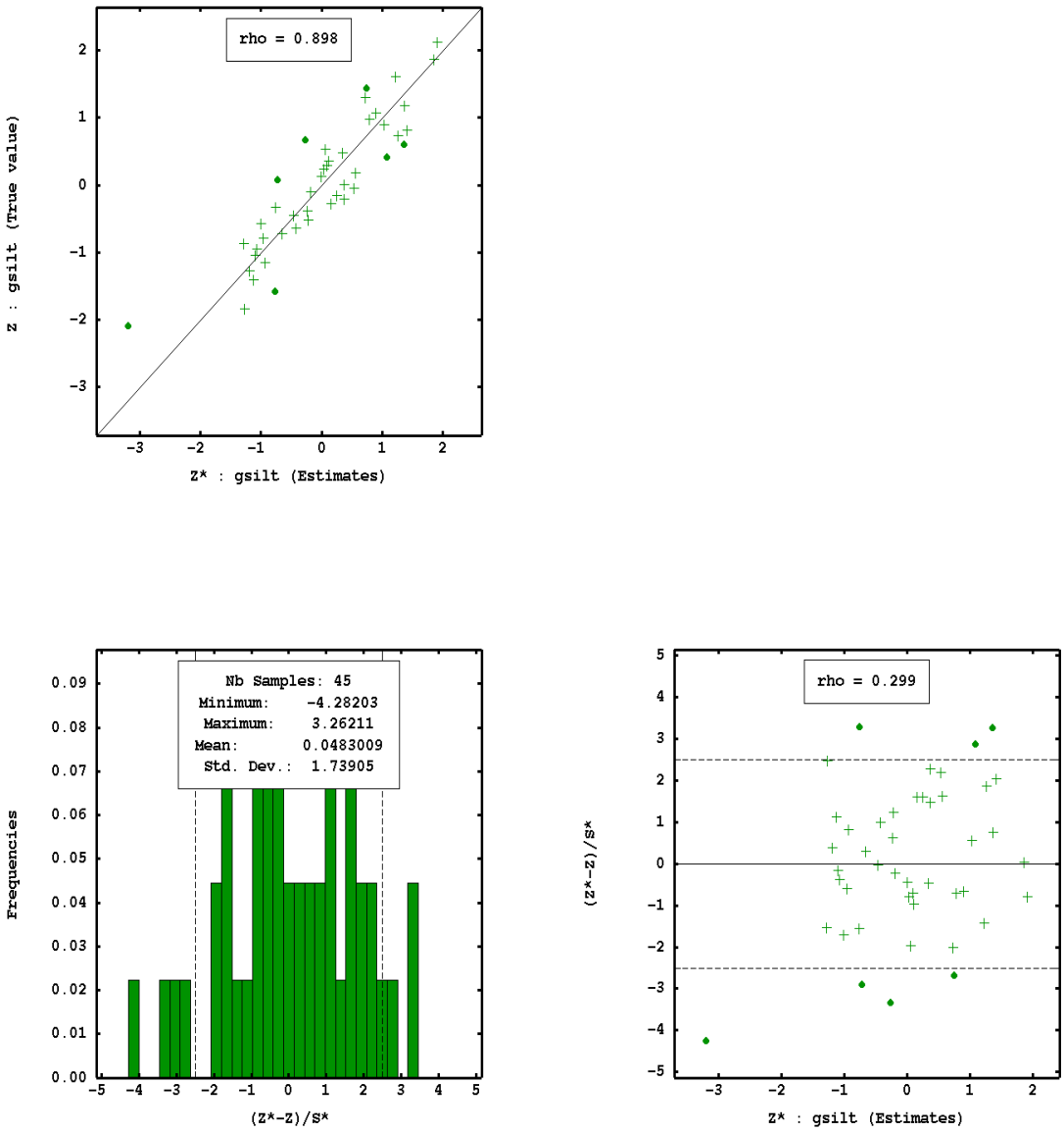

Figure 62. Outcomes of the cross-validation of silt content interpolated by fusion model.

A good correlation between the true and estimated values of TC using the fusion model (Figure 63) was found with r value of 0.868. The scatter diagram of the standardised estimation errors versus the estimated values looks like a round cloud with r value of 0.044 and symmetrical to the horizontal axis. The histogram of the standardised estimation errors is symmetrical with only one outlier less than -2.5. The mean standardized error is close to 0 (0.024) which means that the model is unbiased and the standard deviation of standardized error is close to 1 (1.05) which means that the model estimation is accurate.

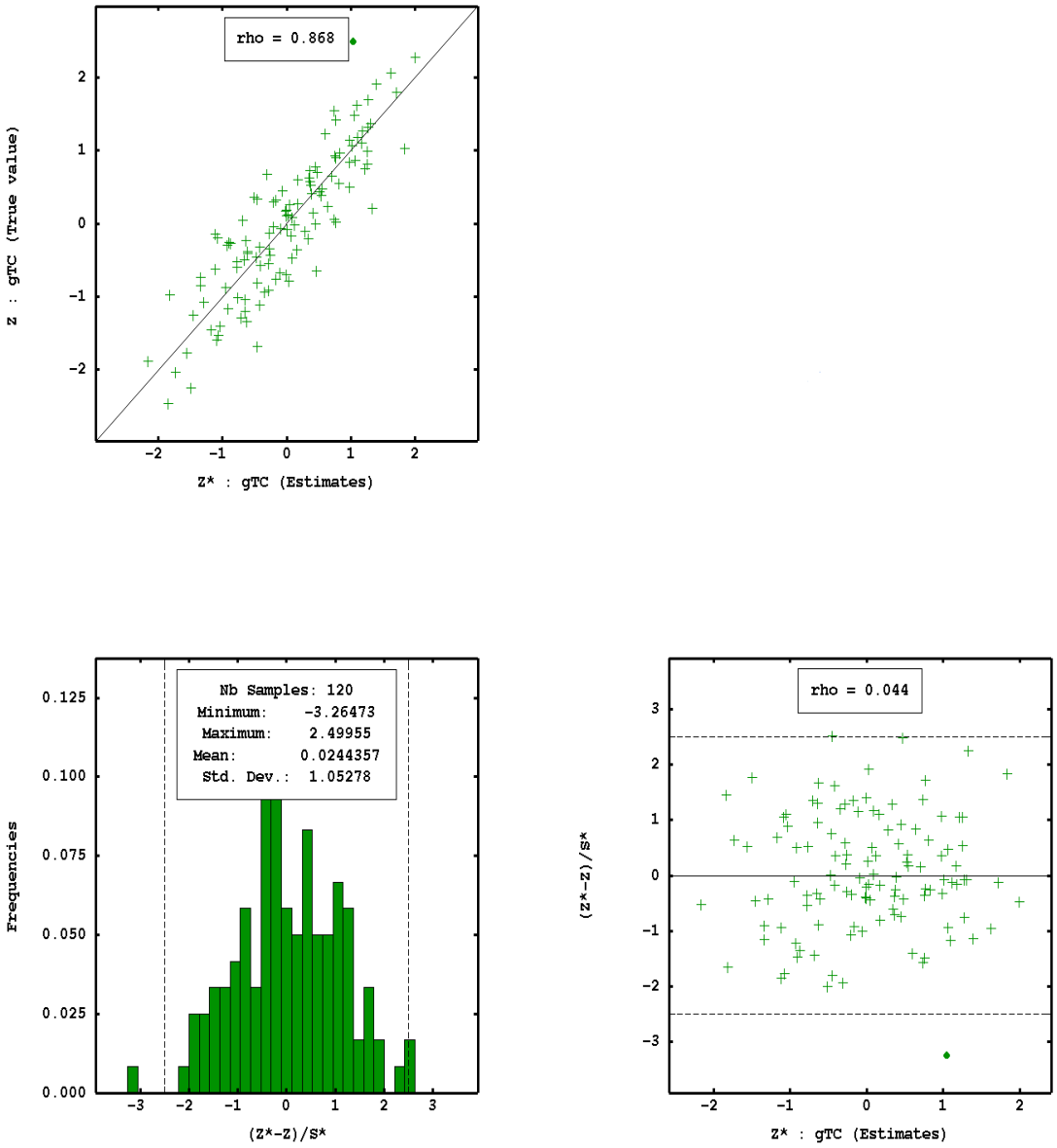

Figure 63. Outcomes of the cross-validation of total carbon interpolated by fusion model.

A slightly weaker correlation between true and estimated values of TN using the fusion model (Figure 64) is found with r value of 0.730 as compared with r value of silt (0.761) estimated by soil model. The scatter diagram of the standardised estimation errors versus the estimated values looks like a round cloud with r value of 0.299 and symmetrical to the horizontal axis. The histogram of the standardised estimation errors is symmetrical but with some points found outside the range (-2.5 – 2.5), most of them less than -2.5. The mean standardized error is close to 0 (-0.012) which means that the model is unbiased and the standard deviation of standardized error is close to 1 (1.22) which means that the model estimation is accurate.

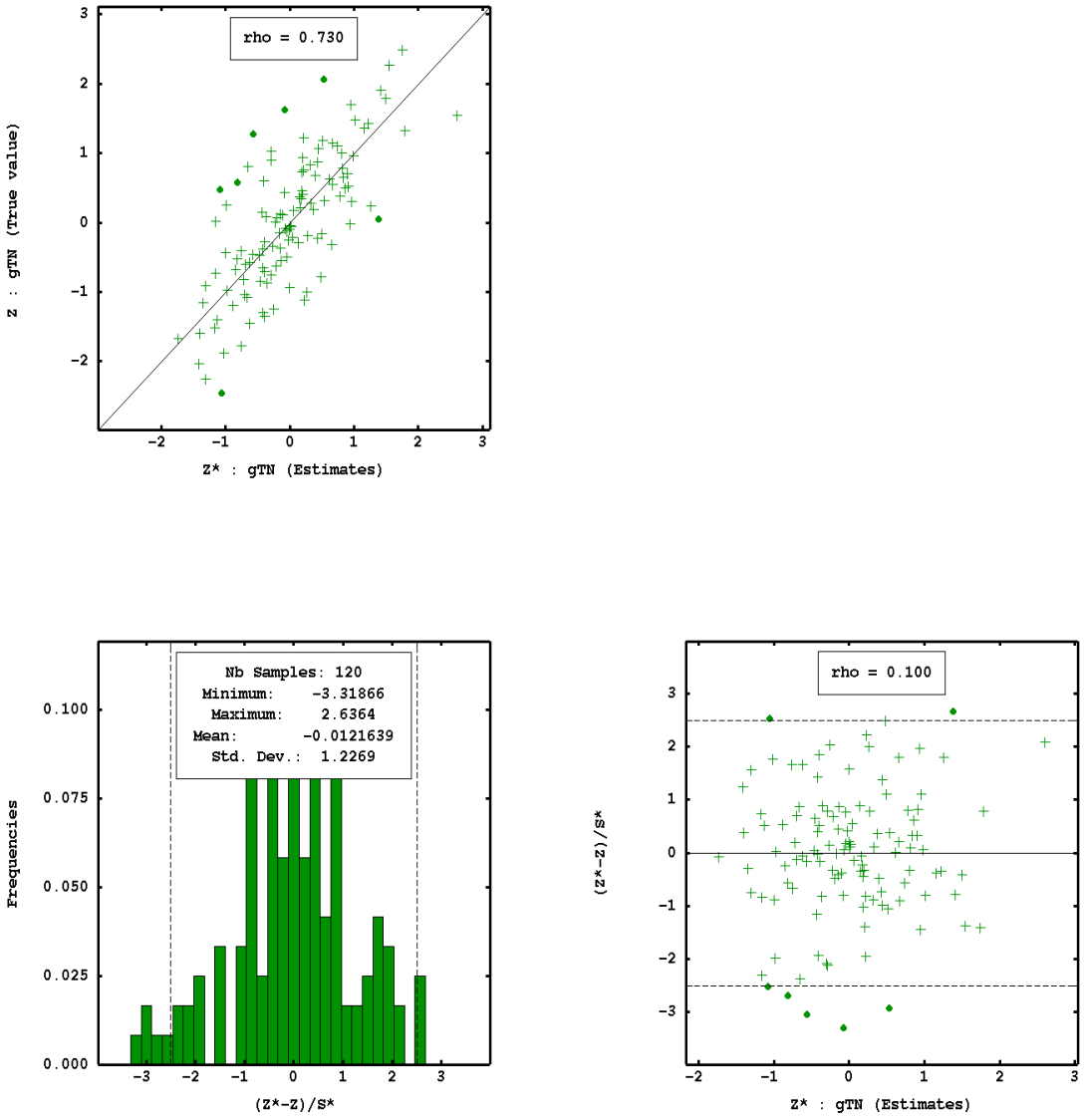

Figure 64. Outcomes of the cross-validation of total nitrogen interpolated by the fusion model.

Summing up the results of cross-validation, we can state that the fitting of the fusion model was satisfactory for all variables, though the model generally underestimates high values and overestimates low values.

6.2.2 Thematic spatial maps of the studied soil properties

6.2.2.1 Soil model

The following maps are interpolated maps using the soil model. CEC spatial map (Figure 65) shows that the CEC values increase from S to N along a direction parallel to the longer axis of the field.

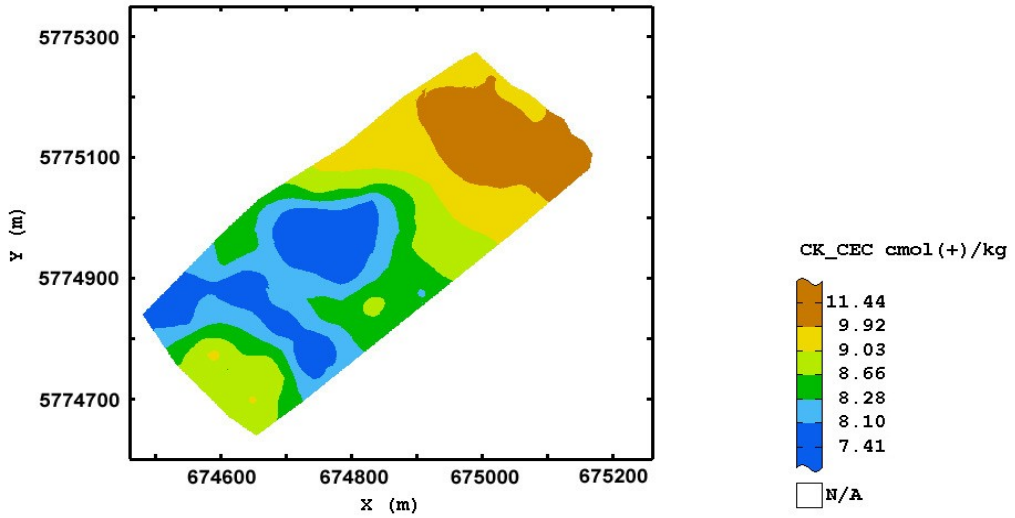

Figure 65. Spatial map of CEC interpolated by the soil model

Exchangeable potassium map (Figure 66) shows that the higher values of potassium occur in the southern part of the field, and vary conversely to what was observed for CEC.

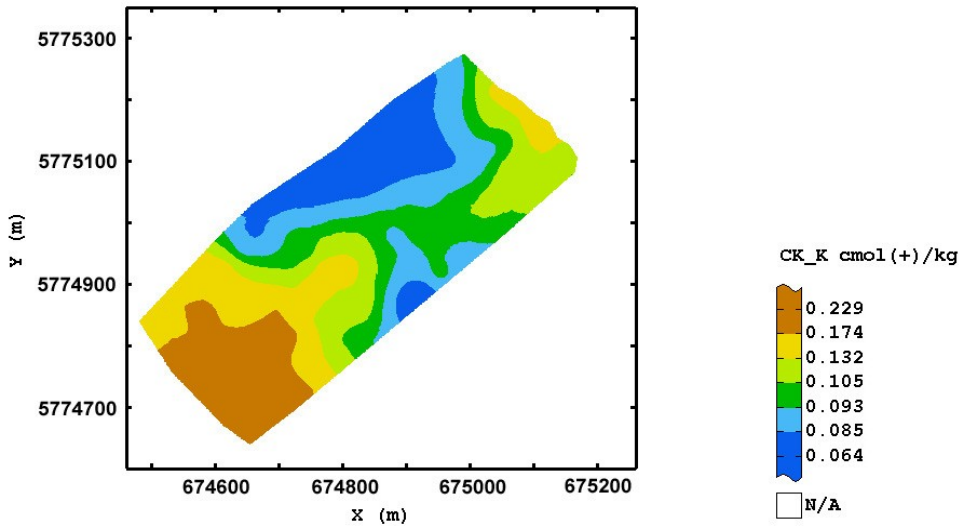

Figure 66. Spatial map of exchangeable potassium interpolated by the soil model

The extractable phosphorus map (Figure 67) shows a distinct increasing gradient, quite similar to the one observed for K.

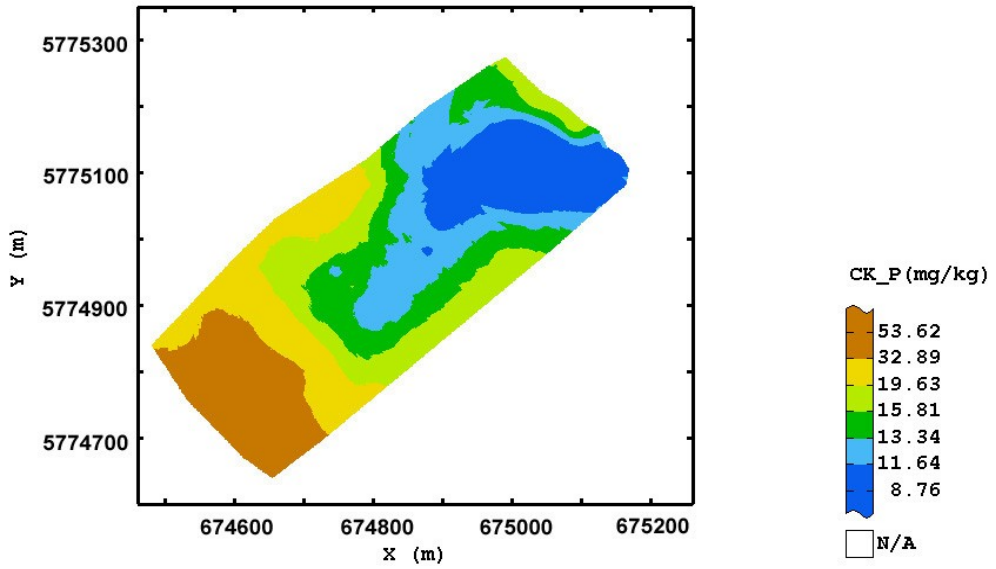

Figure 67. Spatial map of extractable phosphorus interpolated by the soil model

pH spatial map (Figure 68) shows a large spatial heterogeneity. The higher values of pH are found along the western and eastern sides of the field, whereas the central zone has pH values which vary from 6 to 6.8.

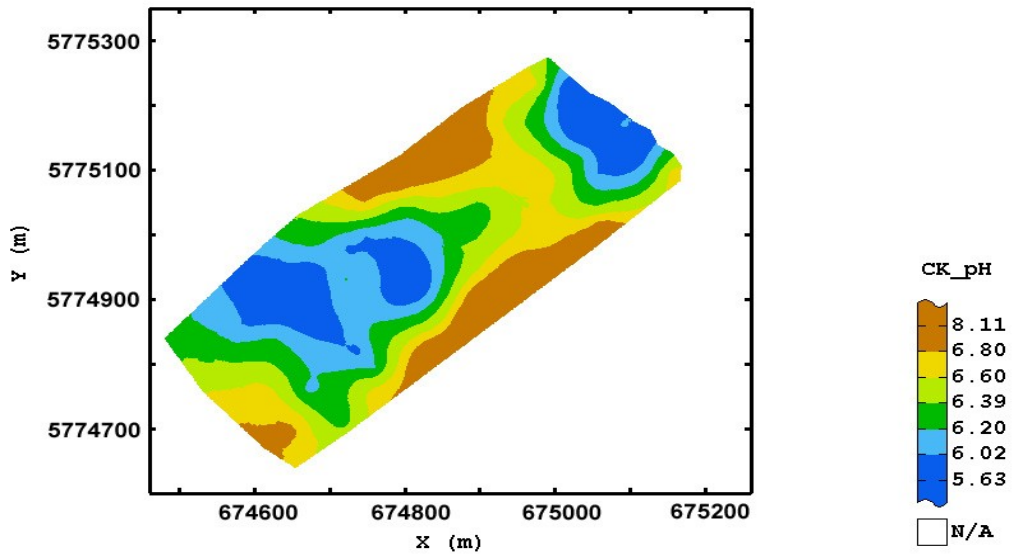

Figure 68. Spatial map of soil pH interpolated by the soil model

Sand spatial map (Figure 69) shows a distinct spatial gradient, increasing from NE to SW, so it appears sensible to split the field into three areas with different sand content.

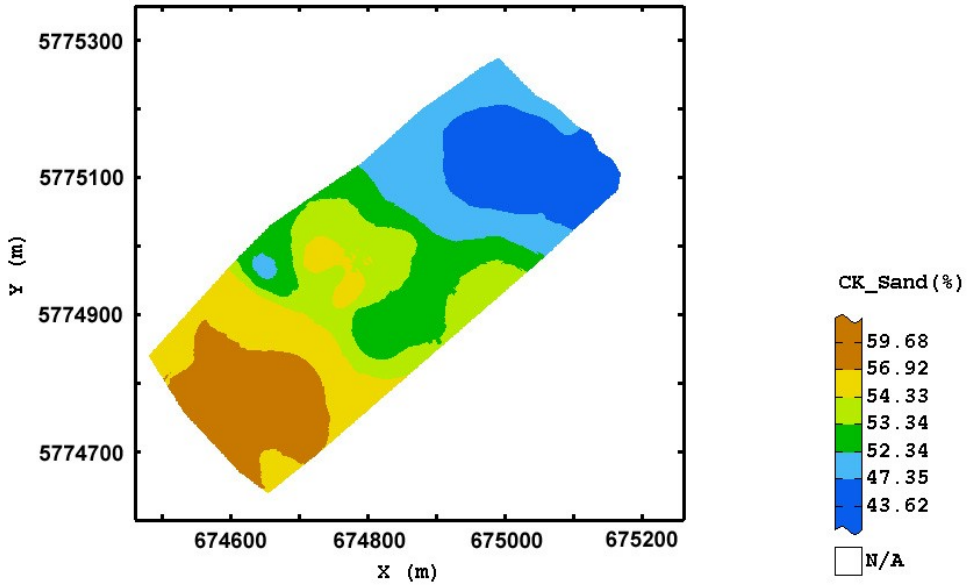

Figure 69. Spatial map of sand content interpolated by the soil model

Unlike the sand map, the silt map (Figure 70) shows a trend of increasing values from S to N content in the SW to NE direction. The highest values are found in the NE part of the field and in a small area centered at X= 674700 m and Y= 5774950 m. Sand spatial map shows a smaller area at the same coordinates that has low content of sand.

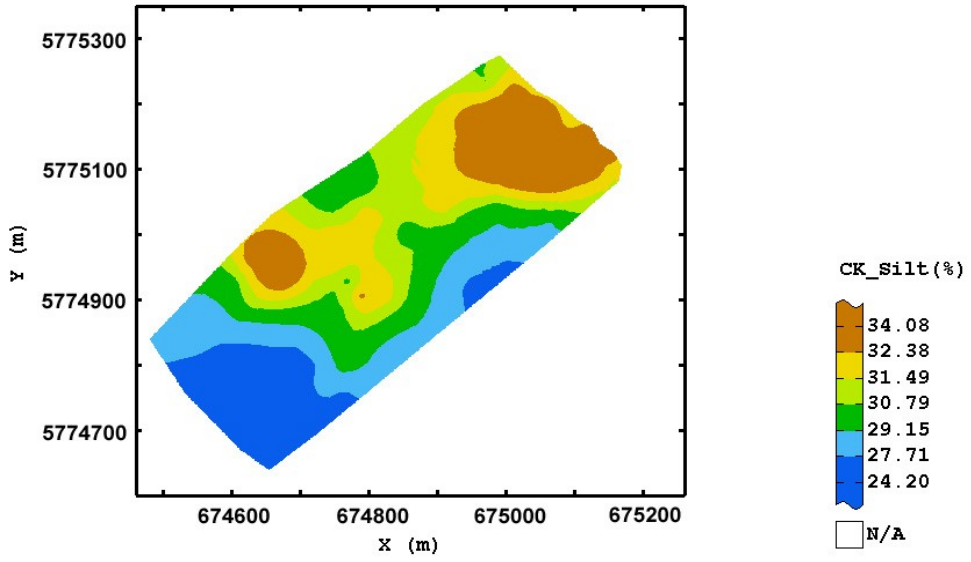

Figure 70. Spatial map of silt content interpolated by the soil model

The spatial map of total carbon (Figure 71) shows that there exists a trend of increasing values of TC from SW to NE with a wide area of the lowest values in the south western part of the field.

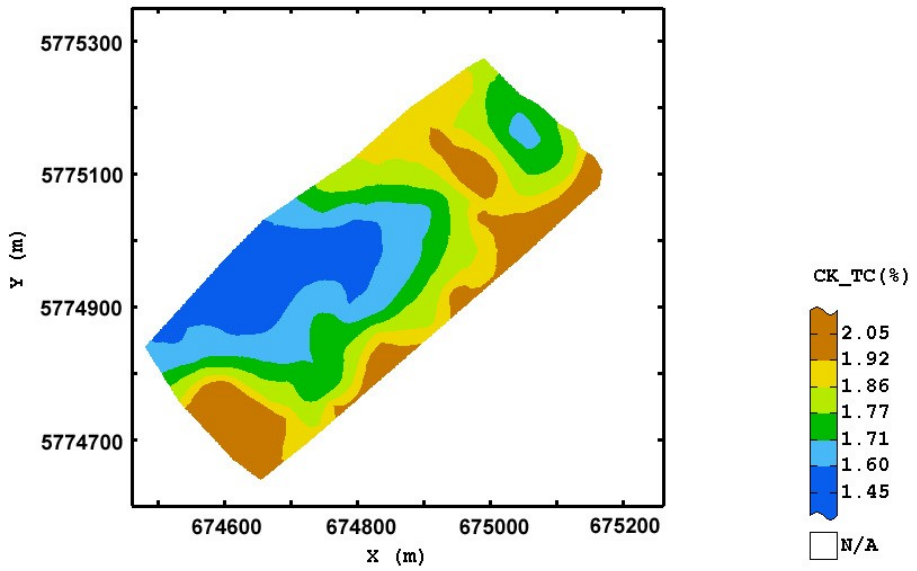

Figure 71. Spatial map of total carbon interpolated by the soil model

The spatial map of total nitrogen (Figure 72) shows quite similar to the previous one with a wide central zone of the field characterized by low contents. A trend of increasing values can be seen from outside to the inside of the field.

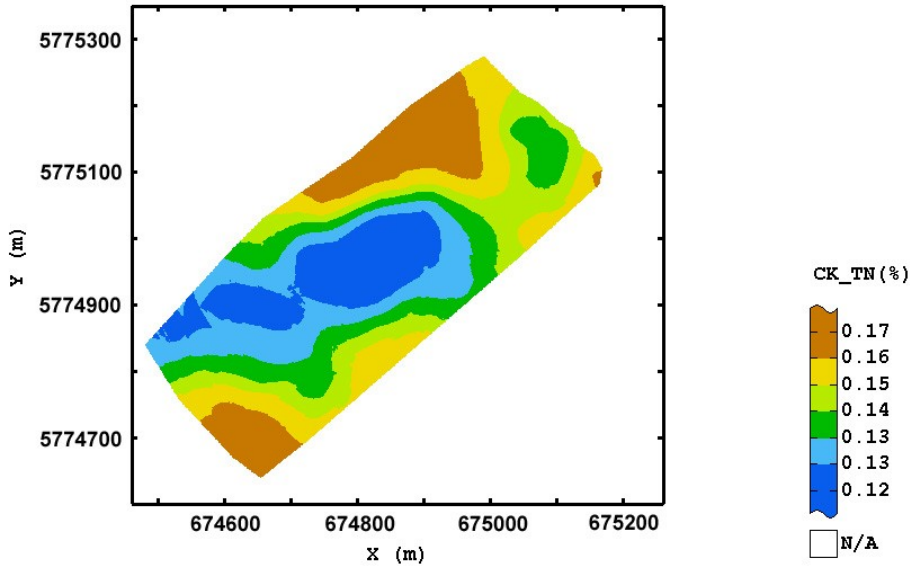

Figure 72. Spatial map of total nitrogen interpolated by the soil model

6.2.2.2 Spectral model

Figures 73-76 show the maps of the soil properties estimated from the spectra and interpolated using the spectral model. By the naked eye, they look much noisier compared to previous maps. This can be attributed to two main types of error: the first one is the error of the calibration model developed to predict the spectral variable; the second one is the error of the spectral model developed for spatial prediction.

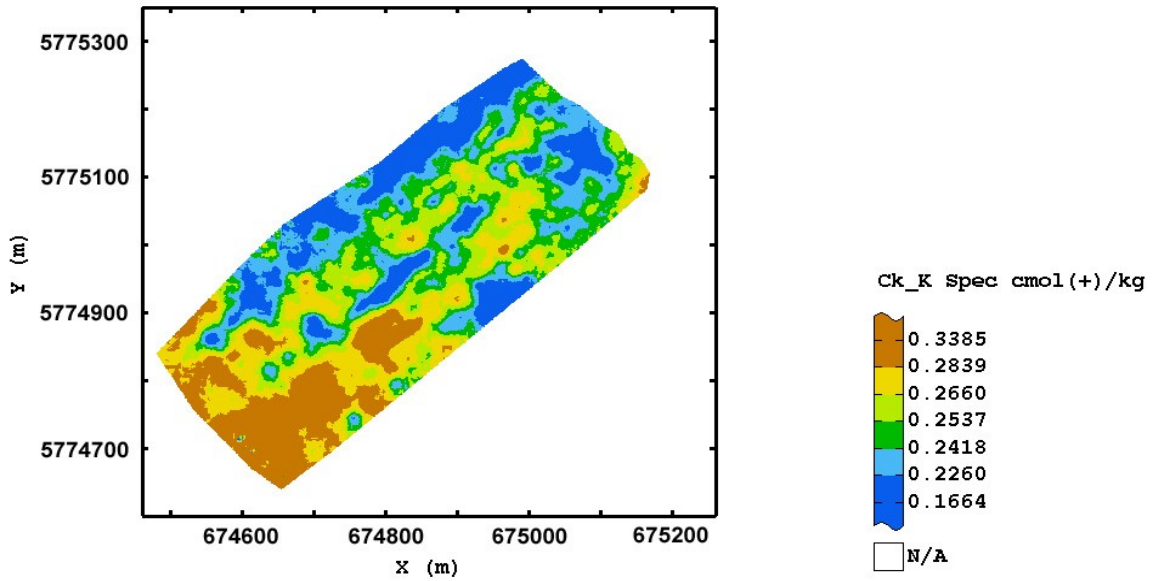

Figure 73. Spatial map of exchangeable potassium interpolated by the spectral model

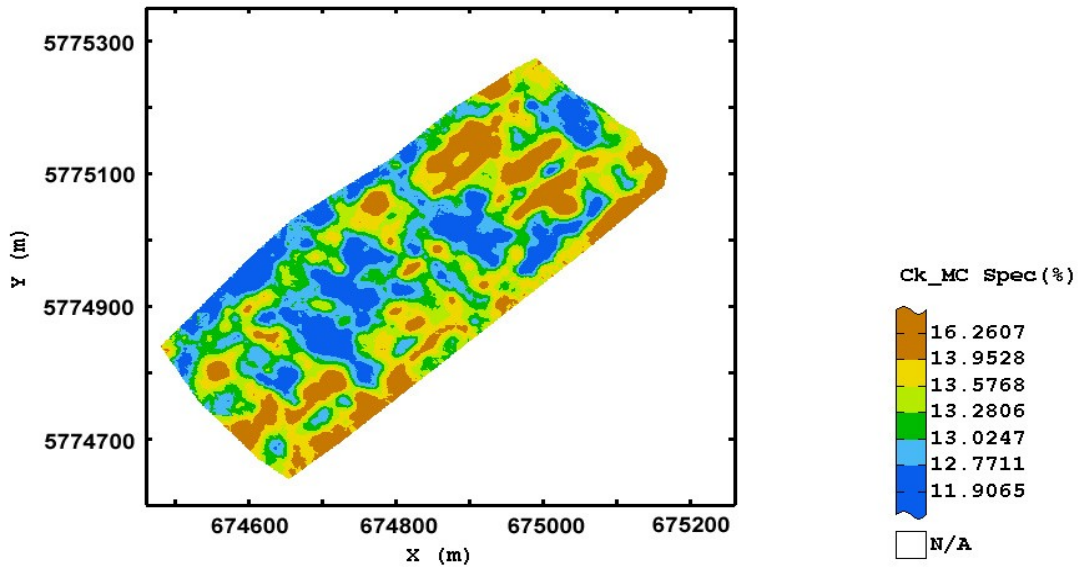

Figure 74. Spatial map of moisture content interpolated by the spectral model

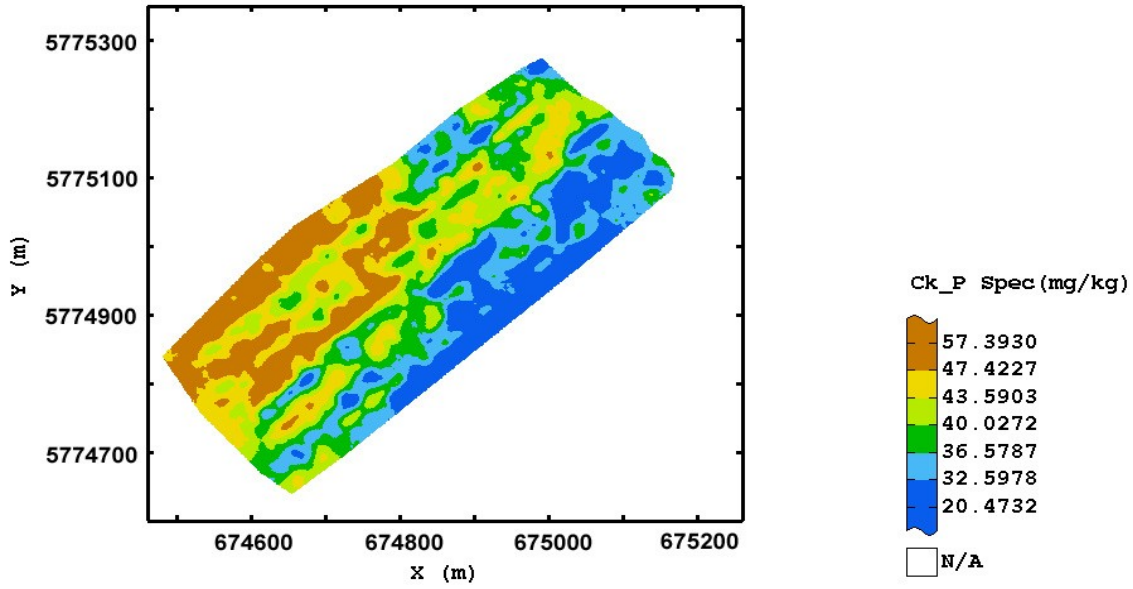

Figure 75. Spatial map of extractable phosphorus interpolated by the spectral model

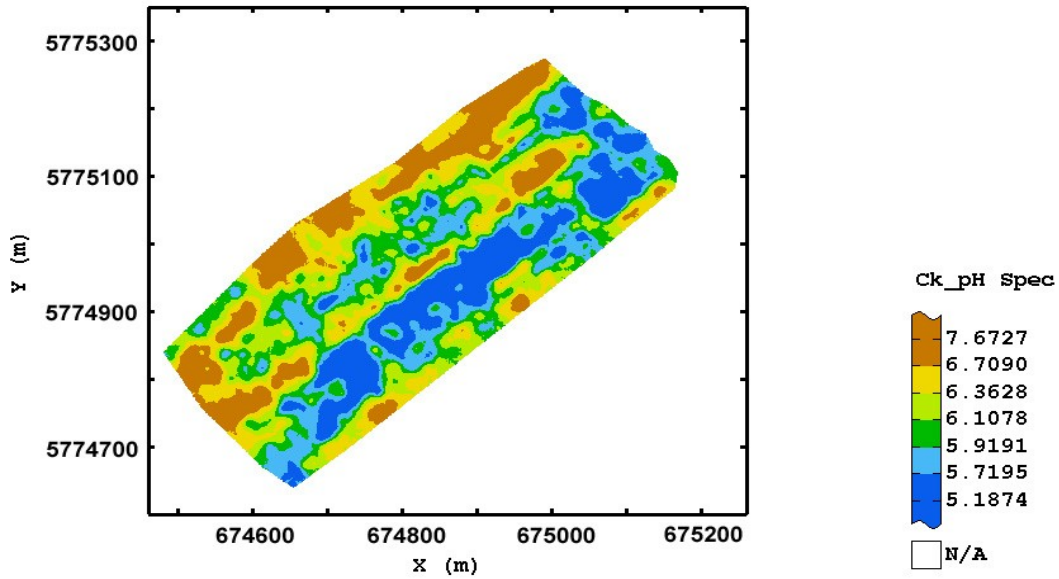

Figure 76. Spatial map of soil pH interpolated by the spectral model

However, the K map shows a decreasing trend from SW to NE similarly to the map of K obtained from soil sample analysis. Based on the MC map it is quite difficult to extract some clear structure of spatial dependence. The P and pH maps are the ones that differ most from the corresponding maps obtained from the soil model. Moreover, in the last four maps some artefacts along the main direction of driving are evident.

6.2.2.3 Fusion model

The following spatial maps were developed using the fusion model.

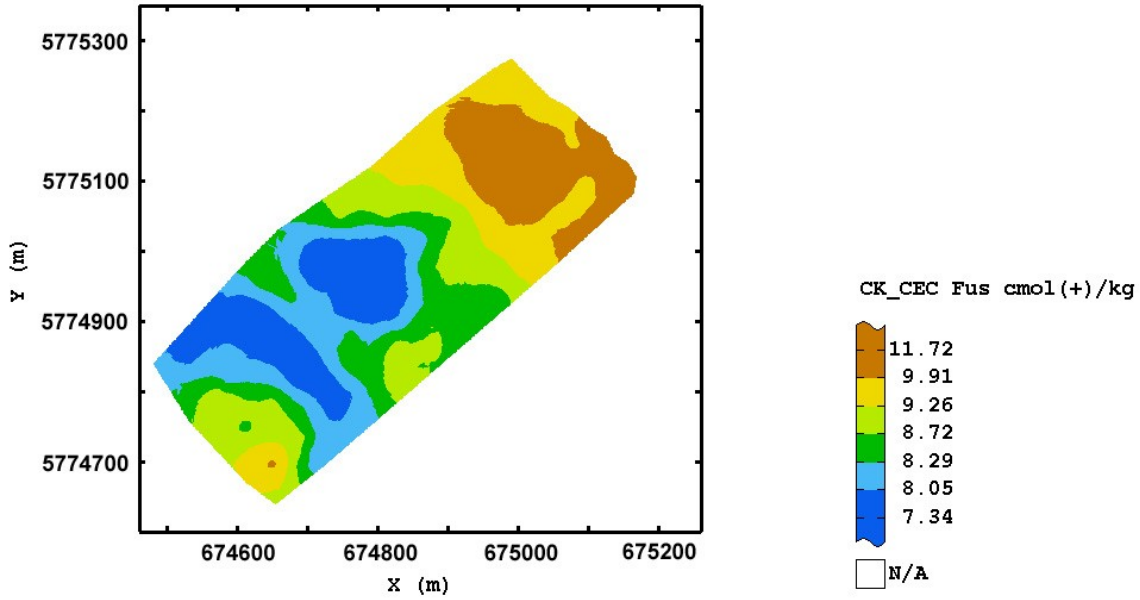

Figure 77. Spatial map of CEC interpolated by the fusion model

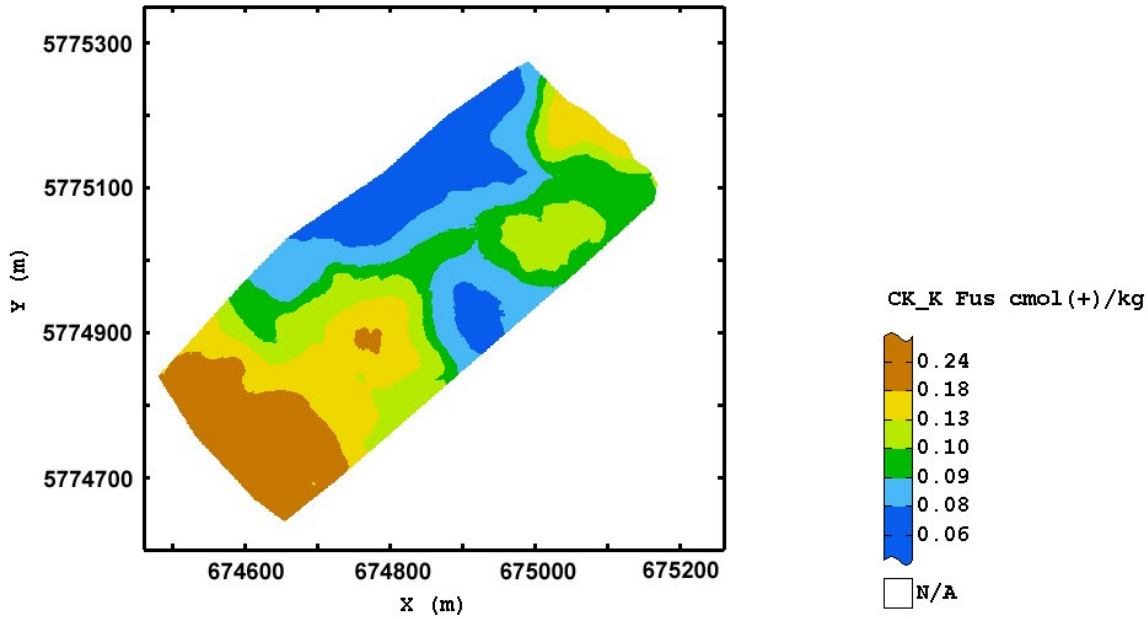

Figure 78. Spatial map of exchangeable potassium interpolated by the fusion model

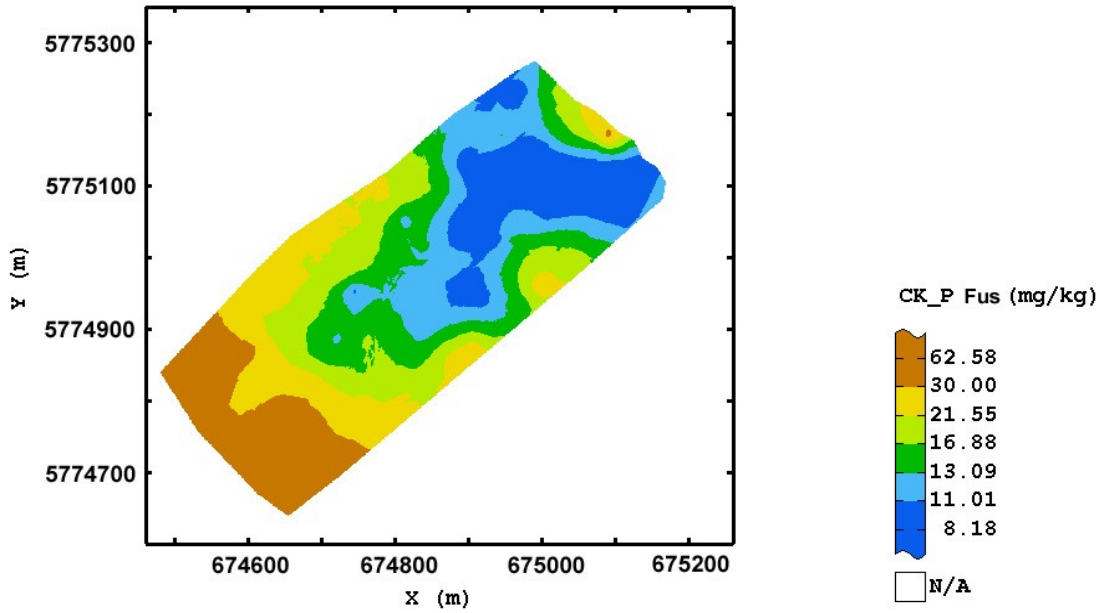

Figure 79. Spatial map of extractable phosphorus interpolated by the fusion model

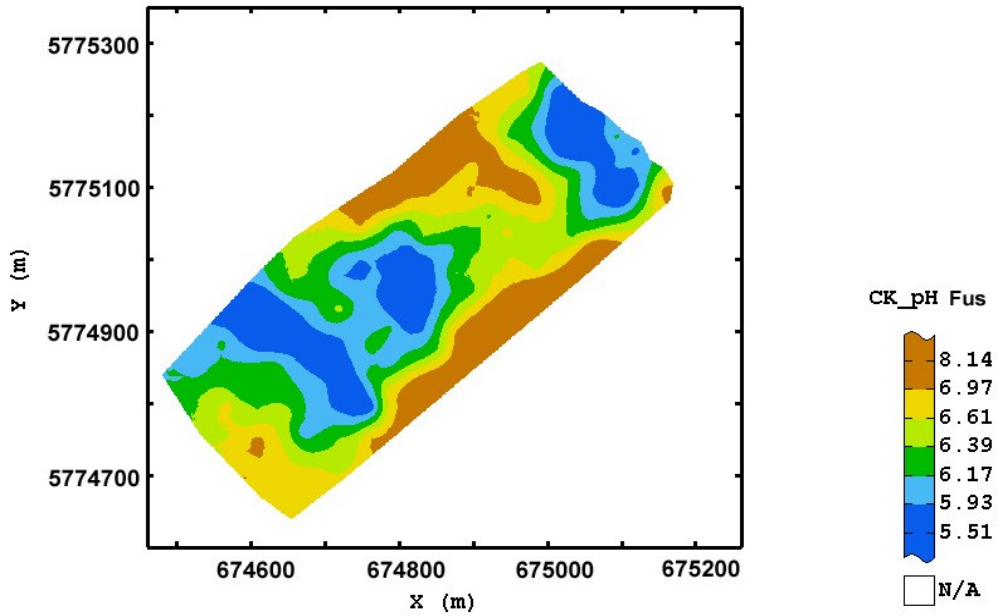

Figure 80. Spatial map of soil pH interpolated by the fusion model

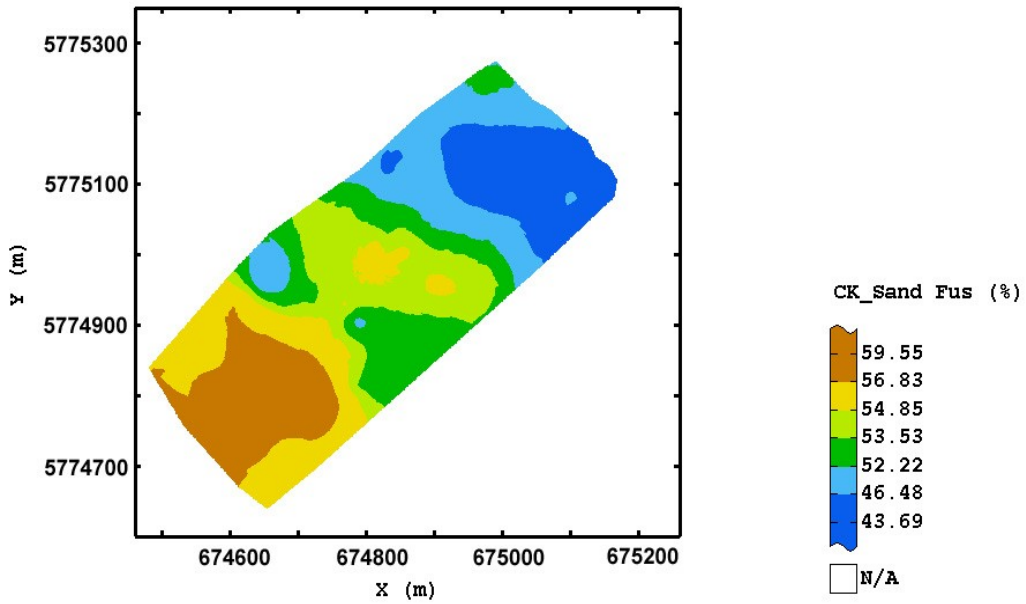

Figure 81. Spatial map of sand content interpolated by the fusion model

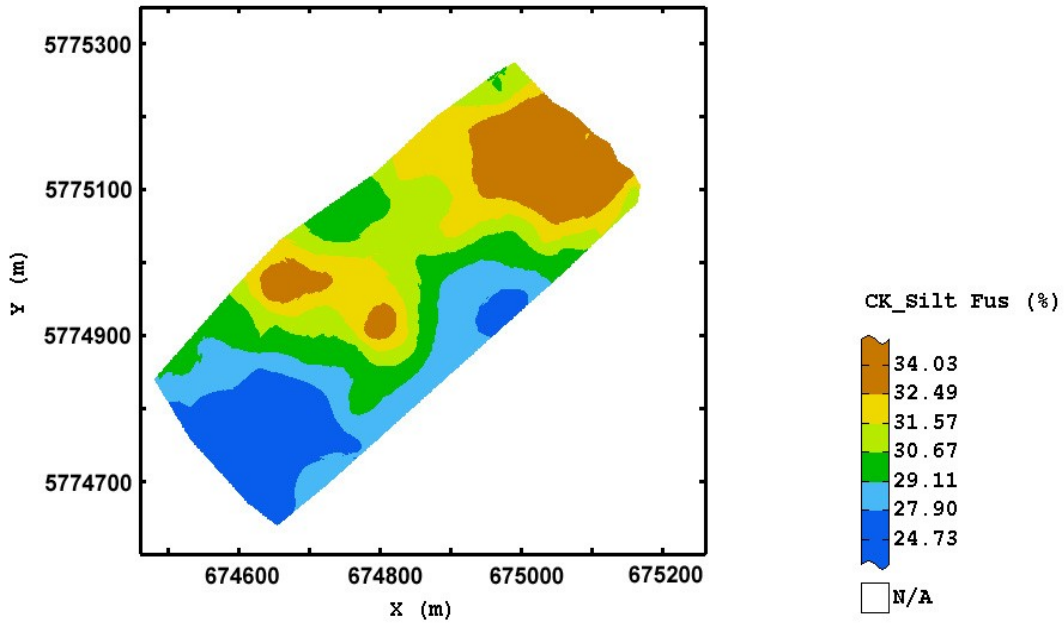

Figure 82. Spatial map of silt content interpolated by the fusion model

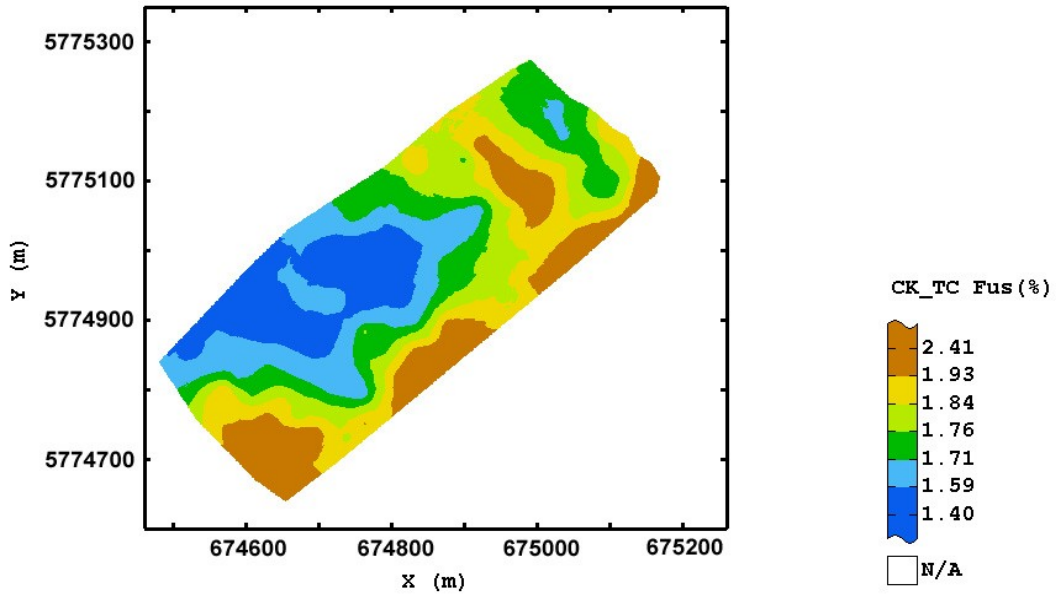

Figure 83. Spatial map of total carbon interpolated by the fusion model

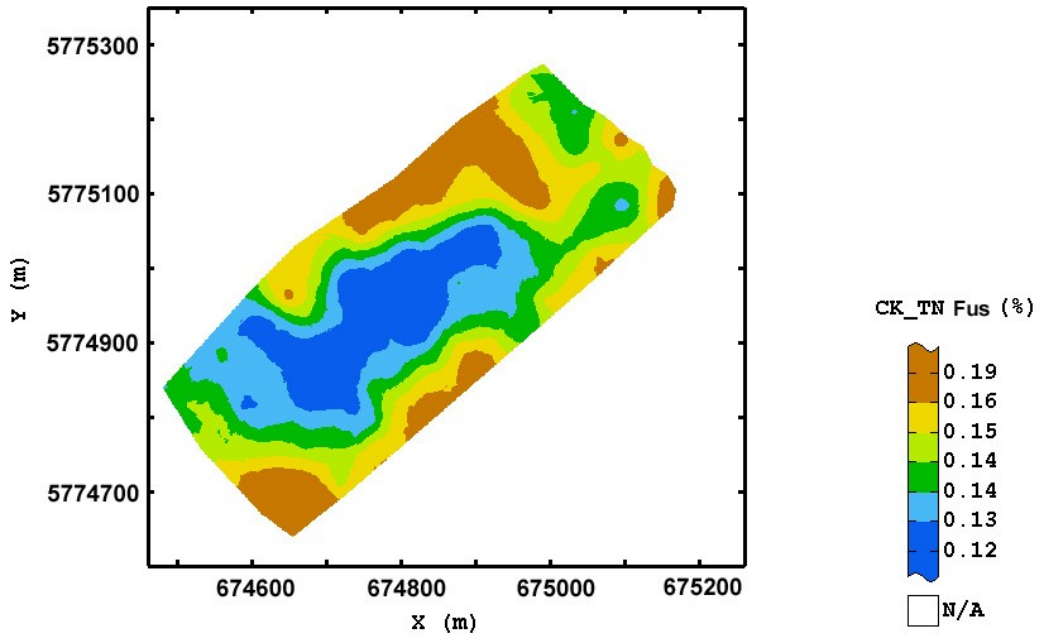

Figure 84. Spatial map of total nitrogen interpolated by the fusion model

The soil property maps (Figures 77-84) obtained with the fusion model look quite similar to the ones calculated from the soil model, with an increased variability at short scale due to the introduction of the fine-surveyed spectral variables.

6.2.3 Validation

Table 13 summarizes the basic statistics of the validation data set. K and P are the variables that show the largest variability. The validation data set includes different numbers of samples for each variable because not all soil samples (183) were subjected to all analyses reported in this study, i.e. 183 soil samples were analyzed for pH, TC, TN and MC, whereas 95 soil samples were analyzed for extractable P, CEC and exchangeable cations. 72 soil samples were analyzed for particle size distribution (sand, silt, clay). Then the whole data set of soil analysis was divided randomly into two subsets: the calibration data set containing 2/3 of soil samples and the validation data set containing 1/3 of soil samples for each variable analysed.

Table 13. Statistics of validation data set

Variable	Count	Minimum	Maximum	Mean	Std. Dev	Variance	CV %
Sand (%)	27	42.52	59	51.46	3.87	14.95	7.52
Silt (%)	27	26.5	33.79	30.16	1.82	3.3	6.03
TC (%)	63	1.03	2.37	1.76	0.28	0.08	15.91
TN (%)	63	0.09	0.2	0.14	0.02	0	14.29
CEC (cmol+.kg^{-1})	34	7.46	11.5	9.11	1.02	1.05	11.20
K (cmol+.kg^{-1})	34	0.06	0.36	0.11	0.05	0	45.45
P (mg.kg^{-1})	34	4.8	47.6	16.45	9.8	96.02	59.57
pH	63	5.16	8.21	6.63	0.82	0.66	12.37

For the validation data set being representative of the calibration data set, the means and the variances of the corresponding variables should not be statistically different. So, firstly a Bartlett's test was performed to verify the homogeneity of variance. The results reported in Table 15 show that only for K and P the assumption of homogeneity of variance cannot be accepted. Therefore, a t test with unequal variances was carried out for these two variables, whereas for the remaining variables the variances were assumed equal, in order to test the equality of means. Table 14 shows the results of the t test, where it results that K and P differ in the two data sets also in terms of means. These results show the necessity to extract another data set, independent of the calibration data set to validate the three models for P and K.

Table 14. T-test between the means of the variables of calibration and validation data sets

Variables	Calibration		Validation		p-values	α	Interpretation
	Mean	Variance	Mean	Variance			
CEC	8.94	1.10	9.11	1.08	0.466	0.05	Means are the same
pH	6.49	0.54	6.63	0.68	0.274	0.05	Means are the same
P	23.38	202.16	16.45	98.93	0.007	0.05	Means are different
TC	1.78	0.05	1.76	0.08	0.745	0.05	Means are the same
TN	0.15	0.00	0.14	0.00	0.564	0.05	Means are the same
Sand	52.93	20.91	51.46	15.53	0.153	0.05	Means are the same
Silt	29.91	5.98	30.16	3.43	0.624	0.05	Means are the same
K	0.23	0.01	0.11	0.00	0.000	0.05	Means are different

Table 15. Bartlett's test / Two-tailed test between variables of calibration and validation data sets

	CEC	pH	TC	TN	Sand	Silt	P	K
Chi-square (Observed value)	0.003	1.002	3.285	0.000	0.691	2.348	4.933	5.082
Chi-square (Critical value)	3.841	3.841	3.841	3.841	3.841	3.841	3.841	3.841
DF	1	1	1	1	1	1	1	1
p-value (one-tailed)	0.954	0.317	0.070	0.986	0.406	0.125	0.026	0.024
alpha	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05

As regards mean error (ME) of the three geostatistical models, the fusion model was better than the soil model for CEC, TC, silt and pH (Table 16). Conversely, the soil model was better than the fusion model for sand, TN and P. For K, both soil and fusion models show similar mean errors. The soil model produces the minimum deviation from 0 of mean error for TN, whereas the fusion model produces the minimum deviation from 0 of mean error for CEC. The minimum deviation from 1 of MSSE was obtained by the soil model for P (0.7), whereas the minimum deviation from 1 of MSSE was obtained by the fusion model for silt content (0.66). For spectral variables, the results show that ME, MSSE are not satisfactory because of the high deviation from 0 and 1 for ME and MSSE respectively with the exception of K that produces the minimum deviation from 0 of ME in the spectral model. Concerning RMSE, the fusion model was better than the soil model for CEC, pH, P and K because it shows the minimum deviation from 0 of RMSE. In the other hand, the soil model was better than the fusion for silt and sand. Both soil and fusion models show the same values of RMSE for TN and TC.

Table 16. Mean error (ME), mean standardized squared error (MSSE), and root mean squared error (RMSE) obtained from the validation data set for variables under study.

	Soil model			Fusion model			Spectral model		
	ME	MSSE	RMSE	ME	MSSE	RMSE	ME	MSSE	RMSE
CEC	-0.014	0.285	0.503	0.001	0.271	0.501	–	–	–
TC	-0.013	2.296	0.195	0.002	1.832	0.195	–	–	–
Silt	-0.534	0.496	1.169	-0.519	0.663	1.371	–	–	–
Sand	-0.005	0.254	2.017	-0.152	0.333	2.334	–	–	–
TN	0.001	1.795	0.018	0.002	1.790	0.018	–	–	–
pH	0.197	2.825	0.678	0.183	1.436	0.651	0.443	2.532	0.712
P	0.019	0.692	6.922	0.600	0.595	6.116	-22.865	11.972	7.516
K	0.000	2.079	0.047	0.000	1.599	0.044	-0.138	42.248	0.055
MC	–	–	–	–	–	–	4.568	69.530	1.491

Summing up, the two soil and fusion models show similar behavior, whereas one needs to underline the very poor performance of the spectral model which largely underestimates P and overestimates MC.

6.2.3.1 Quantile-Quntile (Q-Q) plots

6.2.3.1.1 Soil model

Q-Q plot of CEC estimates (Figure 85) shows that they are normal distributed (left Q-Q plot) with the exception of the lowest values. The right Q-Q plot compares the distribution of estimates with the one of observations. The two distributions are quite comparable, however the samples having CEC values less than 8 cmol+.kg^{-1} are overestimated by the soil model whereas those having CEC values greater than 10.5 cmol+.kg^{-1} are underestimated.

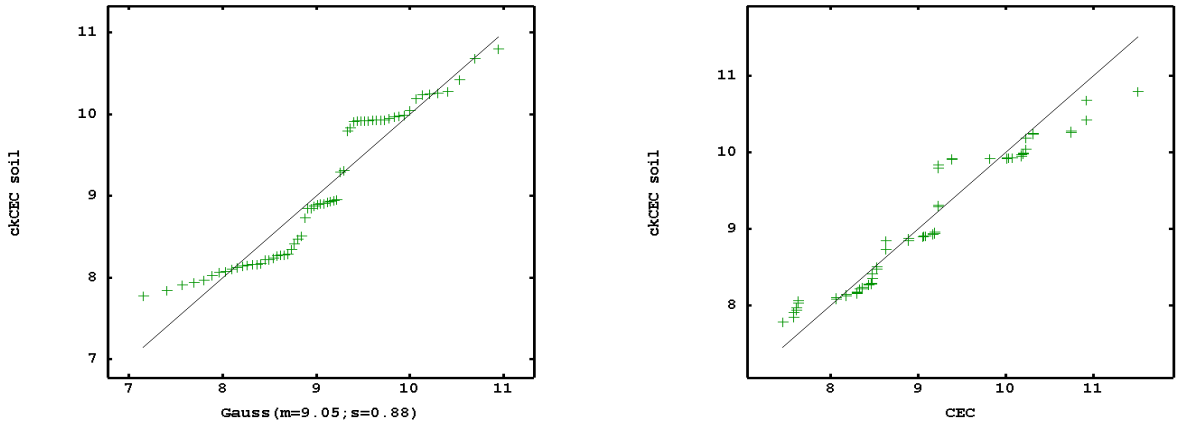

Figure 85. Q-Q plot of CEC interpolated by the soil model

Concerning exchangeable potassium, Q-Q plot (Figure 86) show that its distribution doesn't follow normal distribution (left Q-Q plot). The values of K are quite well estimated (right Q-Q plot) until values smaller than 0.20. Larger values, and more specific, greater than 0.35, is sensibly underestimated.

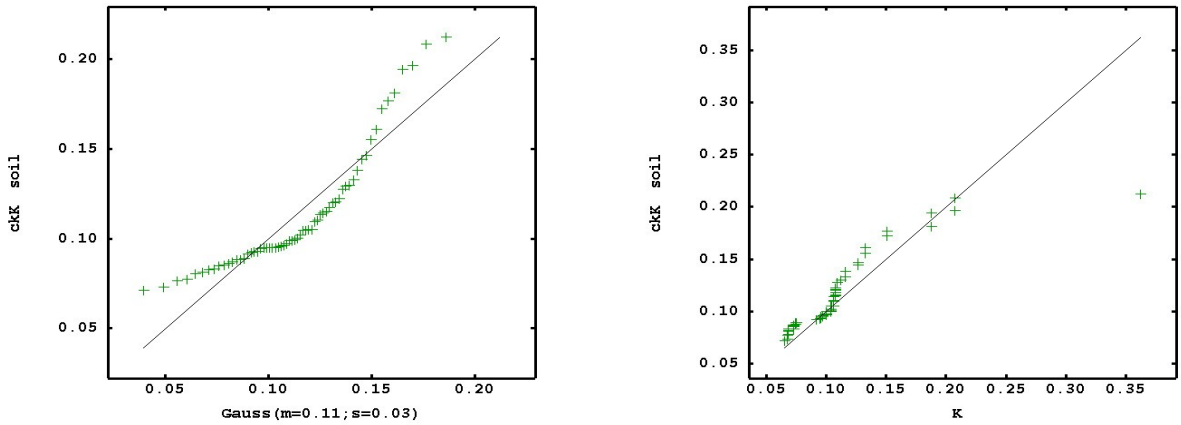

Figure 86. Q-Q plot of K interpolated by the soil model

Figure 87 shows a Q-Q plot of extractable phosphorus where there can be seen a large departure from normal distribution (left Q-Q plot). Samples having values less than 15 mg.kg^{-1} are slightly overestimated, whereas those having P content between 15 and 50 mg.kg^{-1} are well estimated (right Q-Q plot).

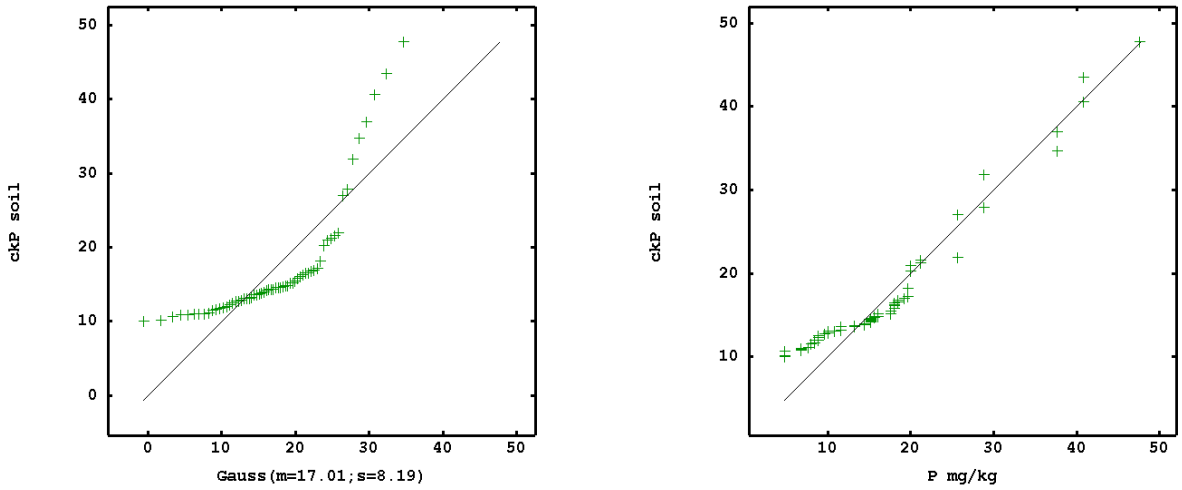

Figure 87. Q-Q plot of P interpolated by the soil model

Soil pH is normally distributed as shown in Figure 88 (left Q-Q plot). Samples that have pH values from 5 to 6 are overestimated, those varying between 6 and 7 are well estimated, whereas pH values greater than 7 are largely underestimated (right Q-Q plot).

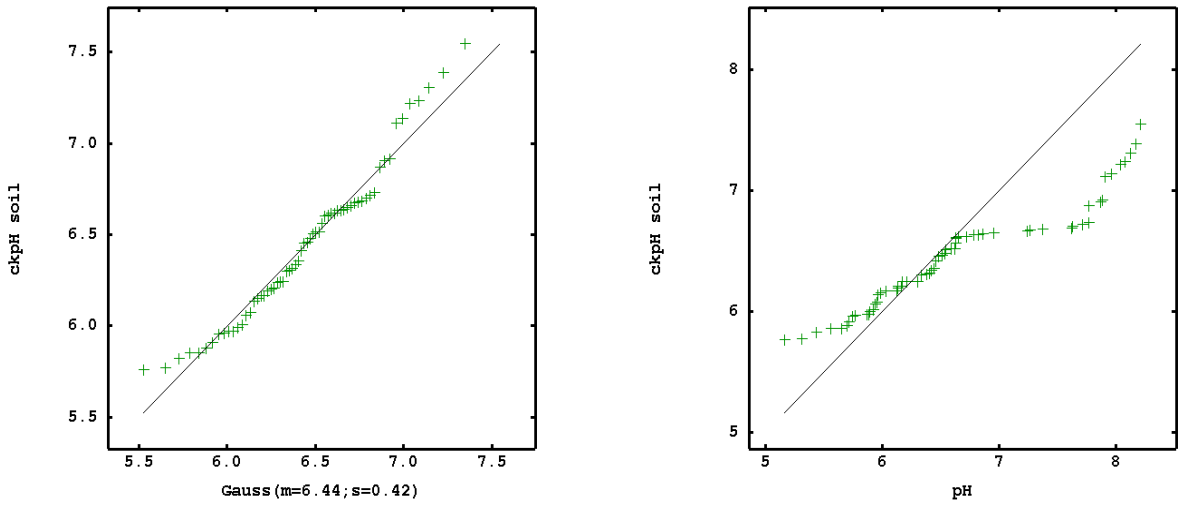

Figure 88. Q-Q plot of pH interpolated by the soil model

Figure 89 shows Q-Q plot of sand content where there can be seen that this property follows normal distribution (left Q-Q plot). The right Q-Q plot shows that sand content is well estimated over the whole range of sand content (45 - 60 %).

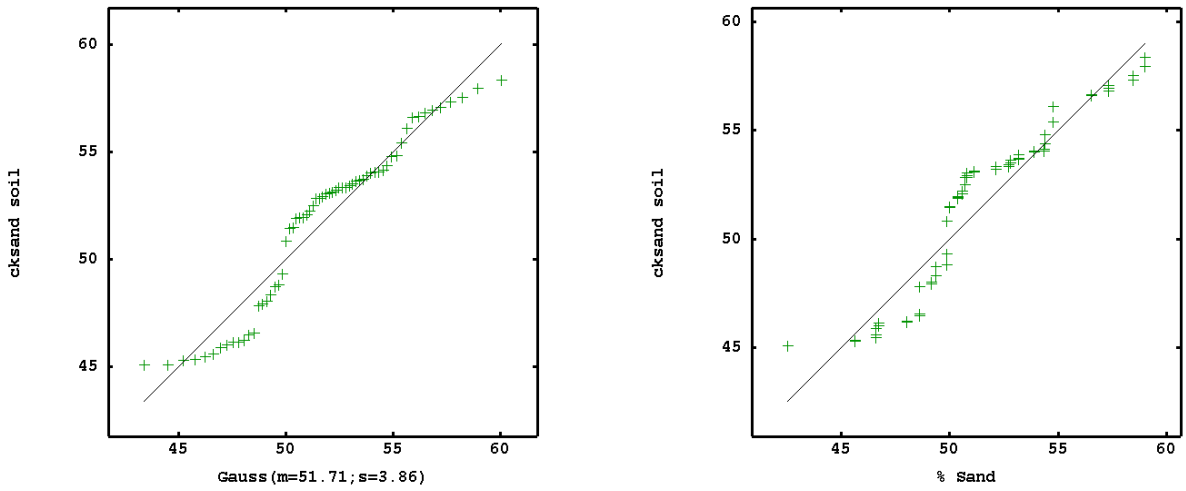

Figure 89. Q-Q plot of sand interpolated by the soil model

Q-Q plot of silt content (Figure 90) shows that it is normally distributed (left Q-Q plot). The estimated values of silt are well estimated over the whole range of silt content (26 -34 %).

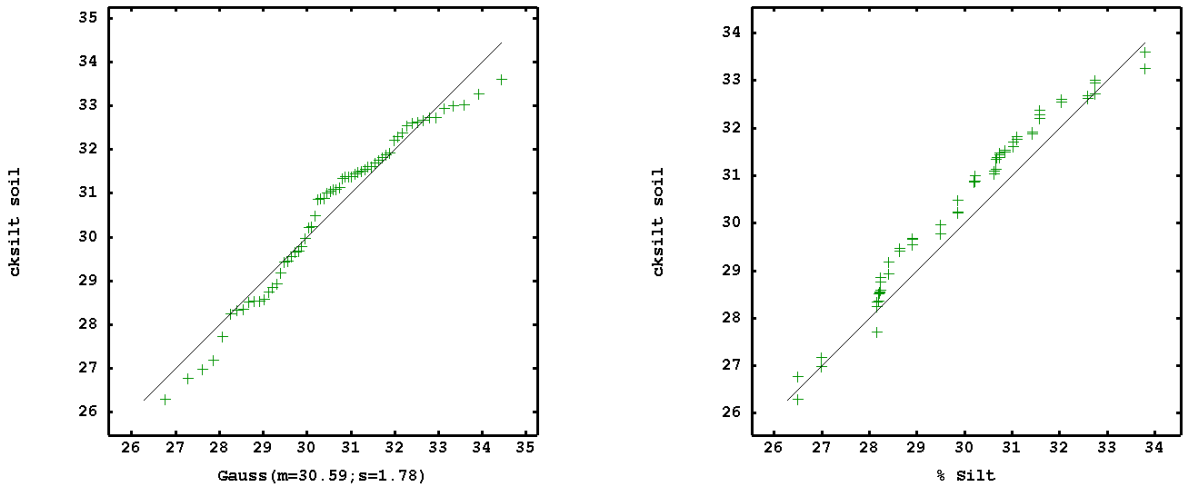

Figure 90. Q-Q plot of silt interpolated by the soil model

For total carbon, Q-Q plot (Figure 91) shows that TC follows normal distribution (left Q-Q plot). However, the TC values between 1 and 1.7 % are overestimated, whereas values approximately between 1.7 and 2 % are well estimated and values between 2 and 2.5 % are underestimated (right Q-Q plot).

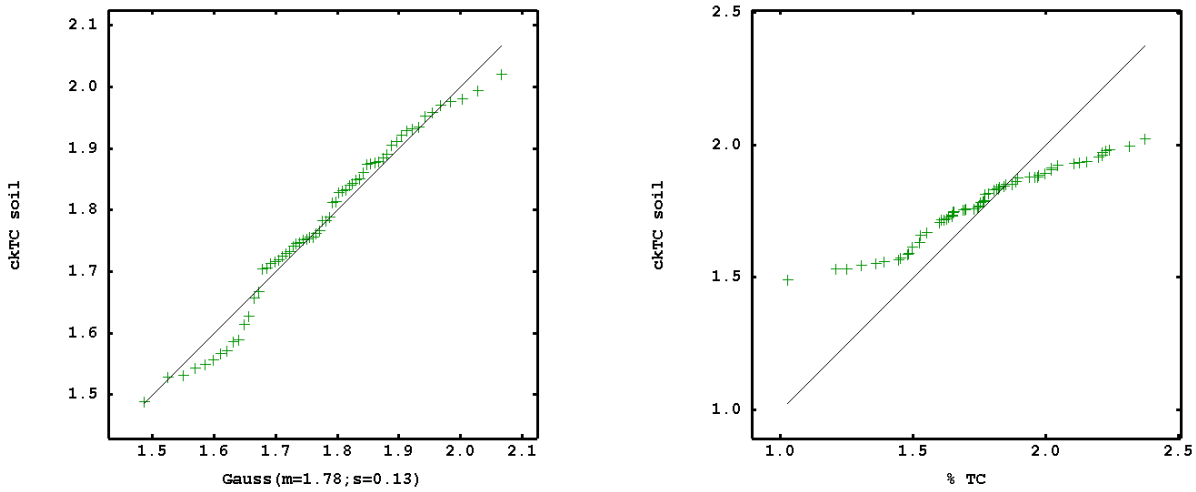

Figure 91. Q-Q plot of TC interpolated by the soil model

Figure 92 shows that total nitrogen is normally distributed (left Q-Q plot). In the right Q-Q plot, there can be seen three types of estimation. The first one is for values between 0.100 and 0.135 % that are overestimated, the second one is for values between 0.135 and 0.160 % that are well estimated and the third one is for values between 0.160 to 2 % that are underestimated.

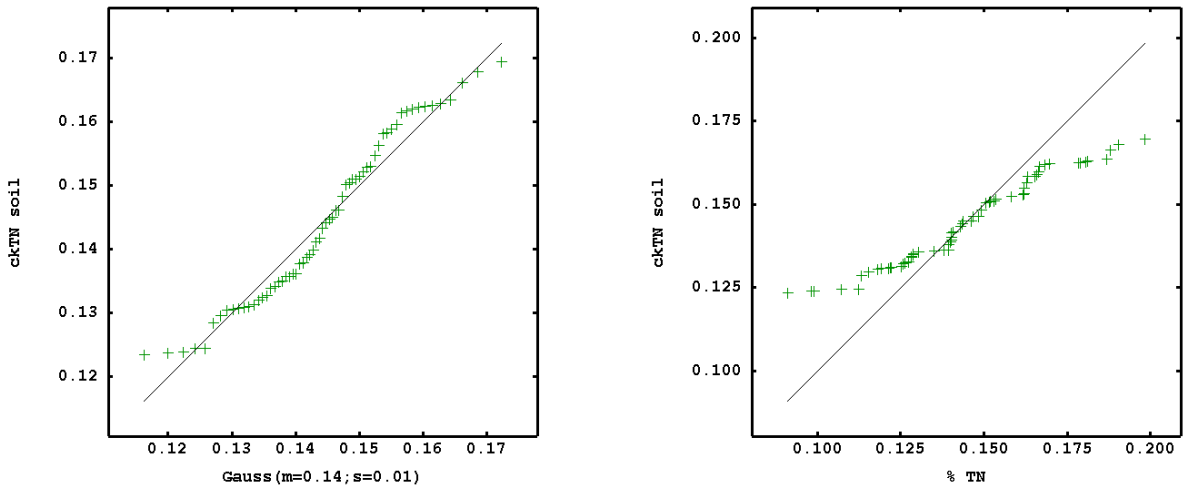

Figure 92. Q-Q plot of TN interpolated by the soil model

6.2.3.1.2 Spectral model

The left Q-Q plot (Figure 93) shows that exchangeable potassium follows normal distribution except for values less than 0.225. The right Q-Q plot shows that all values of exchangeable potassium are overestimated except for the largest value. This bad matching between estimates and observations may be due to different types of errors: one is the error of the calibration model used to predict extractable K from the spectra and the second is the error of the spectral model used for spatial prediction.

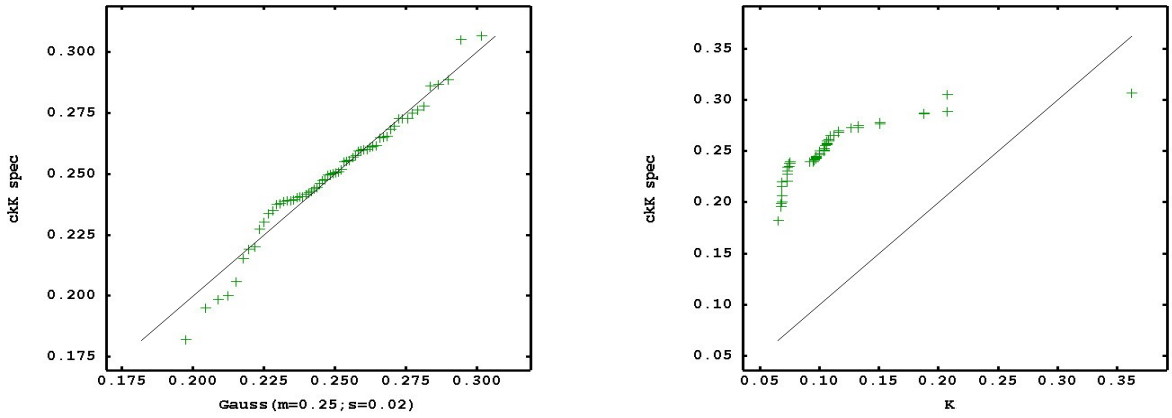

Figure 93. Q-Q plot of K interpolated by the spectral model

Although the estimated values of moisture content follow normal distribution (left Q-Q plot), the MC values (Figure 94) are severely underestimated (right Q-Q plot). This may be attributed to the causes of error reported above.

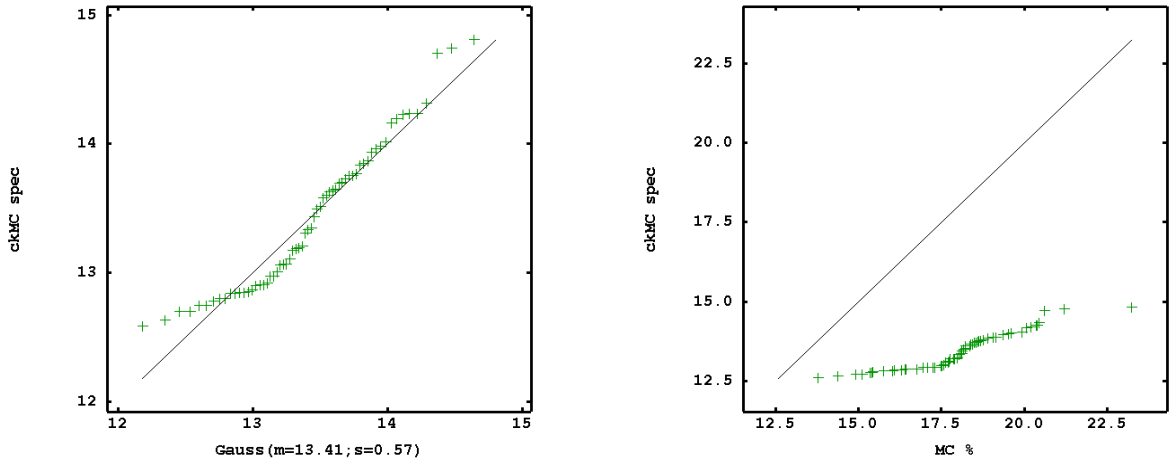

Figure 94. Q-Q plot of MC interpolated by the spectral model

The left Q-Q plot (Figure 95) shows that extractable phosphorus is normally distributed. However, also for P the severe mismatching between estimates and observations, which were largely overestimated, may be due to the approximations of the models used.

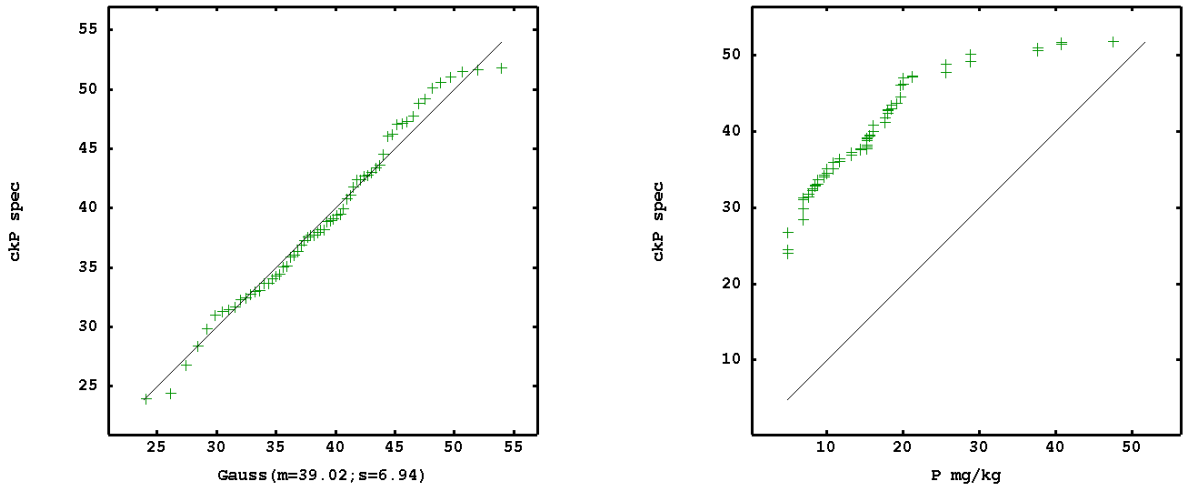

Figure 95. Q-Q plot of P interpolated by the spectral model

The Q-Q plot (Figure 96) shows that pH is normally distributed except for values greater than 7 (left Q-Q plot). The right Q-Q plot shows that pH values between 5 and 6.5 are well estimated but pH values between 7 and 8.5 are sensibly under estimated.

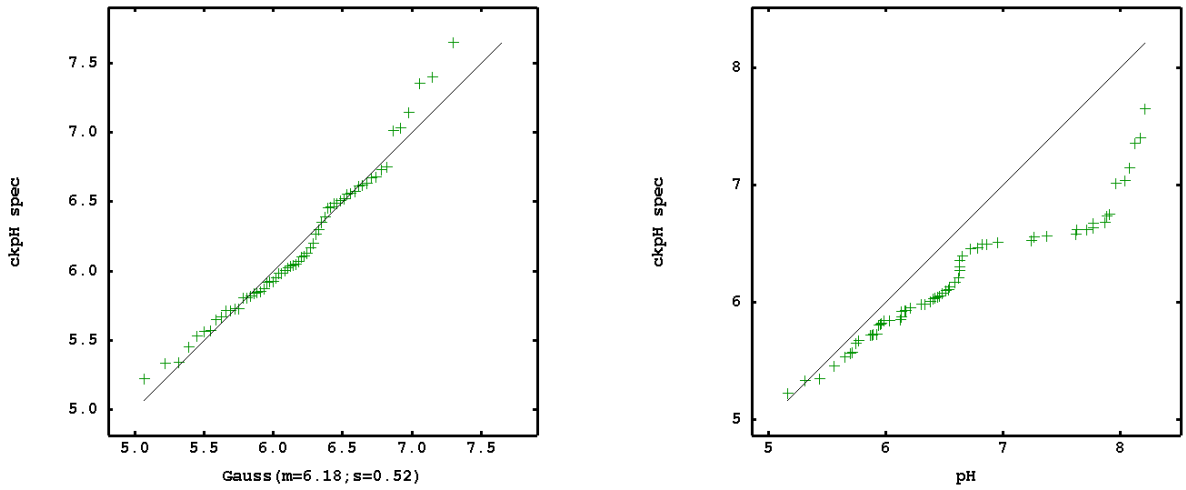

Figure 96. Q-Q plot of pH interpolated by the spectral model

Summing up we can state that the spectral model performs worse than the soil model.

6.2.3.1.3 Fusion model

The left Q-Q plot of Figure 97 shows that CEC follows normal distribution and CEC is also well estimated (right Q-Q plot)

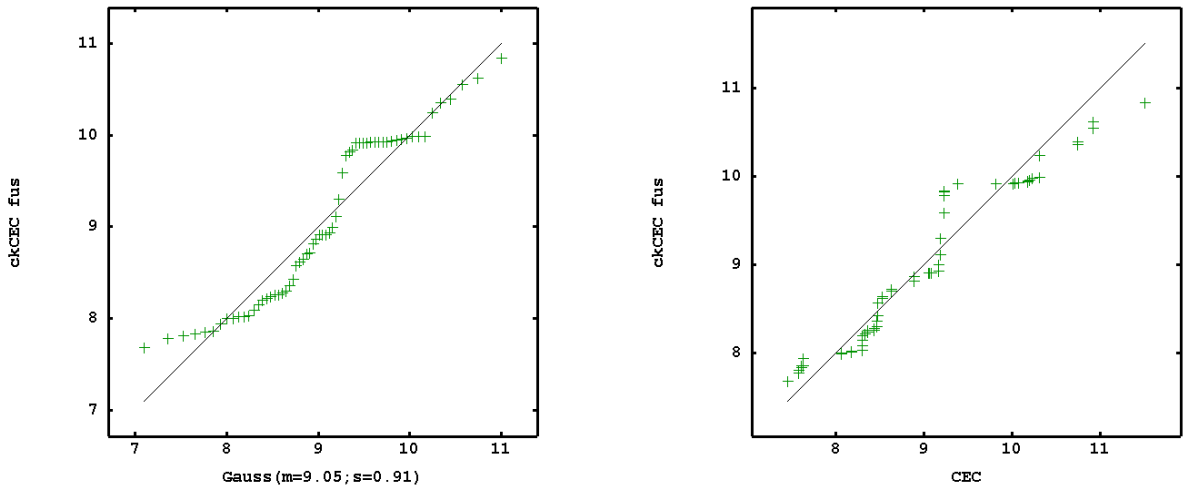

Figure 97. Q-Q plots of CEC interpolated by the fusion model

The left Q-Q plot (Figure 98) shows that exchangeable potassium is normally distributed except for values less than 0.10. Values of K are well estimated except for the larger values (>0.35) (right Q-Q plot).

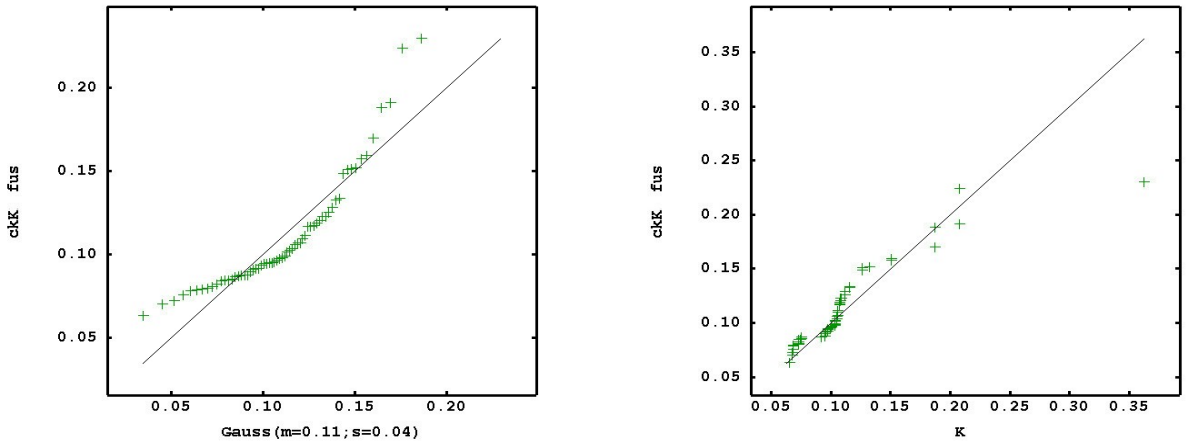

Figure 98. Q-Q plot of K interpolated by the fusion model

Q-Q plot of extractable phosphorus (Figure 99) shows that there is some departure from normal distribution (left Q-Q plot) for P values less than 0.10 or greater than 40. Moreover, P is slightly underestimated for values between 30 and 40 and overestimated for values greater than 40 (right Q-Q plot).

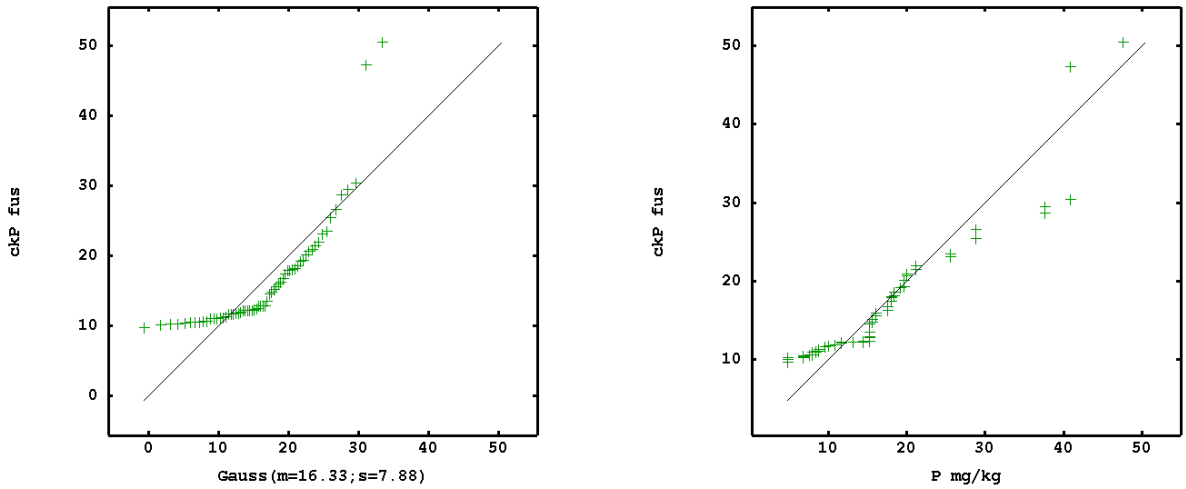

Figure 99. Q-Q plot of P interpolated by the fusion model

For pH, Figure 100 shows that its distribution can be assumed as normal (left Q-Q plot). The right Q-Q plot shows that pH values less than 5.7 are overestimated, whereas values greater than 7 are underestimated.

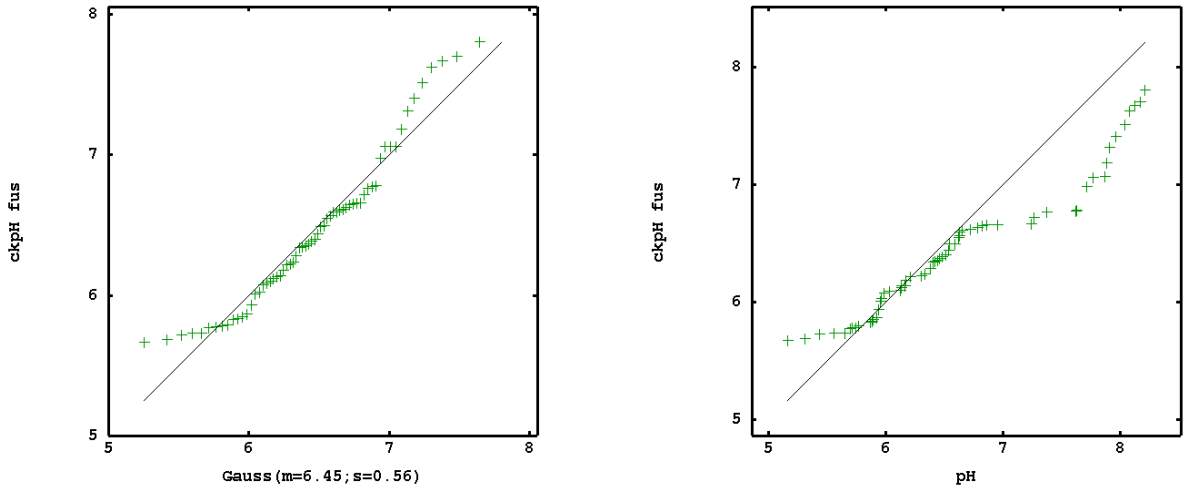

Figure 100. Q-Q plot of pH interpolated by the fusion model

For sand content, Figure 101 shows that the distribution can be assumed as normal (left Q-Q plot). The matching between estimation and observation is quite good (right Q-Q plot).

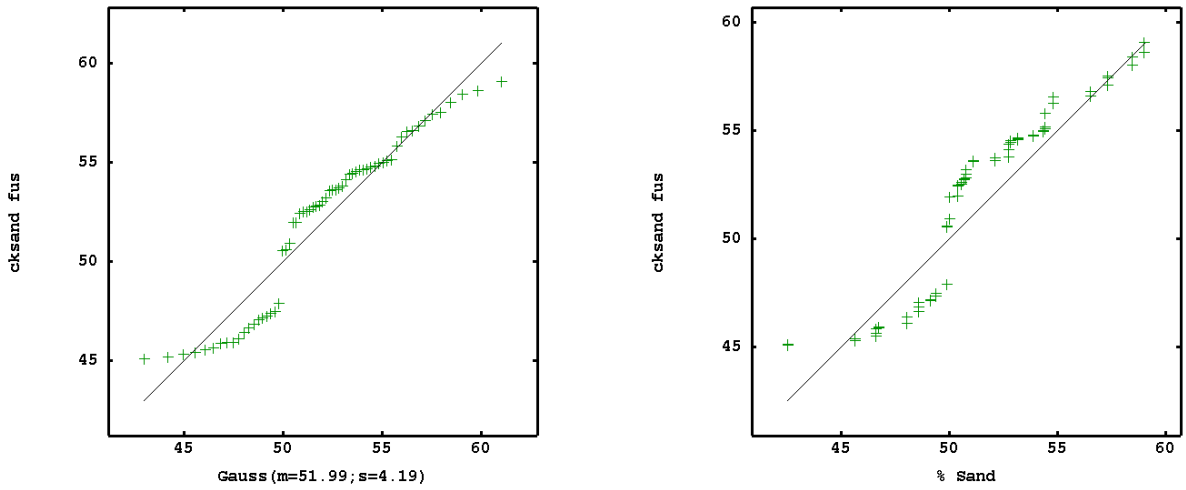

Figure 101. Q-Q plot of sand interpolated by the fusion model

Also for silt (Figure 102) the above considerations for sand hold.

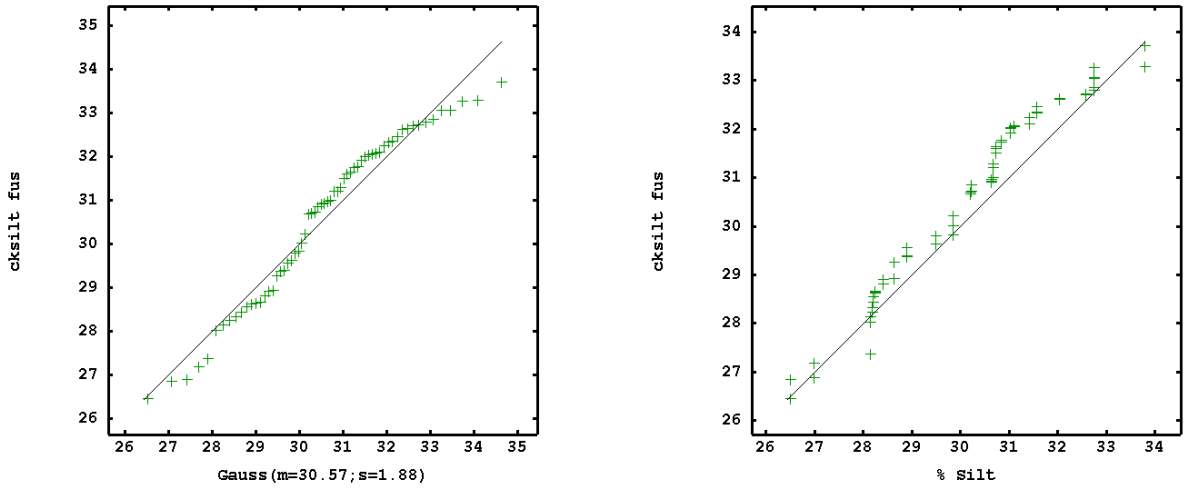

Figure 102. Q-Q plot of silt interpolated by the fusion model

For total carbon, the left Q-Q plot (Figure 103) shows that TC follows normal distribution. However, also for the fusion model there is the same mismatching between estimates and observations noticed in the spectral model (right Q-Q plot).

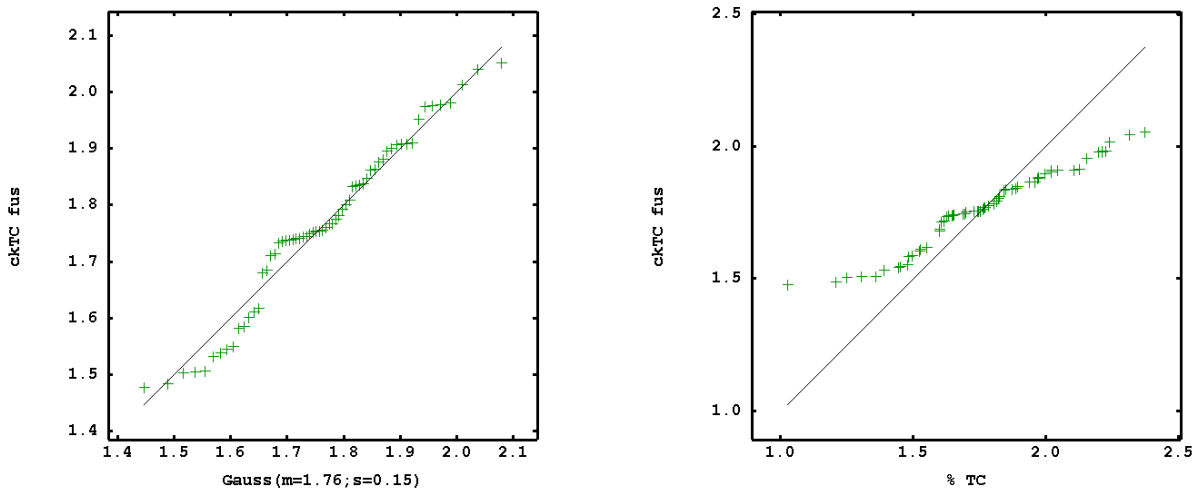

Figure 103. Q-Q plot of TC interpolated by the fusion model

The same considerations as the above hold also for TN (Figure 104).

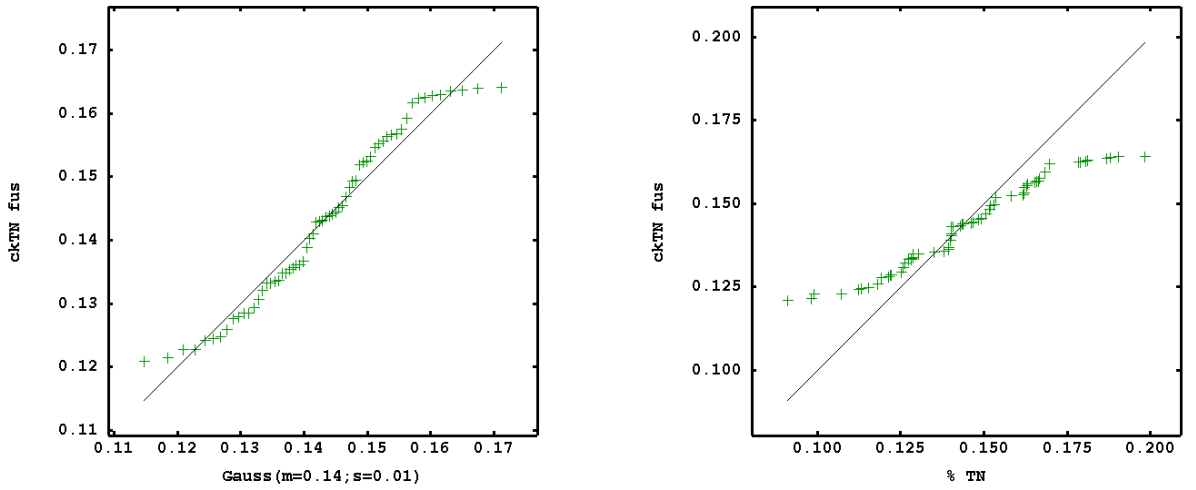

Figure 104. Q-Q plot of TN interpolated by the fusion model

Summing up, the soil and fusion models show similar performance and both perform better than the spectral model.

6.2.3.2 Box plots of differences

6.2.3.2.1 Soil model

All box plots for all soil properties estimated by the soil model show that the mean difference between true and estimated values is close to zero (Figure 105) but there is no evidence if mean differences between true and estimated values are statistically different from zero, so a t test of means with unequal variances is reported following the box plots section.

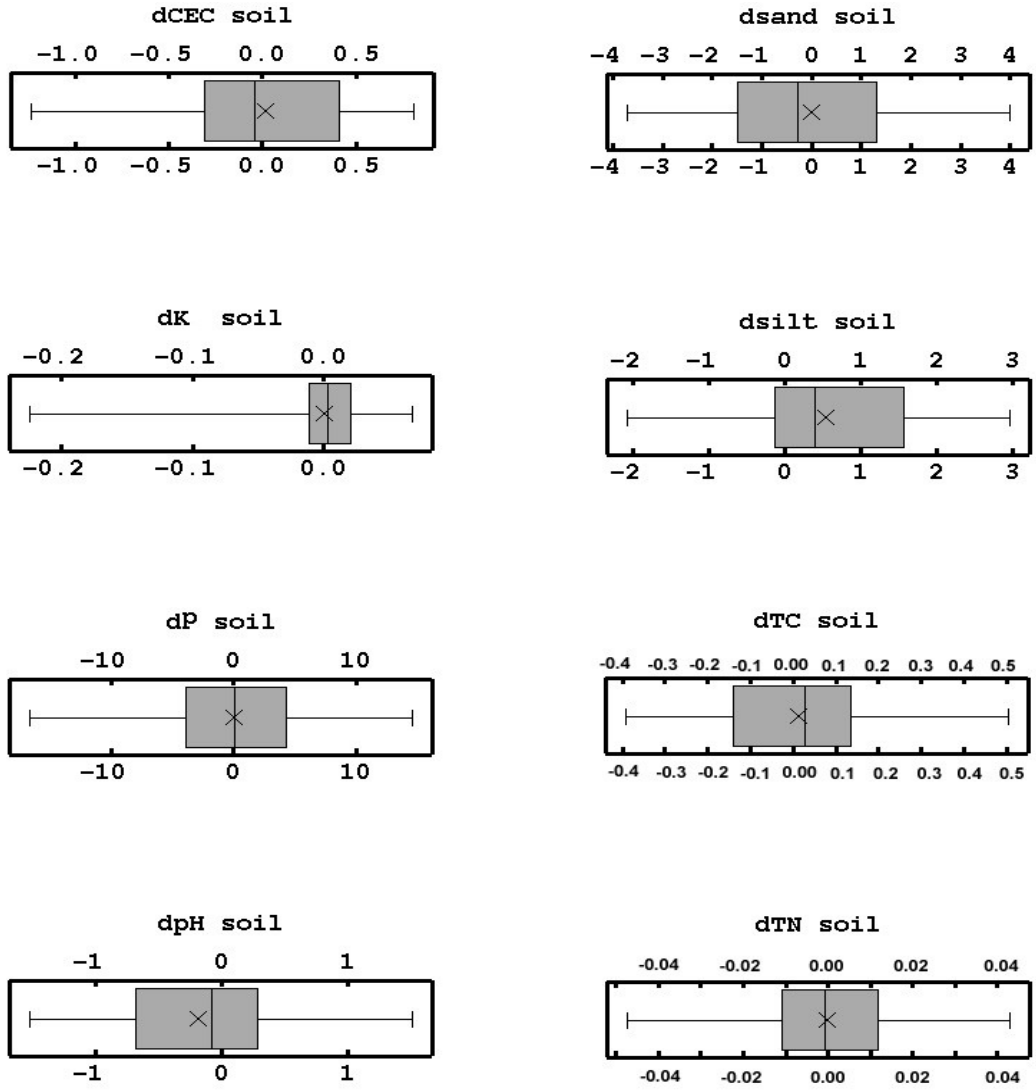

Figure 105. Box plots of differences between true and estimated values of the eight soil variables interpolated by the soil model

6.2.3.2.2 Spectral model

The mean differences between true and estimated values of soil properties using the spectral model are far from zero (Figure 106), which means that the spectral model was biased.

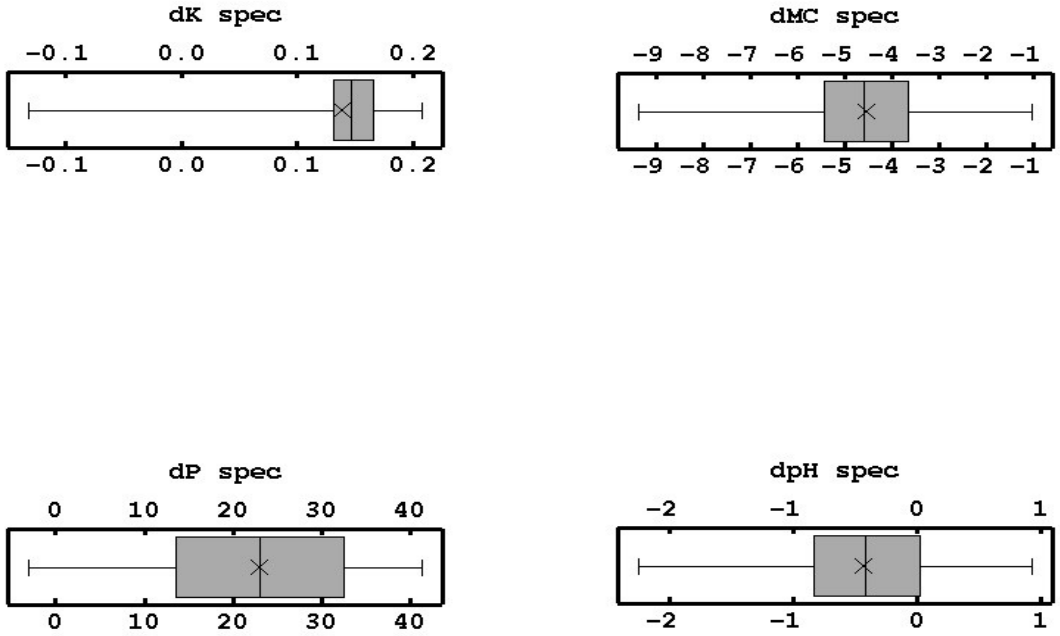

Figure 106. Box plots of differences between true and estimated values of the four soil variables interpolated by the spectral model

6.2.3.2.3 Fusion model

Box plots (Figure 107) of differences between true and estimated values of soil properties using the fusion model show that mean difference is close to zero for all the soil properties.

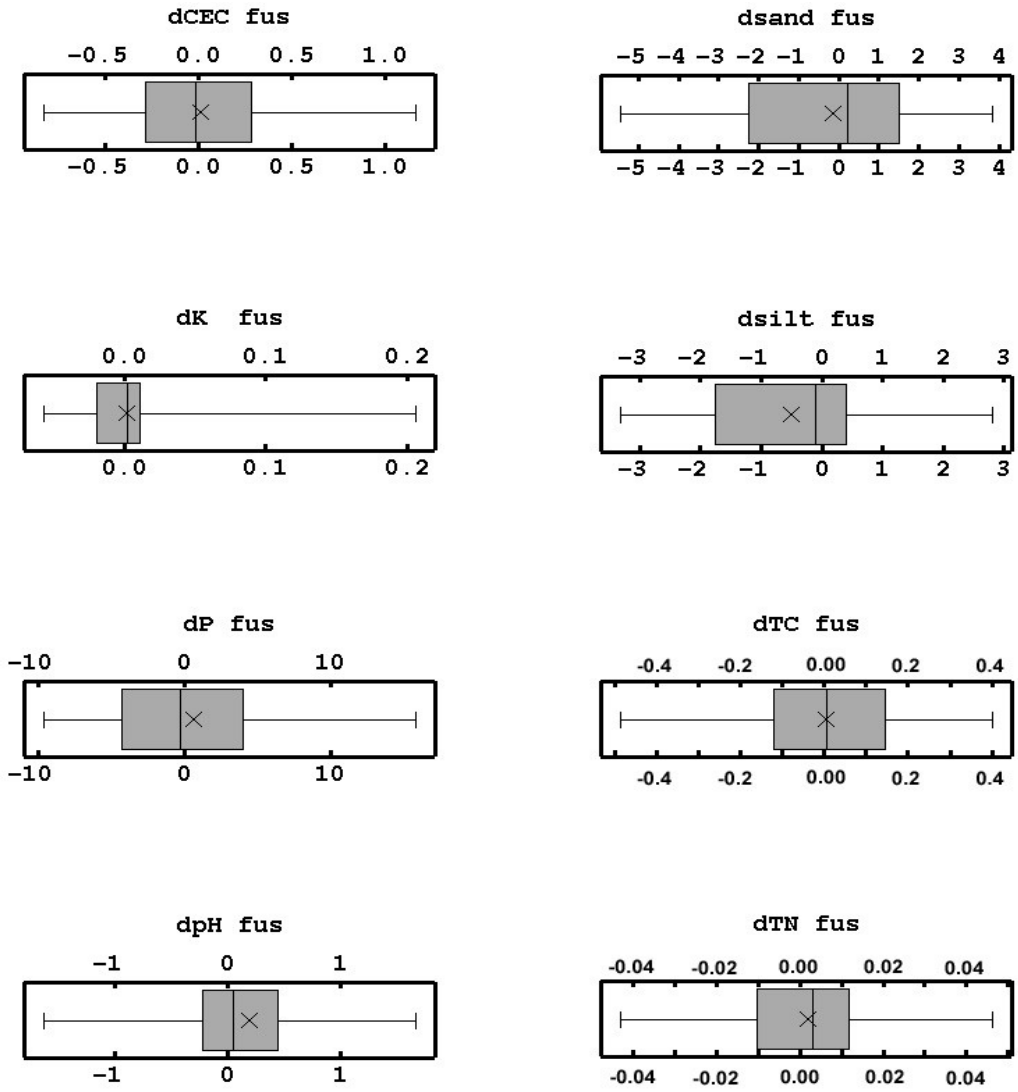

Figure 107. Box plots of differences between true and estimated values of the eight soil variables interpolated by the fusion model

6.2.3.3 t-test

A t-test assuming unequal variances was applied to check if the mean differences between true (validation data set) and estimated values are significantly different from zero.
The results of the t test with confidence level (α) equal to 0.05 for soil properties estimated by the soil model are shown in Table 17. For CEC, pH, exchangeable potassium, extractable phosphorus, sand content and silt content, the means are different, whereas means are the same for total carbon and total nitrogen.

Table 17. t-test results assuming unequal variances for testing mean equality of observations and estimates for soil model.

Variables	Soil model			Interpretation
	T value	*P* value	α	
TC	1.45	0.153	0.05	Means are the same
TN	0.80	0.428	0.05	Means are the same
CEC	4.29	0.000	0.05	Means are different
Sand	-6.32	0.000	0.05	Means are different
Silt	3.68	0.001	0.05	Means are different
pH	3.63	0.001	0.05	Means are different
P	-9.41	0.000	0.05	Means are different
K	-5.59	0.000	0.05	Means are different

Results of t test for soil properties estimated by the spectral model (Table 18) show that the means of the four soil spectral variables differ significantly from observations, which proves, as above, the poor performance of the spectral model.

Table 18. t-test results assuming unequal variances for testing mean equality of observations and estimates for the spectral model.

Variables	Spectral model			Interpretation
	T value	*P* value	α	
pH	3.04	0.003	0.05	Means are different
P	-16.74	0.000	0.05	Means are different
K	-17.05	0.000	0.05	Means are different
MC	21.43	0.000	0.05	Means are different

Similarly to the t test results for the soil model, only for TC and TN there are no significant differences between observation and estimation for the fusion model (Table 19).

Sameh Mohammed Shaddad, Proximal Soil Sensors and Geostatistical Tools in Precision Agriculture Applications, Tesi di dottorato in Scienze e Biotecnologie dei Sistemi Agrari e Forestali e delle Produzioni Alimentari, Indirizzo: Agrometeorologia ed Ecofisiologia dei Sistemi Agrari e Forestali, Università degli Studi di Sassari.

Table 19. t-test results assuming unequal variances testing mean equality of observations and estimates for the fusion model

Variables	Fusion model			Interpretation
	T value	*P* value	α	
TC	1.35	0.183	0.05	Means are the same
TN	0.38	0.703	0.05	Means are the same
CEC	4.18	0.000	0.05	Means are different
Sand	-6.4	0.000	0.05	Means are different
Silt	3.87	0.001	0.05	Means are different
pH	3.57	0.001	0.05	Means are different
P	-9.29	0.000	0.05	Means are different
K	-5.28	0.000	0.05	Means are different

6.2.4 Delineation of management zones

To synthesise the complex multivariate variation of the field in a restricted number of zones to be submitted to differential management, factor kriging analysis was applied to the estimate data sets of the three different models. In the following analysis we retained only the eigenvectors corresponding to eigenvalues greater than one and omitted the ones corresponding to nugget effect, because the latter are mostly affected by measurement errors.

Table 20 shows the decomposition of the soil model into regionalized factors. The loading values for the first factor indicate that TC, pH and TN weigh positively whereas silt weighs negatively, whereas sand content is the most influencing variable for the second factor. This means that the first factor is mostly related to soil fertility, whereas the second factor to soil texture. Figure 108 shows the F1 map split into three isofrequency classes (high, medium and low) that can be interpreted as MZs related to soil fertility. Figure 109 shows the F2 map split into three isofrequency classes (low, medium and high) related to soil texture and sand content in particular. The expected high fertility zones according to the first factor are found in the southern part and along the sides of the field. Low fertility zones are found in the central part of the field and in the northern part.

Table 20. Decomposition into regionalized factors of the soil model

	g[1]P	gTC	gTN	gCEC	gK	gpH	gSand	gSilt	Eigen Val.	Perc. Var.
Factor 1	0.21	0.54	0.35	0.28	-0.04	0.48	0.13	-0.47	2.44	41.59
Factor 2	0.36	-0.11	-0.19	-0.36	0.41	-0.22	0.53	-0.44	1.76	30.06

[1] g means gaussian

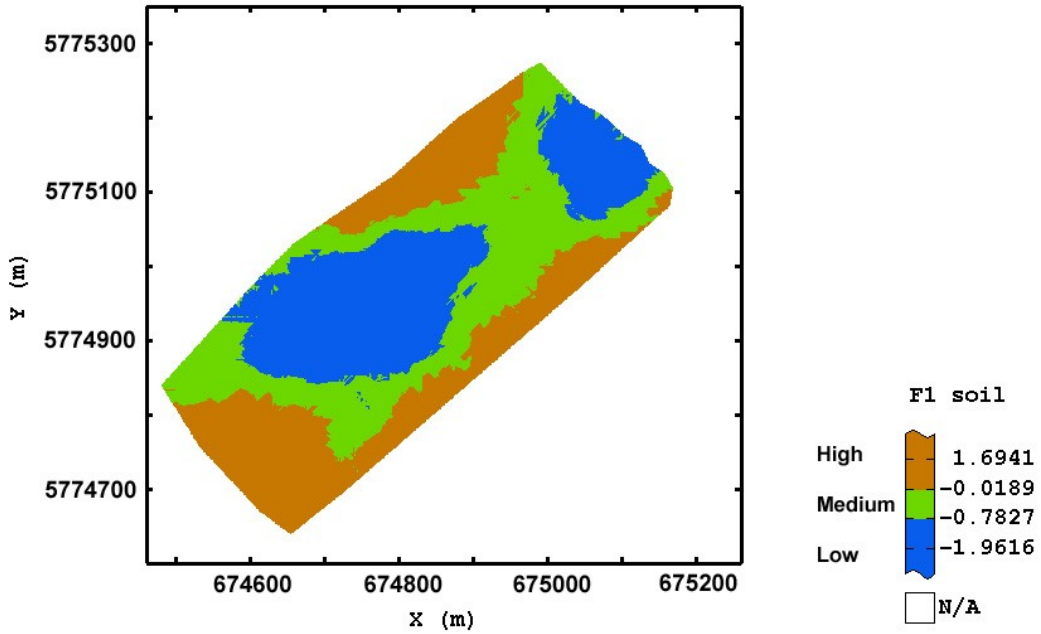

Figure 108. Maps of three management zones (MZ) according to the first factor of the soil model

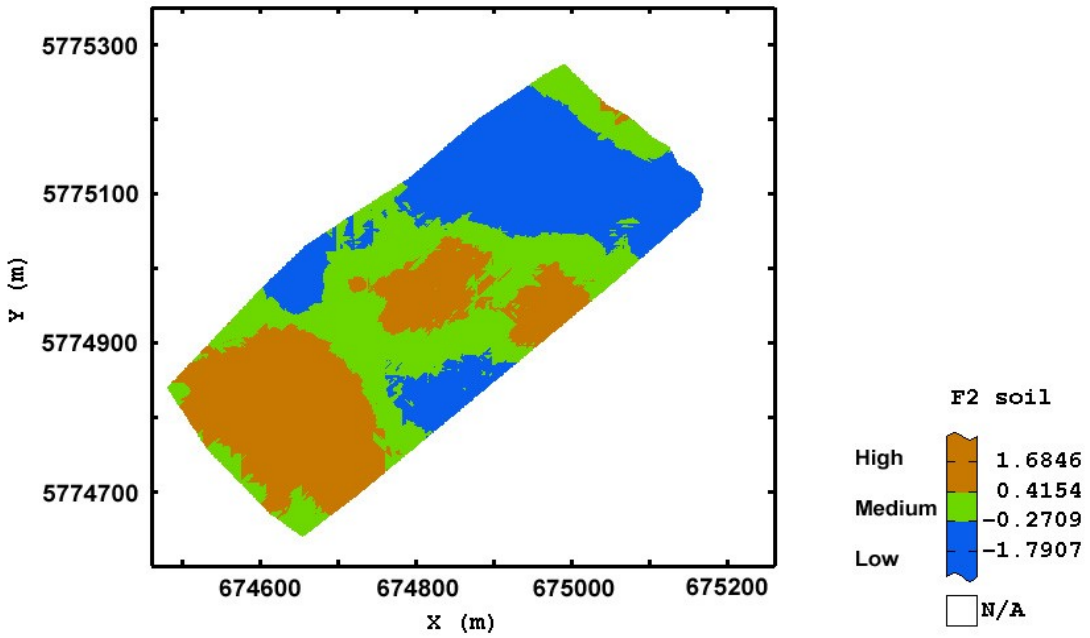

Figure 109. Maps of three management zones (MZ) according to the second factor of the soil model

Table 21 shows the decomposition into regionalized factors at longer scale (210 m) for the spectral model. No factor has an eigenvalue greater than 1, however we restricted our discussion only to the first factor which explains about 80% of the spatial variation at longer scale. The loading values indicate that the first factor is affected negatively by extractable phosphorus. Figure 110 Shows the MZ map for the first factor, where low values correspond to high values of P. However the map shows artefacts along the main direction of driving and the homogeneous areas are too narrow to actually be differentially managed by the farmer.

Table 21. Decomposition into regionalized factors of spectral model at longer range (210 m)

	gK spec[1]	gMC spec	gP spec	gpH spec	Eigen Val	Var. Perc.
Factor 1	0.19	0.08	-0.88	-0.42	0.41	79.89
Factor 2	0.86	0.39	0.07	0.31	0.06	12.03
Factor 3	-0.33	0.16	-0.45	0.82	0.04	8.08
Factor 4	0.33	-0.91	-0.12	0.24	0.00	0.00

[1] spec means spectral variable

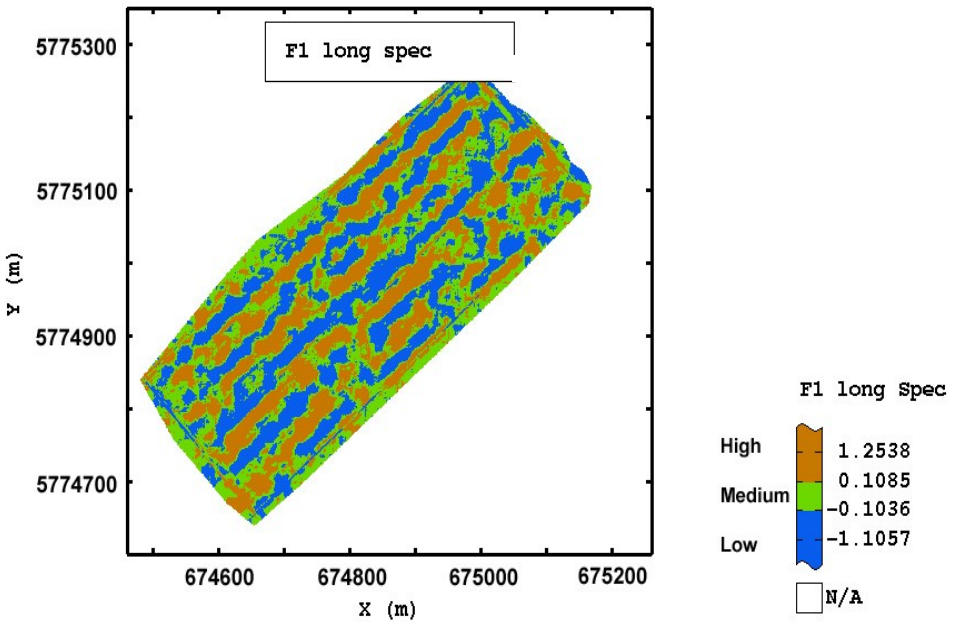

Figure 110. Maps of three management zones (MZ) according to the first factor of the spectral model

Table 22. Decomposition into regionalized factors of the fusion model

	gK spec	gMC spec	gP spec	gpH spec	gP	gTC	gTN	gCEC	gk	gpH	gSand	gSilt	Eigen Val.	Var. Perc.
Factor 1	0.02	0.15	-0.22	0.14	0.15	0.51	0.44	0.34	-0.08	0.51	-0.15	-0.17	2.48	45.81
Factor 2	0.21	0.02	-0.17	-0.17	0.35	0.13	-0.16	-0.21	0.25	0.00	0.55	-0.57	1.07	19.82

Table 22 shows the decomposition into regionalized factors of the fusion model. The same considerations made for the soil model hold here for the first two factors of the fusion model. The maps of the two factors (Figures 111-112) look quite similar to Figure 108 and Figure 109 with a slight increase of variability.

Summing up all the results, we can say that the fusion model, though including more variables, performs as well as the soil model.

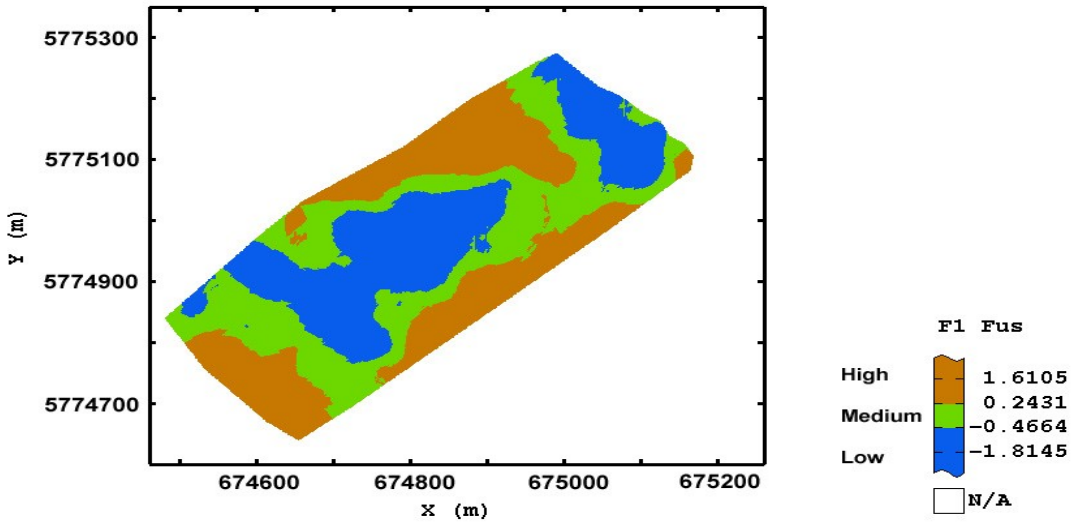

Figure 111. Maps of three management zones (MZ) according to the first factor of the fusion model

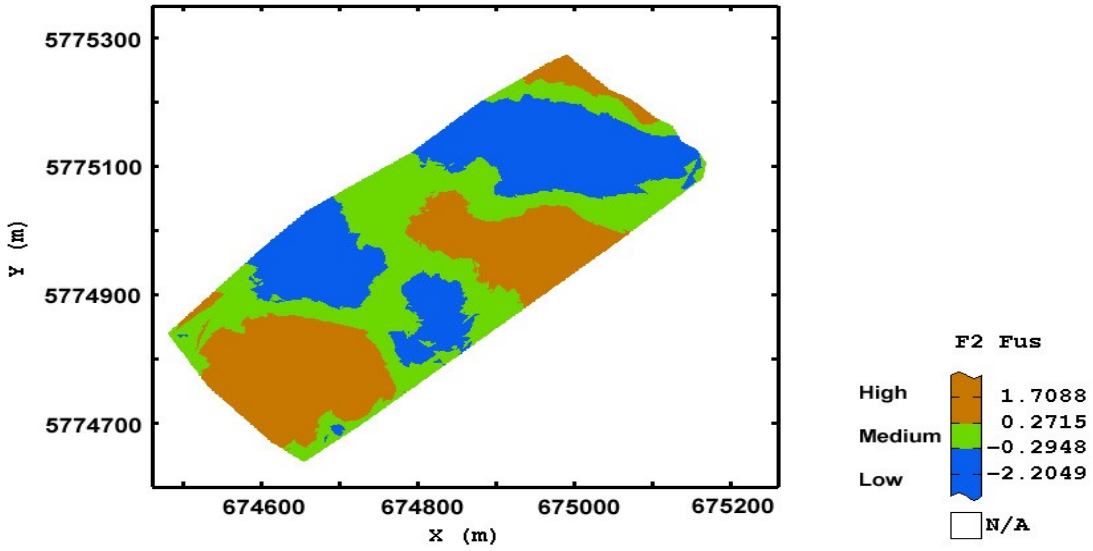

Figure 112. Maps of three management zones (MZ) according to the second factor of the fusion model

6.2.5 Comparison of the management zone with the yield map of barley

A yield map of barley (Figure 113) was produced by using the inverse distance weighting (IDW) method including 3 isofrequency classes of yield production. The first class represents the area of low production (2.59 – 5.73 t/ha), the second class represents the area of medium yield production (5.73 – 7.26 t/ha) and the third class represents the area of high production (7.26 – 8.00 t/ha).

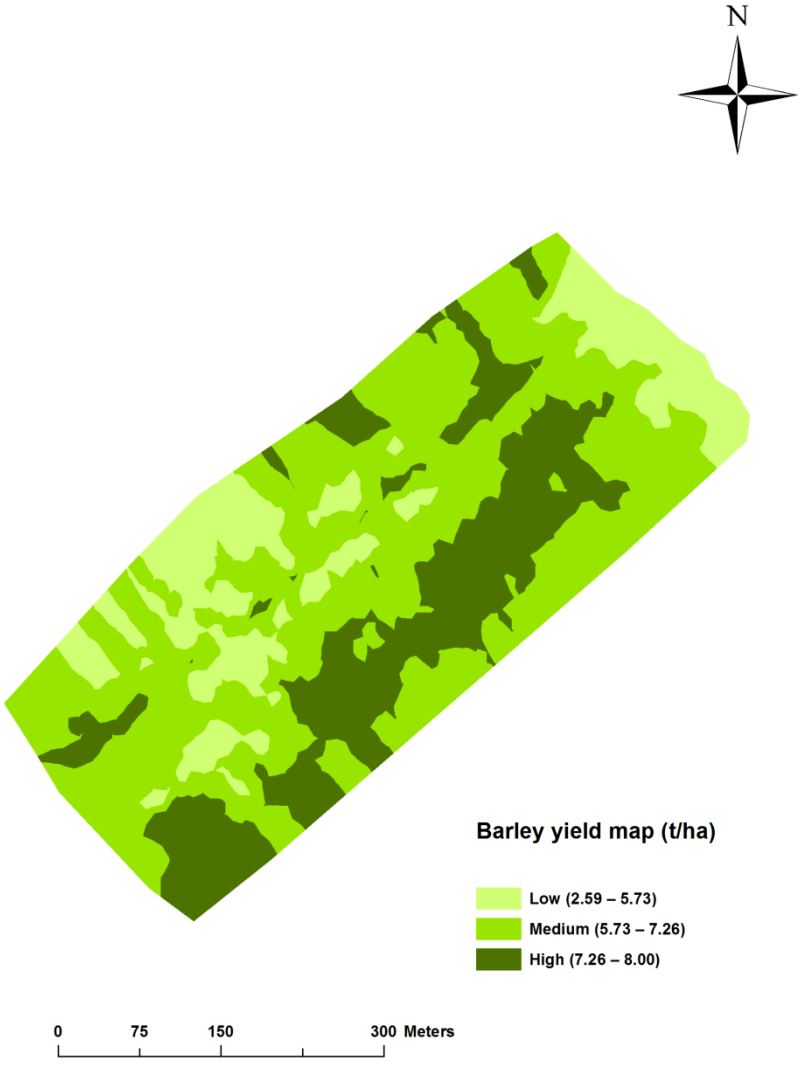

Figure 113. Barley yield map 2013

Table 23. Percentages of similarity between delineated management zones according to F1 of soil and fusion models and yield production of barley winter 2013

Yield t/ha	F1_Soil			F1_Fus		
	Low	Medium	High	Low	Medium	High
Low yield (2.59 - 5.73)	39%	15%	5%	33%	19%	6%
Medium yield (5.73 -7.26)	49%	49%	64%	55%	48%	59%
High yield (7.26 - 8.00)	13%	36%	32%	12%	33%	35%

The spatial association between the yield map and the delineation map for the soil and fusion models is reported in

Table 23. As F1 was interpreted as an indicator of soil fertility, we wanted to evaluate the production potential expressed by the field delineation obtained from the factor. The overall accordance between the two maps was 40.0 % for the soil model and 38.6 % for the fusion model. Again the two models perform quite similarly. The above percentages of similarity mean that the field delineation for the two models can explain only about 40% of the yield variability, whereas more than 50% is ascribable to more dynamic factors not included in this study. These dynamic factors include monitoring of agro-meteorological conditions, plant diseases and nutrition stresses which affect the yield as well as soil productivity. Such monitoring of the previous factors will help to better detect the variation in yield.

Table 24. Mean yield of each management zone for F1 of soil and fusion models

Classes	Mean yield[1] t/ha	
	F1_Soil	F1_Fus
Low	6.12	6.19
Medium	6.77	6.65
High	6.93	6.97

[1] Low (2.59 – 5.73), Medium (5.73 -7.26), High (7.26 – 8.00)

In table 24 the yield mean per each class of F1 for the soil and fusion models was reported. As it can be seen from table 24, the delineation shows a clear trend in yield variation which is consistent with the level of soil fertility. This results stress the utility of the proposed approach for field delineation in precision agriculture and VRT fertilization in particular.

7. Conclusion and future work

Research in precision agriculture needs to have all facilities such as proximal soil sensors that provide us with high spatial resolution information regarding physical, chemical and biological soil properties. Statistical tools are also necessary to investigate the variation nature of soil parameters, as geostatistics which studies spatial dependence of soil variables.

The ultimate goal of this study was to use vis-NIR spectroscopy and geostatistics to develop high spatial resolution maps of the soil properties under study and to delineate the management zones in a field in United Kingdom. To fulfil this goal, several steps have been planned. These included:

1. carrying out an on-the-go survey by a vis-NIR online sensor;
2. sampling the field for physical and chemical analyses;
3. carrying out a laboratory scanning of soil samples by the same vis-NIR sensor
4. developing different calibration models of some soil parameters using partial least square regression (PLSR) applied on the lab based spectra
5. predicting soil properties by using the on-line collected spectra (point 1) and the calibration models (point 4) and then create high resolution interpolated maps of the studied soil properties
6. developing three different geostatistical models: the model for the soil properties, the one for the spectral variables estimated through vis-NIR calibration models and the model for all variables obtained from fusion of soil variables with spectral variables)
7. spatially interpolating the soil variables of each model
8. applying factor cokriging analysis for the three geostatistical models to delineate the management zones (MZ)
9. performing a validation test for the three models and comparing them
10. assessing the potential of productivity prediction expressed by MZ delineation for each model through comparison of MZ map with the yield map of barley harvested after the soil survey.

7.1 Vis-NIR experiment

7.1.1 Cross-validation

The accuracy of the calibration models was tested by cross-validation and independent validation. Results of cross-validation showed that the quality of calibration models can be categorized according to RPD values as class A for extractable phosphorus and total carbon, and class B for pH, total nitrogen, moisture content, exchangeable potassium indicating good to excellent calibration models.

7.1.2 Independent validation

The independent validation of PLSR calibration models provided RPD and RMSEP values illustrating good to excellent calibration models with the exception of K.

The above results encourage us to use on-line vis-NIR sensor as one of the proximal soil sensors, because it can provide a lot of information about soil properties in a non- destructive, cost effective and easier way than the traditional methods.

7.2 Geostatistical analysis

Three different geostatistical models were developed based on measured soil variables (soil model), spectral variables (spectral model) and a fusion model (combining soil and spectral variables). They were developed through: Gaussian anamorphosis, fitting of linear model of coregionalization (LMC), ordinary cokriging to produce thematic spatial maps, and factor cokriging to delineate management zones. For the fusion model, multi-collocated cokriging and factor cokriging were performed using spectral k variable as auxiliary variable because it showed the strongest correlation with the soil variables.

7.2.1 Cross- validation

The goodness of geostatistical models was tested by cross-validation. Summing up the results of cross-validation for the soil model, the goodness of fitting can be considered as satisfactory, whereas the performance of the spectral model was quite poor. As regards the fusion model, it performed quite well, though the model generally underestimated the high values and overestimated the low values.

7.2.2 Independent validation

The performance of the three different models was also evaluated and the models were compared through an independent validation set. Three statistics were calculated: mean error (ME), as an indicator of bias; mean standardized squared error (MSSE), as an indicator of accuracy, and root mean squared error (RMSE), as an indicator of precision of estimation.

Synthetically, the two, soil and fusion, models performed quite similarly, whereas the performance of the spectral model was much poorer.

7.2.3 Thematic spatial maps

With regard to the spatial maps interpolated with the spectral model, they showed some artefacts along the main direction of driving, that were not evident in the maps obtained with the other two models. The soil property maps from the fusion model looked quite similar to the ones from the soil model, with an increased variability at short scale due to the introduction of the fine-surveyed spectral variables.

7.2.4 Delineation of management zones

To synthesise the complex multivariate variation of the field in a restricted number of zones to be submitted to differential management, the factor cokriging analysis was applied to the estimate data sets of the three different models. The first factor was related to soil properties that affect soil fertility. Based on the first factor, three management zones were delineated and assumed as low, medium and high fertility zones for the soil and fusion model.

7.2.5 Spatial similarity between MZs and yield map

A yield map of barley was interpolated by using the inverse distance weighting (IDW) method and was then classified into 3 isofrequency classes (low, medium, high) as for the MZ maps for the soil and fusion models produced by the first regionalised factor, which was interpreted as an indicator of soil fertility. The overall accordance between the two maps was 40.0 % for the soil model and 38.6 % for the fusion model. Again the two models performed quite similarly. These results can be interpreted as more than 50% of the yield variation was ascribable to more dynamic factors than soil parameters, such as agro-meteorological conditions, plant diseases, nutrition stresses, etc.

7.2.6 Future work

Though the above results are quite satisfactory and are encouraging regarding the use of the proposed approach in site-specific management, nevertheless the following improvements in methodology are suggested:

- Calibration of soil variables using vis-NIR spectra should take into account the influence of water content that is generally variable on the field;
- To avoid the error introduced by calibration with spectra, a way could be to use directly the significant latent variables in PLSR as auxiliary variables in the fusion model;
- For a more effective site-specific management, the static field delineation in MZ should be integrated by on-time information about soil or plant status through remote or proximal imaging.

8. References

Adamchuk, V. I., Lund, E. D., Sethuramasamyraja, B., Morgan, M. T., Dobermann, A., and Marx, D. B. (2005). Direct measurement of soil chemical properties on-the-go using ion-selective electrodes. Comput. Electron. Agric. 48, 272–294.

Aggelopooulou, K., Castrignanò, A., Gemtos, T., De Benedetto, D. (2013). Delineation of management zones in an apple orchard in Greece using a multivariate approach. Computers and Electronics in Agriculture 90, 119–130

Al Hagrey, S. A., and Müller, C. (2000). GPR study of pore water content and salinity in sand. Geophys. Prosp. 48, 63–85.

Al-Janobi, A. (2000). A data-acquisition system to monitor performance of fully mounted implements. J. Agric. Eng. Res. 75, 167–175.

Annan, A. P. (2002). GPR—History, trends, and future developments. Subsurf. Sens. Technol. Applic. 3, 253–270.

Arriola-Morales, J., Batlle-Sales, J., Valera, M. A., Linares, G., and Acevedo, O. (2009). Spatial variability analysis of soil salinity and alkalinity in an endorreic volcanic watershed. Int. J. Ecol. Dev. 14, 1–17.

Barrett, L. R. (2002). Spectrophotometric color measurement in situ in well drained sandy soils. Geoderma 108, 49–77.

Baumgardner, M. F., Silva, L. F., Biehl, L. L., and Stoner, E. R. (1985). Reflectance properties of soils. Adv. Agron. 38, 2–44.

Ben-Dor, E., and Banin, A. (1995). Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. Soil Sci. Soc. Am. J. 59, 364–372.

Ben-Dor, E., Irons, J. R., and Epema, G. F. (1999). Soil reflectance. In ‘‘Remote Sensing for the Earth Sciences: Manual of Remote Sensing’’ (A. N. Rencz, Ed.), Vol. 3, pp. 111–188. (3rd edn.). Wiley, New York.

Bergveld, P., Hendrikse, J., and Olthuis, W. (1998). Theory and application of the material work function for chemical on the field principle. Meas. Sci. Technol. 9, 1801–1808.

Birrell, S. J., and Hummel, J. W. (2000). Membrane selection and ISFET configuration evaluation for soil nitrate sensing. Trans. Am. Soc. Agric. Eng. 43, 197–206.

Birrell, S. J., and Hummel, J. W. (2001). Real-time multi ISFET/FIA soil analysis system with automatic sample extraction. Comput. Electron. Agric. 32, 45–67.

Bishop, J. L., Pieters, C. M., and Edwards, J. O. (1994). Infrared spectroscopic analyses on the nature of water in montmorillonite. Clays Clay Minerals 42, 702–716.

Bogrekci, I., and Lee, W. S. (2005). Improving phosphorus sensing by eliminating soil particle size effect in spectral measurement. Trans. ASABE 48, 1971–1978.

Bogrekci, I., and Lee, W. S. (2005a). Spectral soil signatures and sensing phosphorus. Biosyst. Eng. 92, 527–533.

Bokobza, L. (1998). Near infrared spectroscopy. J. Near Infrared Spectrosc. 6, 3–17.

Bowers, S. A., and Hanks, R. J. (1965). Reflection of radiant energy from soils. Soil Sci. 100, 130–138.

Bricklemyer, R. S., and Brown, D. J. (2010). On-the-go VisNIR: Potential and limitations for mapping soil clay and organic carbon. Comput. Electron. Agric. 70, 209–216.

Bricklemyer, R. S., Miller, P. R., Paustian, K., Keck, T., Nielsen, G. A., and Antle, J. M. (2005). Soil organic carbon variability and sampling optimization in Montana dry land wheat fields. J. Soil Water Conserv. 60, 42–51.

Brown, D. J., Shepherd, K. D., Walsh, M. G., Mays, M. D., and Reinsch, T. G. (2006). Global soil characterization withVNIRdiffuse reflectance spectroscopy. Geoderma 132, 273–290.

Bullock, D. S., and Bullock, D. G., (2000). From agronomic research to farm management guidelines: A primer on the economics of information and precision technology. Precis. Agric. 2, 71–101.

Burgess, T. M. and Webster, R., (1980). Optimal interpolation and isarithmic mapping of soil properties II. Block kriging. J. Soil Sci., 31: 333 – 341.

Castrignanò, A. (2011). Introduction to spatial data analysis. Aracne, Rome. ISBN 978-88-548-3978-6.

Castrignanò, A., De Benedetto, D., Girone, G., Guastaferro, F., Sollitto, D. (2010). Characterization, Delineation and Visualization of Agro-Ecozones Using Multivariate Geographical Clustering. Ital. J. Agron. / Riv. Agron., 2010, 5:121-132

Castrignanò, A., Costantini, E. A. C., Barbetti, R., Sollitto, D., (2009). Accounting for extensive topographic and pedologic secondary information to improve soil mapping. Catena 77. 28–38.

Castrignanò, A., Giugliarini, L., Risaliti, R., Martinelli, N., (2000). Study of spatial relationships among some soil physico-chemical properties of a field in central Italy using multivariate geostatistics. Geoderma 97, 39–60.

Castrignanò, A., Wong, M. T. F., Stelluti, M., De Benedetto, D., and Sollitto, D. (2012). Use of EMI, gamma-ray emission and GPS height as multi-sensor data for soil characterisation. Geoderma 175-176 (2012) 78–89.

Chang, C. W., Laird, D. A., Mausbach, M. J., and Hurburgh, C. R. (2001). Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. Soil Sci. Soc. Am. J. 65, 480–490.

Chang, G. W., Laird, D. A., and Hurburgh, G. R. (2005). Influence of soil moisture on near infrared reflectance spectroscopic measurement of soil properties. Soil Sci. 170, 244–255.

Chanzy, A., Tarussov, A., Judge, A., and Bonn, F. (1996). Soil water content determination using a digital ground-penetrating radar. Soil Sci. Soc. Am. J. 60, 1318–1326.

Chilès, J. P., Delfiner, P., (1999). Geostatistics: Modelling Spatial Uncertainty. Wiley, New York. 695 pp.

Christy, C. D. (2008). Real-time measurement of soil attributes using on-the-go near infrared reflectance spectroscopy. Comput. Electron. Agric. 61, 10–19.

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

Clark, R. N., King, T. V. V., Klejwa, M., Swayze, G. A., and Vergo, N. (1990). High spectral resolution reflectance spectroscopy of minerals. J. Geophys. Res. 95, 12653–12680.

Cohen, M. J., Prenger, J. P., and DeBusk, W. F. (2005). Visible-near infrared reflectance spectroscopy for rapid, nondestructive assessment of wetland soil quality. J. Environ. Qual. 34, 1422–1434.

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

Corwin, D. L., and Lesch, S. M., (2005). Characterizing soil spatial variability with apparent soil electrical conductivity: I. Survey protocols. Comput. Electron. Agric. 46, 103–133.

Corwin, D. L., and Rhoades, J. D. (1982). An improved technique for determining soil electrical conductivity-depth relations from above-ground electromagnetic measurements. Soil Sci. Soc. Am. J. 46, 517–520.

Cozzolino, D., and Moron, A. (2003). The potential of near-infrared reflectance spectroscopy to analyze soil chemical and physical characteristics. J. Agric. Sci. 140, 65–71.

Creutin, J. And Obled, C., (1982). Objective analysis and mapping techniques for rainfall fields. An objective comparison. Wat.Resour. Res., 18: 413 – 431.

Dahnke, W. C. (1971). Use of the nitrate specific ion electrode in soil testing. Commun. Soil Sci. Plant Anal. 2, 73–84.

Dalal, R. C., and Henry, R. J. (1986). Simultaneous determination of moisture, organic carbon, and total nitrogen by near infrared reflectance spectrophotometry. Soil Sci. Soc. Am. J. 50, 120–123.

Daniel, K. W., Tripathi, N. K., and Honda, K. (2003). Artificial neural network analysis of laboratory and in situ spectra for the estimation of macronutrients in soils of Lop Buri (Thailand). Aust. J. Soil Res. 41, 47–59.

Daniels, D. J. (2007). Ground penetrating radar. IET Radar, Sonar, Navigation and Avionics Series. The institution of Engineering and Technology, London, United Kingdom, Stevenage, United Kingdom, (726).

Davis, J. L., and Annan, A. P. (1989). Ground-penetrating radar for high-resolution mapping of soil and rock stratigraphy. Geophys. Prosp. 37, 531–551.

De Benedetto, D., Castrignanò, A., Rinaldi, M., Ruggieri, S., Santoro, F., Figorito, B., and Gualano, S. (2012). An approach for delineating homogeneous zones by using multi-sensor data, Geoderma, http://dx.doi.org/10.1016/j.geoderma.2012.08.028

De Jong, E., Ballantyne, A. K., Cameron, D. R., and Read, D. W. L. (1979). Measurement of Apparent Electrical Conductivity of Soils by an Electromagnetic Induction Probe to Aid Salinity Surveys. Soil Sci. Soc. Am. J. 43, 810–812.

Deutsch, C. V. (2002). Geostatistical Reservoir Modeling, Oxford University Press, 376 pages.

Diacono, M., De Benedettob, D., Castrignanò, A., Rubinoa, P., Vittib, C., Abdelrahmanc, H.M., Sollittob, D., Cocozzac, C., and Ventrella, D. (2013). A combined approach of geostatistics and geographical clustering for delineating homogeneous zones in a durum wheat field in organic farming. NJAS - Wageningen J. Life Sci.

Dickson, B. L., and Scott, K. M. (1997). Interpretation of aerial gamma-ray surveys Adding the geochemical factors. AGSO J. Aust. Geol. Geophys. 17, 187–200.

Domsch, H., and Giebel, A. (2004). Estimation of soil textural features from soil electrical conductivity recorded using the EM38. Precis. Agric. 5, 389–409.

Durrant-Whyte, H. (2001). Multi Sensor Data Fusion. Australian Centre for Field Robotics, Univ. Sydney. Version 1.2

Ehsani, M. R., Upadhyaya, S. K., Slaughter, D., Shafii, S., and Pelletier, M. (1999). A NIR technique for rapid determination of soil mineral nitrogen. Precision Agric. 1, 217–234.

Fidencio, P. H., Poppi, R. J., and de Andrade, J. C. (2002). Determination of organic matter in soils using radial basis function networks and near infrared spectroscopy. Anal. Chim. Acta 453, 125–134.

Fystro, G. (2002). The prediction of C and N content and their potential mineralisation in heterogeneous soil samples using Vis-NIR spectroscopy and comparative methods. Plant Soil. 246, 139–149.

Gebbers, R., and Adamchuk, V. I., (2010). Precision agriculture and food security. Sci. Mag. 327, 828–831.

Gill, W. R., and Vandenberg, G. E. (1967). Soil Dynamics in Tillage and Traction. USDA Agricultural Research Service. Agricultural Handbook. p. 316.

Goddu, R. F., and Delker, D. A. (1960). Spectra-structure correlations for the near-infrared region. Anal. Chem. 32, 140–141.

Godwin, R. J. (1975). An extended octagonal ring transducer for use in tillage studies. J. Agric. Eng. Res. 20, 347–352.

Goovaerts, P. (1997). Geostatistics for natural resources evaluation. Oxford University Press. New York, 483 pp.

Goovaerts, P. (1999). Geostatistics in soil science: state-of-the-art and perspectives. Geoderma 89 -1999. 1–45.

Groenigen, J. W., Mutters, C. S., Horwath, W. R., and Kessel, C. (2003). NIR and DRIFT-MIR spectrometry of soils for predicting soil and crop parameters in a flooded field. Plant Soil 250, 155–165.

Henderson, T. L., Baumgardner, M. F., Franzmeier, D. P., Stott, D. E., and Coster, D. C. (1992). High dimensional reflectance analysis of soil organic-matter. Soil Sci. Soc. Am. J. 56, 865–872.

Hendriks, P. H. G. M., Limburg, J., and De Meijer, R. J. (2001). Full-spectrum analysis of natural g-ray spectra. J. Environ. Radioact. 53, 365–380.

Herrero, J., Ba, A. A., and Aragüés, R. (2003). Soil salinity and its distribution determined by soil sampling and electromagnetic techniques. Soil Use Manag. 19, 119–126.

Hezarjaribi, A., and Sourell, H. (2007). Feasibility study of monitoring the total available water content using non-invasive electromagnetic induction-based and electrode-based soil electrical conductivity measurements. Irrig. Drain. 56, 53–65.

Huisman, J. A., Hubbard, S. S., Redman, J. D., and Annan, A. P. (2003). Measuring soil water content with ground penetrating radar: A review. Vadose Zone J. 2, 476–491.

Hunt, G. R. (1977). Spectral signatures of particulate minerals in visible and near-infrared. Trans. Am. Geophys. Union 58, 553.

Hunt, G. R., and Salisbury, J. W. (1970). Visible and near-infrared spectra of minerals and rocks. I. Silicate minerals. Mod. Geol. 1, 283–300.

Hyvönen, E., Turunen, P., Vanhanen, E., Arkimaa, H., and Sutinen, R. (2005). Airborne gamma-ray surveys in Finland. Geological Survey of Finland Special Paper, 39, pp. 119–134.

Isaaks, E.H. and Srivastava, R.M. (1989). An Introduction to Applied Geostatistics. Oxford University Press, 561 pages.

Jenny, H., (1980). The soil resource: origin and behavior. Ecological Studies vol.37, Springer-Verlag, New York. 377 pp.

Jol, H. M. (2009). Ground Penetrating Radar: Theory and Applications. Elsevier Science & Technology, University of Wisconsin, Eau Claire, USA. 524.

Journel, A. G., Huijbregts, C. J., (1978). Mining Geostatistics. Academic Press, London. 600 pp.

Jung, W. K., Kitchen, N. R., Sudduth, K. A., Kremer, R. J., and Motavalli, P. P. (2005). Relationship of apparent soil electrical conductivity to claypan soil properties. Soil Sci. Soc. Am. J. 69, 883–892.

Khosla, R., Westfall, D. G., Reich, R. M., Mahal, J. S., and Gangloff, W. J. (2010). Spatial Variation and Site-Specific Management Zones. In M.A. Oliver (ed.), Geostatistical Applications for Precision Agriculture (pp. 195 – 219). Reading, United Kingdom.

Kim, H. J., Hummel, J. W., and Birrell, S. J. (2006). Evaluation of nitrate and potassium ionselective membranes for soil macronutrient sensing. Trans. ASABE 49, 597–606.

Kitchen, N. R., Sudduth, K. A., and Drummond, S. T. (1996). Mapping of sand deposition from 1993 midwest floods with electromagnetic induction measurements. J. Soil Water Conserv. 51, 336–340.

Kodaira, M., and Shibusawa, S., (2012) Using a mobile real-time soil visible-near infrared sensor for high resolution soil property mapping, Geoderma. http://dx.doi.org/10.1016/j.geoderma.2012.09.007

Kooistra, L., Wehrens, R., Leuven, R. S. E. W., and Buydens, L. M. C. (2001). Possibilities of visible-near-infrared spectroscopy for the assessment of soil contamination in river floodplains. Anal. Chim. Acta 446, 97–105.

Koolen, A. J., and Kuipers, H. (1983). Agricultural Soil Mechanics. Springer-Verlag, Berlin.

Krischenko, V. P., Samokhvalov, S. G., Fomina, L. G., and Novikova, G. A. (1992). Use of infrared spectroscopy for the determination of some properties of soil. In ‘‘Making Light Work: Advances in near Infrared Spectroscopy’’ (I. Murray and L. A. Cowe, Eds.), pp. 239–249 (Aberdeen, Scotland).

Krishnan, P., Alexander, D. J., Butler, B., and Hummel, J. W. (1980). Reflectance technique for predicting soil organic matter. Soil Sci. Soc. Am. J. 44, 1282–1285.

Kuang, B., and Mouazen, A. M. (2011). Calibration of a visible and near infrared spectroscopy for soil analysis at field scales across three European farms. Eur. J. Soil Sci. 62, 629–636.

Kuang, B., Mahmood, S. H., Quraishi, Z. M., Hoogmoed, B. W., Mouazen, A. M., and Eldert J. van Henten, (2012). Sensing Soil Properties in the Laboratory, In Situ, and On-Line: A Review. In Donald Sparks, editor: Advances in Agronomy, Vol. 114, Burlington: Academic Press, pp. 155-223. ISBN: 978-0-12-394275-3.

Laslett, G. M., McBrathney, A. B., Pahl, P. J. and Hutchison, M. F., (1987). Comparison of several spatial methods for soil pH. J. Soil Sci., 38: 325 – 341.

Liu, W. D., Baret, F., Gu, X. F., Tong, Q. X., Zheng, L. F., and Zhang, B. (2002). Relating soil surface moisture to reflectance. Remote Sens. Environ. 81, 238–246.

Mahmood, H. S., Hoogmoed, W. B., and Van Henten, E. J. (2011). Estimating soil properties with a proximal gamma-ray spectrometer using windows and full-spectrum analysis methods. In "Proceedings of the Second Global Workshop on Proximal Soil Sensing" (V. I. Adamchuk and R. A. Viscarra Rossel, Eds.), pp. 132–135. Montreal, Quebec, Canada.

Maleki, M. R., Van Holm, L., Ramon, H., Merckx, R., De Baerdemaeker, J., and Mouazen, A. M. (2006). Phosphorus sensing for fresh soils using visible and near infrared spectroscopy. Biosyst. Eng. 95, 425–436.

Malley, D. F., and Williams, P. C. (1997). Use of near-infrared reflectance spectroscopy in prediction of heavy metals in freshwater sediment by their association with organic matter. Environ. Sci. Technol. 31, 3461–3467.

Malley, D. F., Martin, P. D., McClintock, L. M., Yesmin, L., Eilers, R. G., and Haluschak, P. (2000). Feasibility of analysing archived Canadian prairie agricultural soils by near infrared reflectance spectroscopy. In ''Near Infrared Spectroscopy: Proceedings of the 9th International Conference'' (A. M. C. Davies and R. Giangiacomo, Eds.), pp. 579–585. NIR Publications, Chichester, UK.

Martens, H., and Naes, T. (1989). Multivariate Calibration John Wiley & Sons, Chichester, UK. 419 pp.

Martin, P. D., Malley, D. F., Manning, G., and Fuller, L. (2002). Determination of soil organic carbon and nitrogen at the field level using near-infrared spectroscopy. Can. J. Soil. Sci. 82, 413–422.

Matheron, G., (1973). The intrinsic random functions and their applications. Adv. Appl. Probab., 5: 239 – 465.

McBratney, A. B., MendonVa Santos, M. L., and Minasny, B., (2003). On digital soil mapping. Geoderma 117, 3–52.

McKyes, E. (1989). Agricultural Engineering Soil Mechanics. Elsevier Sciences, Amsterdam.

McLeod, M. K., Slavich, P. G., Irhas, Y., Moore, N., Rachman, A., Ali, N., Iskandar, T., Hunt, C., and Caniago, C. (2010). Soil salinity in Aceh after the December 2004 Indian Ocean tsunami. Agric. Water Manag. 97, 605–613.

McNeill, J. D. (1992). Rapid, accurate mapping of soil salinity by electromagnetic ground conductivity meters. In "Advances in measurement of soil physical properties: bringing theory into practice", pp. 209–229. ASA, CSSA, USA, SSSA, Madison, WI.

Melendez-Pastor, I., Navarro-Pedreño, J., Gómez, I., and Koch, M. (2008). Identifying optimal spectral bands to assess soil properties with VNIR radiometry in semi-arid soils. Geoderma 147, 126–132.

Minet, J., Agung, W., Patrick, B., Marnik, V., and Sébastien, L. (2010) Mapping shallow soil moisture profiles at the field scale using full-waveform inversion of ground penetrating radar data. In: *Geoderma*, Vol. 161, p. 225-237. doi:10.1016/j.geoderma.2010.12.023.

Moron, A., and Cozzolino, D. (2003). Exploring the use of near infrared reflectance spectroscopy to study physical properties and microelements in soils. J. Near Infrared Spectrosc. 11, 145–154.

Morra, M. J., Hall, M. H., and Freeborn, L. L. (1991). Carbon and nitrogen analysis of soil fractions using near-infrared reflectance spectroscopy. Soil Sci. Soc. Am. J. 55, 288–291.

Morris, R. V., Lauer, H. V., Lawson, C. A., Gibson, E. K., Nace, G. A., and Stewart, C. (1985). Spectral and other physicochemical properties of submicron powders of hematite (Alpha-Fe_2O_3), maghemite (Gamma-Fe_2O_3), magnetite (Fe_3O_4), goethite (Alpha- Feooh), and lepidocrocite (Gamma-Feooh). J. Geophys. Res. 90, 3126–3144.

Mortimore, J. L., Marshall, L. J. R., Almond, M. J., Hollins, P., and Matthews, W. (2004). Analysis of red and yellow ochre samples from Clearwell Caves and Catalhoyuk by vibrational spectroscopy and other techniques. Spectrochim. Acta A Mol. Biomol. Spectrosc. 60, 1179–1188.

Mouazen, A. M. (2009). The future of for on-line measurement of soil properties with sensor fusion. Landwards. Nelson Publishing, Natural Resources Department, Cranfield University, MK43 OAL, United Kingdom, 14–16.

Mouazen, A. M., and Ramon, H. (2006). Development of on-line measurement system of bulk density based on on-line measured draught, depth and soil moisture content. Soil Till. Res. 86, 218–229.

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

Mouazen, A. M., De Baerdemaeker, J., and Ramon, H. (2006). Effect of wavelength range on the measurement accuracy of some selected soil constituents using visual-near infrared spectroscopy. J. Near Infrared Spectrosc. 14, 189–199.

Mouazen, A. M., Dumont, K., Maertens, K., and Ramon, H. (2003). Two-dimensional prediction of spatial variation in topsoil compaction of a sandy loam field-based on measured horizontal force of compaction sensor, cutting depth and moisture content. Soil Till. Res. 74, 91–102.

Mouazen, A. M., Karoui, R., De Baerdemaeker, J., and Ramon, H. (2006a). Characterization of soil water content using measured visible and near infrared spectra. Soil Sci. Soc. Am. J. 70, 1295–1302.

Mouazen, A. M., Kuang, B., De Baerdemaeker, J., and Ramon, H. (2010). Comparison between principal component, partial least squares and artificial neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. Geoderma. doi:10.1016/J. Geoderma. 2010.03.001.

Mouazen, A. M., Maleki, M. R., Cockx, L., Van Meirvenne, M., Van Holm, L. H. J., Merckx, R., de Baerdemaeker, J., and Ramon, H. (2009). Optimum three-point linkage set up for improving the quality of soil spectra and the accuracy of soil phosphorus measured using an on-line visible and near infrared sensor. Soil Tillage Res. 103, 144–152.

Mouazen, A. M., Maleki, M. R., De Baerdemaeker, J., and Ramon, H. (2007). On-line measurement of some selected soil properties using a VIS-NIR sensor. Soil Till. Res. 93, 13–27.

Mouazen, A. M., Ramon, H., and Baerdemaeker, J. D. (2002). Effects of bulk density and moisture content on selected mechanical properties of sandy loam soil. Biosyst. Eng. 83, 217–224.

Mouazen, A. M. (2006). Soil Survey Device. International publication published under the patent cooperation treaty (PCT). World Intellectual Property Organization, International Bureau. International Publication Number: WO2006/015463; PCT/BE2005/000129; IPC: G01N21/00; G01N21/00.

Murphy, J. and Riley, J. P. (1962). A modified single solution method for determination of phosphate in natural waters. Annals Chem. Acta 27: 31 – 36.

National Research Council (1997). Precision Agriculture in the 21st Century. Wadhington DC, USA: National Academy Press.

Olea, R. A., (1999). Geostatistics for Engineers and Earth Scientists, Kluwer Academic Publishers, 303 pages.

Oliver, M. A. (Ed.) (2010). Geostatistical Applications for Precision Agriculture. Springer, Dordrecht. ISBN 978-90-481-9132-1. DOI 10.1007/978-90-481-9133-8 Springer Dordrecht Heidelberg London New York.

Olsen, S. R., Cole, C. V., Watanabe, F. C., and Dean, L. S. (1954). Estimation of available phosphorus in soil by extraction with sodium bicarbonate. U.S. Dept. Agric. Circ. 939.

Pereira, A. G., Gómez, A. H., and He, Y. (2004). Near-infrared spectroscopy potential to predict N, P, K and OM content in a loamy mixed soil and its combination with precision farming. In ‘‘CIGR International Conference Beijing,’’ Beijing, China.

Petersen, H., Fleige, H., Rabbel, W., and Horn, R. (2005). Applicability of geophysical prospecting methods for mapping of soil compaction and variability of soil texture on farm land. J. Plant Nutr. Soil 168, 68–79.

Pierce, F. J., and Nowak, P., (1999). Aspects of Precision Agriculture. Adv. Agron. 67, 1–85.

Piikki, K., M. Söderström., B. Stenberg., (2013). Sensor data fusion for topsoil clay mapping, Geoderma 199, 106–116.

Piper, C. S. (1950). Soil and plant analysis. Interscience. Publ. Inc. New York.

Post, J. L., and Noble, P. N. (1993). The near-infrared combination band frequencies of dioctahedral smectites, micas, and illites. Clays Clay Minerals 41, 639–644.

Pracilio, G., Adams, M. L., and Harper, R. J. (2005). Soil properties mapped from gammaray spectrometry. In "9th Annual Symposium on Precision Agriculture Research & Application in Australasia," Nedlands WA, Australia.

Pracilio, G., Adams, M. L., Smettem, K. R. J., and Harper, R. J. (2006). Determination of spatial distribution patterns of clay and plant available potassium contents in surface soils at the farm scale using high resolution gamma ray spectrometry. Plant Soil 282, 67–82.

Pungor, E. (1998). The Theory of Ion-Selective Electrodes. Anal. Sci. 14, 249–256.

Pungor, E. (1999). The new theory of Ion-Selective Electrodes. Crit. Rev. Anal. Chem. 29, 111–120.

Pungor, E. (2001). The new theory of ion-selective electrodes. Sensors 1, 1–12.

Reeves, J. B. III, and McCarty, G. W. (2001). Quantitative analysis of agricultural soils using near infrared reflectance spectroscopy and a fiber-optic probe. J. Near Infrared Spectrosc. 9, 25–34.

Reeves, J. B. III, McCarty, G. W., and Reeves, V. B. (2001). Mid-infrared diffuse reflectance spectroscopy for the quantitative analysis of agricultural soils. J. Agric. Food Chem. 49, 766–772.

Rhoades, J. D., and Corwin, D. L. (1981). Determining soil electrical conductivity-depth relations using an inductive electromagnetic soil conductivity meter. Soil Sci. Soc. Am. J. 45, 255–260.

Rhoades, J. D., Chanduvi, F., and Lesch, S. (1999). Soil salinity assessment: Methods and interpretation of electrical conductivity measurements. FAO irrigation and drainage paper; 57. FAO, Rome.

Richards, T. (2000). Development of a System for Mapping the Performance of Agricultural Field Operations (Unpublished EngD Thesis). Institue of AgriTechnology, UK.

Rivoirard, J. (2001). Which models for collocated cokriging. Mathematical Geology 33, 117–131.

Robert, P. C. (1999). Precision agriculture: research needs and status in the USA. In: J. V. Stafford (Ed.), Precision Agriculture ’99 (pp. 19–33). Sheffield, UK: Sheffield Academic Press.

Roberts, L. M., Wilford, J. R., Field, J. B., and Greene, R. S. B. (2003). High-Resolution Ground Based Gamma-Ray Spectrometry and Electromagnetics to Assess Regolith Properties."Advances in Regolith", Boorowa, NSW, pp. 352–357.

Saey, T., Van Meirvenne, M., Vermeersch, H., Ameloot, N., and Cockx, L. (2009). A pedotransfer function to evaluate the soil profile textural heterogeneity using proximally sensed apparent electrical conductivity. Geoderma 150, 389–395.

Saeys, W., Mouazen, A. M., Ramon, H. (2005). Potential for on-site and on-line analysis of hog manure using visual and near-infrared reflectance spectroscopy. Biosyst. Eng. 91 (4), 393–402.

Schafer, R. L., Young, S. C., Hendrick, J. G., and Johnson, C. E. (1984). Control concepts for tillage systems. *Soil & Tillage Research*, *4*, 313–320.

Schueller, J. K. (1997). Technology for precision agriculture. In: J. V. Stafford (Ed.), Precision Agriculture ’97 (pp. 19–33). Oxford, UK: BIOS Scientific Publishers.

Schuller, H. (1969). Die CAL-Methode, eine neue Methode zur Bestimmung des Pflanzenverfugbaren Phosphates in Böden. Z Pflanzenkr Bodenkunde 123, 48–63.

Schulze, D. G. (2002). An introduction to soil mineralogy. In ‘‘Soil Mineralogy with Environmental Applications’’ (J. B. Dixon and D. G. Schulze, Eds.), pp. 1–35. Soil Science Society of America Inc., Madison, WI.

Shaddad, S. M., Kuang, B., Madrau, S., and Mouazen, A. M. (2013). On-line measurement of soil extractable P and pH with visible and near infrared spectroscopy. 3rd Global Workshop on Proximal Soil Sensing. Potsdam. Germany.

Shepherd, K. D., and Walsh, M. G. (2002). Development of reflectance spectral libraries for characterization of soil properties. Soil Sci. Soc. Am. J. 66, 988–998.

Sherman, D. M., and Waite, T. D. (1985). Electronic spectra of Fe^{3+} oxides and oxyhydroxides in the near infrared to ultraviolet. Am. Mineral. 70, 1262–1269.

Shibusawa, S., Made Anom, S. W., Sato, H. P., and Sasao, A. (2001). Soil mapping using the real-time soil spectrometer. In ‘‘ECPA 2001’’ (G. Gerenier and S. Blackmore, Eds.), Vol. 2, pp. 485–490. agro Montpellier, Montpellier, France.

Shonk, J. L., Gaultney, L. D., Schulze, D. G., and Scoyoc, G. E. V. (1991). Spectroscopic sensing of soil organic matter content. Trans. ASAE 34, 1978–1984.

Sinfield, J. V., Fagerman, D., and Colic, O. (2010). Evaluation of sensing technologies for on-the-go detection of macro-nutrients in cultivated soils. Comput. Electron. Agric. 70, 1–18.

Smith, M. C., Vellidis, G., Thomas, D. L., and Breve, M. A. (1992). Measurement of water table fluctuations in a sandy soil using ground penetrating radar. Trans. Am. Soc. Agric. Eng. 35, 1161–1166.

Stenberg, B. (2010). Effects of soil sample pretreatments and standardised rewetting as interacted with sand classes on Vis-NIR predictions of clay and soil organic carbon. Geoderma. doi:10.1016/j.geoderma.2010.04.008.

Stenberg, B., Viscarra Rossel, R. A., Mouazen, M. A., and Wetterlind, J. (2010). Visible and Near Infrared Spectroscopy in Soil Science. Adv. Agron. 107, 163–215.

Sudduth, K. A., Kitchen, N. R., Wiebold, W. J., Batchelor, W. D., Bollero, G. A., Bullock, D. G., Clay, D. E., Palm, H. L., Pierce, F. J., Schuler, R. T., and Thelen, K. D. (2005). Relating apparent electrical conductivity to soil properties across the north-central USA. Comput. Electron. Agric. 46, 263–283.

Taylor, M. J., Smettem, K. R. J., Pracilio, G., and Verboom, W. H. (2002). Relationships between soil properties and high-resolution radiometrics, central eastern Wheatbelt, Western Australia. Explor. Geophys. 33, 95–102.

Triantafilis, J., Ahmed, M. F., and Odeh, I. O. A. (2002). Application of a mobile electromagnetic sensing system (MESS) to assess cause and management of soil salinization in an irrigated cotton-growing field. Soil Use Manag. 18, 330–339.

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

Van den Bosch, I., Lambot, S., Druyts, P. ; Huynen, I., Acheroy, M. (2006). Buried target signature extraction from ground-penetrating radar signal: measurements and simulations. In: Near Surface Geophysics, Vol. 4, no. 1, p. 31-38.

Van Egmond, F. M., Loonstra, E. H., and Limburg, J. (2010). Gamma ray sensor for topsoil mapping: The mole. In "Proximal Soil Sensing" (R. A. V. Rossel, A. B. McBratney, and B. Minasny, Eds.), Progress in Soil Science, pp. 323–332. Springer Science and Business Media B.V. 2010, Dordrecht, Heidelberg, London, New York.

Van Overmeeren, R. A., Sariowan, S. V., and Gehrels, J. C. (1997). Ground penetrating radar for determining volumetric soil water content; results of comparative measurements at two test sites. J. Hydrol. 197, 316–338.

Viscarra Rossel, R. A. (2007). Robust modelling of soil diffuse reflectance spectra by ''bagging-partial least squares regression''. J. Near Infrared Spectrosc. 15, 39–47.

Viscarra Rossel, R. A., & McBratney, A. B. (1997). Preliminary experiments towards the evaluation of a suitable soil sensor for continuous 'on-the-go' field pH measurements. In J. V. Stafford (Ed.), *Precision agriculture'97* (pp. 493–501). Oxford, UK: BIOS Scientific Publishers.

Viscarra Rossel, R. A., and McBratney, A. B. (1998). Soil chemical analytical accuracy and costs: implications from precision agriculture. Australian Journal of Experimental Agriculture, 38, 765–775.

Viscarra Rossel, R. A., and Behrens, T. (2010). Using data mining to model and interpret soil diffuse reflectance spectra. Geoderma. doi: 10.1016/j.geoderma.2009.12.025.

Viscarra Rossel, R. A., and Lark, R. M. (2009). Improved analysis and modelling of soil diffuse reflectance spectra using wavelets. Eur. J. Soil Sci. 60, 453–464.

Viscarra Rossel, R. A., and McBratney, A. B., Minasny, B. (2010). Proximal soil sensing. ISBN 978-90-481-8858-1.DOI 10.1007/978-90-481-8859-8 Springer Dordrecht Heidelberg London New York.

Viscarra Rossel, R. A., Taylor, H. J., and McBratney, A. B. (2007). Multivariate calibration of hyperspectral g-ray energy spectra for proximal soil sensing. Eur. J. Soil Sci. 58, 343–353.

Viscarra Rossel, R. A., Walvoort, D. J. J., McBratney, A. B., Janik, L. J., and Skjemstad, J. O. (2006). Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131, 59–75.

Wackernagel, H., (1988). Geostatistical techniques for interpreting multivariate spatial information. In: Quantative analysis of mineral and energy resources. Chung, C. F., Fabbri, A. G., and Sinding-Larsen, R. (eds). Reidal, Dordrecht, The Netherlands, 393 – 409.

Wackernagel, H., (2003). Multivariate Geostatistics: An Introduction with Applications, 3rd ed. Springer Verlag, Berlin. 388 pp.

Weiler, K. W., Steenhuis, T. S., Boll, J., and Kung, K. J. S. (1998). Comparison of ground penetrating radar and time-domain reflectometry as soil water sensors. Soil Sci. Soc. Am. J. 62, 1237–1239.

Wetterlind, J., Stenberg, B., and Söderström, M. (2010). Increased sample point density in farm soil mapping by local calibration of visible and near infrared prediction models. Geoderma 156, 152–160.

Wilford, J. (2002). Airborne gamma-ray spectrometry. Cooperative Research Centre for Landscape Environments and Mineral Exploration, Commonwealth Scientific and Indusctrial Research Organization, Bertley, WA, Australia. Open File Rep. 144, 46–52.

Wilford, J. R., Bierwirth, P. N., and Craig, M. A. (1997). Application of airborne gamma-ray spectrometry in soil/regolith mapping and applied geomorphology. AGSO J. Aust. Geo. Geophy. 17, 201–216.

Wilford, J., and Minty, B. (2006). The use of airborne gamma-ray imagery for mapping soils and understanding landscape processes. In "Digital Soil Mapping: An Introductory Perspective" (P. Lagacherie, A. B. McBratney, and M. Voltz, Eds.), pp. 207–218. Elsevier B.V. Amsterdam, Oxford.

Williams, P. C., and Norris, K. (2001). Variable affecting near-infrared spectroscopic analysis. In: Near-Infrared Technology in Agricultural and Food Industries (Williams P C; Norris K, eds), 2nd Ed, p 295. American Association of Cereal Chemists, Inc. St. Paul Minnesota, USA.

Williams, B. G., and Baker, G. C. (1982). An electromagnetic induction technique for reconnaissance surveys of soil salinity hazards. Aust. J. Soil Res. 20, 107–118.

Williams, B. G., and Hoey, D. (1987). The use of electromagnetic induction to detect the spatial variability of the salt and clay contents of soils. Aust. J.Soil Res. 25, 21–27.

Williams, P., (2003). Near-infrared Technology—Getting the Best out of Light. PDK Projects, Nanaimo, Canada.

Wong, M. T. F., Asseng, S., Robertson, M. J., Oliver, Y., (2008). Mapping subsoil acidity and shallow soil across a field with information from yield maps, geophysical sensing and the grower. Precision Agriculture 9, 3–15.

Wong, M.T.F., Harper, R.J., (1999). Use of on-ground gamma-ray spectrometry to measure plant-available potassium and other topsoil attributes. Australian Journal of Soil Research 37, 267–277.

Wong, M.T.F., Oliver, Y.M., Robertson, M.J., (2009). Gamma-radiometric assessment of soil depth across a landscape not measurable using electromagnetic surveys. Soil Science Society of America Journal 73, 1261–1267.

Yan, L., Zhoub, S., Feng, L., Hong-Yi, L. (2007). Delineation of site-specific management zones using fuzzy clustering analysis in a coastal saline land. Computers and Electronics in Agriculture 56, 174–186

Yang, H., Kuang, B., Mouazen, A.M., (2012). Quantitative analysis of soil nitrogen and carbon at a farm scale using visible and near infrared spectroscopy coupled with wavelength reduction. European Journal of Soil Science, 63(3): 410-420.

Yitagesu, F. A., Van der Meer, F. D., and Van der Werff, H., (2009). Quantifying engineering parameters of expansive soil from their reflectance spectra: Engineering Geology, v. 105, p. 151 – 160.

Zornoza, R., Guerrero, C., Mataix-Solera, G., Scow, K. M., Arcenegui, V., and Mataix-Beneyto, J. (2008). Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. Soil Biol. Biochem. 40–7, 1923–1930.