



Ghent University
Faculty of Bioscience Engineering
Department of Soil Management

High resolution soil inventory using a dual signal electromagnetic induction sensor

Liesbet Cockx

Thesis submitted in fulfillment of the requirements for the degree
of Doctor (PhD) in Applied Biological Sciences
Academic Year 2009-2010

Promotor	Prof. dr. ir. Marc Van Meirvenne Department of Soil Management, Faculty of Bioscience Engineering, Ghent University
Dean	Prof. dr. ir. Guido Van Huylenbroeck
Rector	Prof. dr. Paul Van Cauwenberge

LIESBET COCKX

HIGH RESOLUTION SOIL INVENTORY USING A DUAL
SIGNAL ELECTROMAGNETIC INDUCTION SENSOR

Thesis submitted in fulfillment of the requirements for the degree of
Doctor (PhD) in Applied Biological Sciences: Land and Forest
Management

Dutch translation of the title:

Hoge resolutie bodeminventarisatie met behulp van het tweevoudig signaal van een elektromagnetische inductie sensor

Cover: Photograph of an EM38DD mobile survey on an agricultural field in Watervliet (photo: U.W.A. Vitharana).

Reference:

L. Cockx. *High resolution soil inventory using a dual signal electromagnetic induction sensor*. PhD thesis, Ghent University, February 2010.

ISBN-number:

The author and the promotor give the authorization to consult and to copy parts of this work for personal use only. Every other use is subject to the copyright laws. Permission to reproduce any material contained in this work should be obtained from the author.

Dankwoord

Het behalen van een doctoraat doe je niet alleen. Allereerst zou ik mijn appreciatie voor mijn promotor Prof. Marc Van Meirvenne willen uiten. De afgelopen zes jaar heeft hij mij zowat elke opportuniteit laten benutten en stond hij steeds klaar met het nodige advies. Het voornaamste bewijs hiervan is mijn verworven wetenschappelijke kennis, met als kers op de taart dit doctoraat. Op wetenschappelijk vlak bezit hij een niet-te-evenaren enthousiasme waar hij al zijn werknemers in onderdompelt en telkens opnieuw uitdagingen voorschotelt. Ook op persoonlijk vlak waren deze jaren een verrijking. Ik zou Marc dan ook willen bedanken voor zijn vertrouwen in mijn doen en laten.

De ORBit-ers verdienen hier ook een eervolle vermelding. Udaya, not only to be my sparring partner in the office, also for the pleasure of knowing you and your warm family. Meklit for being the coolest lady I know. Hans en Valentijn voor hun assistentie. David, Timothy en Philippe voor de collegiale samenwerking. Eef en Islam for sharing in the stress and easing the final efforts.

De sfeer op de vakgroep zorgde ook steeds voor aangename pauzes. In het bijzonder zou ik Sofie willen bedanken voor de middaguitjes, Tina voor het delen van tweeling-lief (en leed) en Anita voor haar oprechte interesse.

Verder wil ook volgende (ex-)collega's bedanken: Lieven en Frieke om mij te introduceren in de neurale netwerk wereld, Gunther en Irene om mij te betrekken in hun kwartaire geologie passie, Jan Valckx en Bruno De Vos voor de aangename samenwerking en Sam om al mijn computervragen geduldig te beantwoorden.

Voor de nodige ontspanning op het thuisfront kon ik steeds rekenen op een steunende (schoon)familie: merci iedereen voor de helpende handen, luisterende oren en achter kindjes-lopende-benen! Ook ben ik supercontent met de vrienden die me omringen: voor de deugddoende sport-uitdagingen (voor zover er geen buik in de weg zat), de lekkere etentjes, de weekeindjes weg, de Minart & Marimain-avondjes, tout court voor alle gezellige momenten samen.

Eén persoon verdient hier speciale aandacht en dat is natuurlijk Jan, mijn eeuwige soulmate. Hij was de ideale persoon om dit werk tot een goed eind te brengen: Latex-expert, uitstekende wetenschapper maar ook - en vooral - de beste vader voor onze drie schatjes. Jan, uit heel mijn hart: bedankt!

*Gent, februari 2010
Liesbet Cockx*

Table of Contents

Dankwoord	i
Table of Contents	iii
List of Figures	vii
List of Tables	xi
List of Abbreviations and Acronyms	xiii
Samenvatting	xvii
Summary	xxi
1 Introduction	1-1
1.1 Framing soil sensors within a soil inventory context	1-2
1.2 Aim and research questions	1-4
1.3 Outline of the research	1-5
2 Quantifying soil variability	2-1
2.1 Soil variability: the scale issue	2-2
2.2 Quantifying soil variability	2-3
2.2.1 Direct information	2-4
2.2.2 Sensed information	2-5
2.2.2.1 Remote sensing	2-6
2.2.2.2 Proximal soil sensors	2-7
2.3 Factors influencing the soil apparent electrical conductivity	2-16
2.4 The EM38DD soil sensor	2-17
2.4.1 Background on the EM38 sensor	2-17
2.4.2 Introducing the EM38DD soil sensor	2-20
2.4.3 Calibration of an EM38DD	2-21
2.4.4 Mobile EM38DD soil survey	2-23
3 Analyzing soil sensor data	3-1
3.1 Geostatistical methods: from data points to a map	3-2
3.1.1 Introduction	3-2

3.1.2	Variography: modeling spatial variation	3-2
3.1.3	Ordinary kriging	3-5
3.1.4	Validation of the prediction	3-7
3.2	Classification and map comparison	3-8
3.2.1	Fuzzy- <i>k</i> -means classification	3-8
3.2.2	Comparing categorical maps	3-11
4	Characterizing soil nitrate dynamics using the EM38DD soil sensor	4-1
4.1	Introduction	4-2
4.2	Material and Methods	4-3
4.2.1	Study site	4-3
4.2.2	Data collection	4-3
4.2.3	Mapping EC _a	4-5
4.2.4	Mapping NO ₃ ⁻ -N	4-6
4.2.5	Validation	4-6
4.3	Results and Discussion	4-7
4.4	Conclusions	4-15
5	Explaining the within-field variability of earthworm populations using the EM38DD	5-1
5.1	Introduction	5-2
5.2	Material and Methods	5-3
5.2.1	Study Site	5-3
5.2.2	Data collection	5-3
5.2.2.1	Earthworm sampling	5-3
5.2.2.2	EC _a measurements	5-4
5.2.2.3	Soil sampling	5-5
5.2.3	Data analysis	5-5
5.2.4	Measures of (dis)association	5-6
5.3	Results	5-9
5.3.1	Earthworm variability	5-9
5.3.2	EC _a variability	5-11
5.3.3	Measures of (dis)association	5-13
5.3.3.1	Abiotic-biotic comparisons	5-13
5.3.3.2	Community level comparisons	5-15
5.3.3.3	Population level comparisons	5-15
5.4	Discussion	5-15
5.5	Conclusion	5-18
6	Prospecting quaternary frost-wedges and their polygonal network with the EM38DD soil sensor	6-1
6.1	Introduction	6-2
6.2	Material and Methods	6-3
6.2.1	Study site	6-3
6.2.2	Methods	6-4

6.2.2.1	EM38DD survey	6-4
6.2.2.2	Validation of the polygonal network	6-5
6.2.2.3	Accuracy of EM38DD based frost-wedge allocation	6-6
6.3	Results and Discussion	6-6
6.3.1	Soil texture and EC _a measurements	6-6
6.3.2	Predictive quality of the classified PR map	6-10
6.4	Conclusion	6-12
7	Locating clay in a sandy forest soil using the EM38DD soil sensor	7-1
7.1	Introduction	7-2
7.2	Material and Methods	7-3
7.2.1	Study site	7-3
7.2.2	EC _a survey	7-4
7.2.3	Soil sampling	7-5
7.2.4	Interpolation techniques	7-5
7.2.5	Classification and delination methods	7-6
7.2.6	Accuracy measurements	7-7
7.3	Results and Discussion	7-9
7.3.1	EC _a measurements	7-9
7.3.2	Clay content and EC _a variability	7-11
7.3.3	Delineation of the clay lens	7-13
7.3.4	Validation	7-15
7.3.5	Depth of the clay lens	7-17
7.4	Conclusions	7-18
8	Analyzing high resolution soil sensor data to optimize texture predictions	8-1
8.1	Introduction	8-2
8.2	Material and Methods	8-3
8.2.1	Site Description	8-3
8.2.2	Electromagnetic Induction: Survey and Processing	8-5
8.2.3	Soil Data Set	8-5
8.2.4	Artificial Neural Networks: A Brief Introduction	8-6
8.2.5	Datasets, Network Architecture and Data Preparation	8-8
8.2.6	Validation Procedure	8-9
8.2.7	Multiple Linear Regression	8-10
8.3	Results and Discussion	8-11
8.3.1	EC _a and Texture	8-11
8.3.2	Artificial Neural Network Analysis	8-14
8.3.3	Comparison with SMLR	8-17
8.3.4	Mapping the topsoil clay content	8-18
8.4	Conclusions	8-19

9	General conclusions and future research	9-1
9.1	Introduction	9-2
9.2	General conclusions	9-2
9.2.1	Answers to the research questions	9-2
9.2.2	Impact of conclusions	9-8
9.3	Future perspectives	9-9
	Bibliography	Bib-1
	Curriculum vitae	

List of Figures

1.1	The EM38DD soil sensor.	1-3
1.2	Outline of the research.	1-5
2.1	Hypothesized model of spatial variation of soil properties, adapted from Benayas et al. (2004).	2-2
2.2	The electromagnetic spectrum.	2-7
2.3	(a) Basic principle of galvanic electrical resistivity measurements and (b) illustration of the MUCEP system.	2-8
2.4	(a) Basic principle of electromagnetic induction and (b) illustration of the mobile DUALEM-21S sensor configuration with below a photograph of the DUALEM-21S sensor outside its sled.	2-10
2.5	Schematic of the Geometrics, Inc. Ohmmapper CCR system (Walker and Houser, 2002).	2-12
2.6	Examples of (a) the VERIS VIS-NIR sensor and (b) a GPR	2-13
2.7	Principle of a mechanical soil sensor	2-14
2.8	Electrical conductivity pathways: (1) solid-liquid, (2) liquid and (3) solid. Modified from Rhoades et al. (1989).	2-16
2.9	Schematic diagram of the EM38 showing the principle of electromagnetic induction. Adapted from Robinson et al. (2004).	2-18
2.10	The response curves as a function of the depth for the two orientations of the EM38 sensor: (a) the relative response and (b) the cumulative response.	2-19
2.11	Principle of operation of the EM38 (in vertical orientation). Adapted from Sudduth et al. (2001).	2-20
2.12	The EM38DD with the receiver and transmitter coils of each orientation.	2-20
2.13	Calibration of the EM38DD.	2-22
2.14	Mobile configuration of an EM38DD soil survey.	2-23
3.1	An experimental variogram (grey dots) fitted by a spherical model (black line) with indication of the three variogram parameters ($C_0 = 0.1, C_1 = 0.9, a = 10$).	3-4
3.2	Bounded variogram models with the same variogram parameters ($C_0 = 0.1, C_1 = 0.9, a = 10$).	3-5

4.1	(a) Belgian soil map with indication of the study site and (b) Location of the study site in the sandy region of Flanders (Belgium).	4-4
4.2	(a) Indication of the sampling sites and (b) the EC_a measurements. The coordinates are expressed in meters, conform the Lambert72 projections used in Belgium.	4-5
4.3	(a) Interpolated map of EC_a -V and (b) EC_a -H.	4-7
4.4	Fuzzy- k -means classification based on EC_a with indication of the texture samples.	4-8
4.5	Mean sand fraction (in %) in both EC_a zones, with to the left of the graph a soil profile located in EC_a zone 1 and to the right a soil profile located in EC_a zone 1.	4-8
4.6	EC_a -V measurements obtained in (a) November 2003, (b) November 2004 and (c) November 2005.	4-10
4.7	Variograms of the NO_3^- -N residuals at the three sampling times (a) 4 November 2002, (b) 27 January 2003 and (c) 12 March 2003. On the X-axis the lag distance is expressed in m, on the Y axis the semivariogram values are expressed in $(kg\ NO_3^- - N\ ha^{-1})^2$	4-11
4.8	Maps of the NO_3^- -N content [$kg\ ha^{-1}$] in the top 60 cm at the three sampling times (a) 4 November 2002, (b) 27 January 2003 and (c) 12 March 2003.	4-12
4.9	Histogram of the ratio of the NO_3^- -N content on November 4 th 2003 to the NO_3^- -N content on January 27 th 2003 for (a) EC_a -zone 1 and (b) EC_a -zone 2.	4-13
4.10	Fuzzy- k -means classification based on OK interpolated NO_3^- -N measurements at three sampling times.	4-14
5.1	Indication of the study site in the Belgian sandy loam region . . .	5-4
5.2	Location of the earthworm samples (grid and random) and the soil samples. The coordinates are expressed in meters, conform the Belgian Lambert72 projection used in Belgium.	5-5
5.3	Numbered rings within a two pixel radius.	5-7
5.4	Experimental (dots) and fitted (line) variograms of (a) <i>L. terrestris</i> densities, (b) <i>A. caliginosa</i> densities and (c) EC_{av}	5-11
5.5	Kriged contour maps of the densities [$individuals\ m^{-2}$] of (a) <i>A. longa</i> , (b) <i>L. terrestris</i> , (c) <i>A. caliginosa</i> and (d) <i>A. rosea</i>	5-12
5.6	(a) Kriged contour map of EC_{av} [$mS\ m^{-1}$] and (b) Fuzzy- k -means classification of EC_{av}	5-13
5.7	Clay content [%] per depth interval in each EC_{av} class, the error bars depict the standard deviations from the mean.	5-13
5.8	Fuzzy- k -means classification (with $k = 2$) of (a) <i>A. longa</i> , (b) <i>L. terrestris</i> , (c) <i>A. caliginosa</i> and (d) <i>A. rosea</i> . Centroid values per class are given in the legend [$individuals\ m^{-2}$].	5-14

5.9	Fuzzy- k -means classification (with $k = 2$) of the developmental stages for one representative species per ecological category: (a) <i>L. terrestris</i> adults, (b) <i>L. terrestris</i> juveniles, (c) <i>A. caliginosa</i> adults and (d) <i>A. caliginosa</i> juveniles. Centroid values per class are given in the legend [individuals m^{-2}].	5-16
6.1	Oblique aerial photograph of large scale polygonal crop marks in Northwest Belgium (photo: J. Semey, source: Department of Archeology and Ancient History of Europe, Ghent University).	6-2
6.2	Location of the study site in the sandy region of Flanders (Belgium).	6-4
6.3	(a) Study area showing the polygonal pattern and the selected rectangle with the EM38DD sensor in the centre and (b) a vertical cross-section of a sand-filled wedge (the horizontal scale bar shows intervals of 0.20 m).	6-5
6.4	(a) Drawing of the polygonal pattern and (b) Validation raster of the polygonal pattern in which only wedges wider than 0.25 m were retained. The axes represent grid coordinates expressed in cm.	6-6
6.5	Experimental and modeled variograms of (a) $\text{EC}_a\text{-V}$ and (b) $\text{EC}_a\text{-H}$. The left part of the figure shows the experimental variogram till long lag distances, while the right graph has a maximum lag distance of 400 m. Based on the left graph the variogram was modeled.	6-8
6.6	Map of (a) $\text{EC}_a\text{-V}$ and (b) $\text{EC}_a\text{-H}$. The EC_a maps are expressed in mS m^{-1}	6-9
6.7	(a) Map of PR and (b) fuzzy- k -means classification of PR.	6-10
6.8	(a) Map of the PR coefficient of variation (CV), expressed as percentage [%], overlaid with a vector of the wedges wider than 0.25 m and (b) fuzzy- k -means classification of the PR map with an uncertainty zone added based on the CV.	6-12
7.1	Location of the study site in the Campine region of Flanders (Belgium).	7-4
7.2	Illustration of precision and recall with their trade-off.	7-9
7.3	Point ordinary kriging maps of (a) $\text{EC}_a\text{-H}$ with indication of the EM38DD measurement locations, (b) $\text{EC}_a\text{-V}$ with indication of the texture samples and (c) PR. The metric Lambert coordinates [m] are used on the X and Y axis.	7-10
7.4	Mean clay content of the heterogeneous and homogeneous soil profiles, shown in intervals of 0.50 m (errors bars represent the standard deviation).	7-12
7.5	Fuziness performance index (FPI) and normalized classification entropy (NCE) as a function of the number of classes.	7-13

7.6	Map of (a) the fuzzy classification of the profile ratio into five classes, (b) the predicted presence of the clay substratum according the fuzzy- <i>k</i> -means reclassification, (c) the IK probability showing the probability (expressed as percentage) that the profile ratio ≤ 0.81 and (d) the predicted presence of the clay substratum according the classification of the IK probability map with a probability threshold of 0.5.	7-14
7.7	(a) Indicator experimental (dots) and modeled variogram of the validation data and (b) the validation image with the validation samples (black dots are presences, while empty dots are absences).	7-15
7.8	The <i>F</i> -measure of the harmonic mean of precision and recall weighted ($F_{0.5}$) in function of the probability thresholds.	7-16
7.9	The ROC curve of the IK probability method (AUC = 0.770) is represented by the blue line. The green line indicates a model with a classification performance no better than chance (AUC = 0.5).	7-17
7.10	(a) Scatterplot of EC _a -V vs the depth of the clay substratum (D_c) and (b) the depth of the textural discontinuity (clay substratum in meters below the surface).	7-18
8.1	Location of the study site in the Polder region of East-Flanders (Belgium).	8-4
8.2	Oblique aerial image of the arable field, taken in July 1989 (photo: J. Semey, source: Department of Archeology and Ancient History of Europe, Ghent University).	8-4
8.3	Basic network architecture; the input layer (x_j) is connected with the hidden layer (z_i) through the weights w_{ij} , while w_{ki} are the weighted links that connect the hidden layer with the output layer (y_k). w_{i0} , w_{k0} are the bias terms added to the hidden, respectively output layer.	8-6
8.4	Error curve with illustration of the early-stopping cross-validation method. The dotted line indicates when the training should stop.	8-8
8.5	Maps of (a) EC _a -H and (b) EC _a -V. The map units are ms m ⁻¹	8-12
8.6	Variograms of (a) EC _a -H before and after the application of a Z-score filter, (b) EC _a -V.	8-13
8.7	Map indicating the location and the clay content [%] of (a) the topsoil samples and (b) the subsoil samples.	8-14
8.8	Scatterplots of (a) the topsoil clay content and EC _a -H, (b) the topsoil clay content and EC _a -V.	8-15
8.9	The R_n -score of all models.	8-17
8.10	(a) Map of the tclay content predicted by the HV_WS ₁ model and (b) Location map of the estimation error of the HV_WS ₁ model, projected on a classified EC _a -V map.	8-19
8.11	(a) Map of the tclay content predicted by the H_WS ₁ model and (b) Location map of the estimation error of the H_WS ₁ model, projected on a classified EC _a -V map.	8-20

List of Tables

2.1	Type of soil sensor and its targeted soil properties, adapted from Adamchuk et al. (2004). (P = Proximal, R = Remote, I = Invasive, NI = Non-invasive)	2-6
3.1	Confusion matrix.	3-12
3.2	Scale of Landis and Koch (1977) to interpret κ .	3-13
4.1	Descriptive EC_a statistics ($n = 833$).	4-7
4.2	Descriptive nitrate statistics, stratified according to the EC_a -zones.	4-10
4.3	Confusion matrix obtained after comparison of the NO_3^- -N-zones and the EC_a -zones.	4-14
5.1	Ring characteristics for a two pixel radius.	5-7
5.2	Mean ($n = 100$) earthworm densities as collected by mustard extraction in a conventionally ploughed arable field. Data are given for each developmental stage per earthworm species and per ecological category	5-10
5.3	Variogram model parameters for each developmental stage of the representative earthworm species density. All data were best fit by a spherical model	5-10
5.4	Kappa indices (κ and κ^*) and average fuzzy similarity (S_a) of classified maps related to biotic and abiotic interactions.	5-15
6.1	Textural fractions and Fe-content of two samples taken at 0.30 m intervals in the host material and in a wedge.	6-7
6.2	Descriptive EC_a statistics ($n = 480$).	6-7
6.3	Descriptive EC_a statistics, stratified according to the validation image.	6-9
6.4	Confusion matrix obtained after cross-tabulation of the pixels of the validation image (columns) and the PR classification (rows).	6-10
6.5	Confusion matrix obtained after cross-tabulation of the pixels of the validation image (columns) and the PR classification with an uncertainty zone added (rows).	6-11
7.1	A two by two confusion matrix.	7-7

7.2	Indices related to the ROC curve and the F -measure. Their formula based on a two by two confusion matrix is given (in conformity with Table 7.1).	7-8
7.3	Descriptive statistics of EC_a -H, EC_a -V and PR	7-11
7.4	Pearson correlation coefficients for mean clay content ($n = 23$) of different soil layers and EC_a -H and EC_a -V.	7-12
7.5	Accuracy measures of the two classification methods	7-16
8.1	Descriptive statistics of the EC_a -H, EC_a -V ($n = 19694$) and the textural fractions of top- and subsoil.	8-11
8.2	Pearson and Spearman correlation coefficients between EC_a and the textural fractions ($n = 78$).	8-15
8.3	Validation indices of the different network architectures, ordered according to a change in the input window size or a change in the number of sensor orientations.	8-16

List of Abbreviations and Acronyms

ATV	All terrain vehicle
ANN	Artificial neural network
AUC	Area under curve
CCR	Capacitively coupled resistivity
CEC	Cation exchange capacity
CI	Confusion index
CV	Coefficient of variation
D_c	Depth to a clay layer
DC	Direct current
DOE	Depth of exploration
EC	Electrical conductivity
EC_a	Apparent electrical conductivity
EC_{av}	Apparent electrical conductivity, averaged over the two orientations
EC_a-H	Horizontal apparent electrical conductivity
EC_a-V	Vertical apparent electrical conductivity
EM	Electromagnetic
EMI	Electromagnetic induction
ER	Electrical resistivity
FIR	Far infra-red
FPI	Fuziness performance index
GPR	Ground penetrating radar
GRS	Gamma ray spectrometry
IK	Indicator kriging
LSD	Least significance difference
GPS	Global positioning system
MEE	Mean estimation error
MIR	Mid-infrared
MS	Magnetic susceptibility
MSEE	Mean squared estimation error
NCE	Normalized classification entropy
NIR	Near infra-red
OK	Ordinary kriging
PR	Profile ratio
radar	Radio detecting and ranging
RI	Relative improvement

ROC	Receiver operating characteristic
RNE	Relative nugget effect
Rx	Receiver coil
SK	Simple kriging
tclay	Clay content in the topsoil
Tx	Transmitter coil
WS	Window Size

Samenvatting

De karakterisatie van bodemeigenschappen is essentieel voor een duurzaam land- en waterbeheer. Gedetailleerde bodeminventarisaties worden echter bemoeilijkt door de complexe en multivariate natuur van bodems. Het analyseren van bodemstalen is een fundamenteel onderdeel van een bodeminventarisatie. Meestal wordt het aantal bodemstalen echter gelimiteerd door de kost en tijd nodig voor het nemen en analyseren van de stalen. Hierdoor worden de nauwkeurigheid en resolutie van de verzamelde bodeminformatie ook begrensd. Met een toenemende vraag naar gedetailleerde informatie is ook de interesse voor secundaire bodemdatabronnen gestegen. In deze context hebben recente technologische ontwikkelingen het verzamelen van hoge resolutie bodeminformatie mogelijk gemaakt door de introductie van proximale geofysische bodemsensoren.

Dit werk omvat de studie van een nieuw type bodemsensor waarbij de nadruk lag op de verwerking van duale sensor signalen met een hoge resolutie. De EM38DD bodemsensor meet de schijnbare elektrische geleidbaarheid (EC_a) van een bodem door elektromagnetische inductie (EMI). Deze sensor bestaat in feite uit twee enkelvoudige EM38 toestellen loodrecht op elkaar met elk hun eigen diepte gevoeligheid. Hierdoor worden er simultaan twee EC_a metingen bekomen: één in de horizontale oriëntatie (EC_a-H) en één in de verticale (EC_a-V). De mogelijkheden van dit tweevoudig signaal om bodemeigenschappen te inventariseren en karakteriseren werden onderzocht.

In een eerste luik werd er nagegaan of de EM38DD aangewend kan worden om de nitraathuishouding van een weiland te karakteriseren op een binneninperceelsniveau. In de Vlaamse Zandstreek werden EC_a kaarten gemaakt voor een 1 ha weiland waaruit de texturele bodemvariabiliteit kon afgeleid worden. Beide sensor-oriëntaties weerspiegelden hetzelfde ruimtelijk patroon van bodemvariatie. Met een fuzzy- k -means classificatie werden er twee EC_a zones afgebakend: één zone werd gekenmerkt door lage EC_a waarden en een homogeen zandig bodemprofiel terwijl in de andere zone hogere EC_a waarden voorkwamen door een reductie van het zandgehalte in de diepte (60-90 cm). Het nitraatgehalte (0-60 cm) werd bepaald op drie tijdstippen: in november, januari en maart. Een stratificatie van deze nitraatdata volgens de EC_a zones wees op een differentiële nitraathuishouding in deze zones. Daarom werd gestratificeerde simple kriging gebruikt als methode om het nitraatgehalte in kaart te brengen op de drie tijdstippen. Aan de hand van de gemiddelde gekwadrateerde schattingsfout (MSEE) als validatieparameter werd deze methode ook beter geëvalueerd dan ordinary kriging. Bovendien werden er voor de periode van november tot januari meer nitraatverliezen vastgesteld

in de zone met lage EC_a waarden, terwijl in de andere EC_a zone, gekenmerkt door meer klei en silt in de ondergrond, de nitraatverliezen geringer waren. Er werd dus een verschillende evolutie van het residueel bodemnitraatgehalte tijdens de winterperiode geobserveerd in de EC_a zones. Verder werden de drie ordinary kriging kaarten van het nitraatgehalte ook geclassificeerd en een kappa waarde van 0.73 werd gevonden wat wijst op een redelijke overeenkomst tussen de nitraatzones en de EC_a zones. Deze resultaten toonden de mogelijkheden van een EM38DD bodemsensor aan voor plaats-specifiek stikstofbeheer in Vlaanderen.

Een volgend luik was gericht op de ruimtelijke karakterisatie van het voorkomen van regenwormen aan de hand van de EM38DD. Kennis over de ruimtelijke variabiliteit van regenwormen is essentieel bij tot het optimaliseren van het bodembeheer. Zowel biotische als abiotische factoren dragen bij tot het ruimtelijk voorkomen van regenwormen maar deze invloeden zijn niet gemakkelijk te onderscheiden. Zowel een intensieve staalnamecampagne van de regenwormsoorten als een EC_a bodeminventarisatie werden uitgevoerd om de ruimtelijke variabiliteit van regenwormgemeenschappen en populaties te beschrijven en gelijkenissen tussen de patronen van regenworm-voorkomen en de bodem EC_a als abiotische component werden bepaald. Aan de hand van een geostatistische analyse kon besloten worden dat de endogeïsche soorten (bodemwoelers) voorkwamen in overlappende clusters, terwijl het voorkomen van anekische soorten (diepgravers) ruimtelijk gescheiden was. Bovendien was de EC_a variabiliteit alleen gerelateerd met het voorkomen van diepgravende regenwormsoorten. Deze bevindingen werden bevestigd in een uitgebreide categorische vergelijking van de bodem EC_a kaart en kaarten van het voorkomen van de regenwormsoorten. Een grote graad van gelijkenis werd gevonden tussen de geklassificeerde kaart van de anekische soorten en de EC_a wat erop wijst dat hun variabiliteit gedreven werd door de ondergrondse bodemvariabiliteit. Deze studie bewijst dat een bodemonderzoek gebaseerd op EM38DD data een veelbelovende aanpak is om de ruimtelijke variabiliteit van diepgravende regenwormsoorten te identificeren.

Vorst-wig pseudomorfen zijn belangrijke indicatoren van het paleoklimaat, maar methodes om hun polygonaal patroon te observeren zijn eerder beperkt. In de volgende studie werd er hiervoor een nieuwe methode gebaseerd op EM38DD gegevens voorgesteld en geëvalueerd. Een polygonaal network van met zand gevulde wigen was tijdelijk zichtbaar aan de oppervlakte: de diameter van de polygonale cellen varieerde tussen 0.2 m en 10 m. Op basis van de wigen groter dan 0.25 m werd er een gedetailleerde tekening gemaakt van het network die na digitalisatie en rasterisatie gebruikt werd als validatiekaart. Een bodem EC_a onderzoek werd manueel uitgevoerd met een resolutie van 0.5 m op 0.5 m. Tegen alle verwachting in was er noch op de EC_a -H kaart, noch op de EC_a -V kaart een polygonaal network zichtbaar. De wigen waren nochtans gekenmerkt door een andere bodemtextuur dan het gastmateriaal. In de wigen bevond zich eolisch zand terwijl het materiaal van de bodem errond bestond uit marien, Fe-rijk zand en klei. Door de verhouding te nemen van beide EC_a signalen wordt de profiel ratio (PR) verkregen en deze bleek zeer informatief: het polygonaal patroon was duidelijk

aanwezig op een kaart van de PR. De laterale texturele variabiliteit van het gastmateriaal bleek groter dan het textuurcontrast tussen wig en gastmateriaal en een enkelvoudig EM38DD signaal werd dus voornamelijk beïnvloed door de bodemvariabiliteit van het gastmateriaal. De duale dipool configuratie van de EM38DD werd fundamenteel bevonden om een netwerk van vorst-wig pseudomorfen waar te nemen. De nauwkeurigheid van de voorgestelde methode ter observatie van polygonale netwerken werd bevestigd door een algemene nauwkeurigheid (P_o) van 81% bij vergelijking van de geklassificeerde PR kaart en de validatiekaart.

Naast laterale bodemvariabiliteit, is ook kennis van de verticale bodemvariabiliteit bepalend voor de meeste bodembeheersmaatregelen. Daarom werd de capaciteit van de EM38DD om texturele discontinuïteiten af te bakenen onderzocht. Het onderzoek vond plaats op een deel van een bosplantage in de Kempen waar in de zandgronden een kleisubstraat op variabele diepte aanwezig is. Aan de hand van een EM38DD bodemonderzoek werd alleen de PR kaart informatief bevonden voor de afbakening van het kleisubstraat: een cirkelvormig patroon met lage PR waarden (<1) werd geïdentificeerd als het kleisubstraat. De variatie in een enkelvoudig EM38DD signaal werd voornamelijk bepaald door de laterale variabiliteit van de bodemtextuur en hieruit kon de aanwezigheid van een kleisubstraat niet afgeleid worden. Om de omvang van het kleisubstraat af te bakenen werden twee numerische methodes gebruikt: (i) een fuzzy- k -means classificatie van de PR kaart waarna de klasse met de laagste centroid waarde weerhouden werd, en (ii) een probaliteitsaanpak aan de hand van indicator kriging. Een indicator validatiekaart werd verkregen door een reeks boringen en op basis hiervan werden er voor beide methodes kaartvergelijkingstatistieken berekend die aangevuld werden met ruimtelijke nauwkeurigheidstatistieken. De indicator kriging methode werd best bevonden en de nauwkeurigheid werd gekwantificeerd met een P_o van 83.8%, de locationele kappa was 0.864 en S_a was 0.841. Beide methodes onderschatten echter de grootte van het kleisubstraat. Deze resultaten benadrukten dat het tweevoudig signaal onontbeerlijk is om texturele discontinuïteiten te karakteriseren.

Het laatste deel van de thesis behandelt de integratie van de duale EM38DD signalen om de voorspelling van de textuur van de bovengrond te verbeteren. Op een 10.5 ha polderveld werd met een EM38DD een bodemonderzoek uitgevoerd met een resolutie van 2 op 2 m en 78 stalen van de boven- en ondergrond werden geanalyseerd op textuur. Door de aanwezigheid van oude waterkanalen in de ondergrond was de variatiecoëfficiënt (CV) van het kleigehalte in de ondergrond veel groter (45%) in vergelijking met de bovengrond (13%). De EC_a-H en EC_a-V metingen vertoonden een analoog patroon wat wijst op een dominante invloed van de grote texturele variabiliteit van de ondergrond op beide signalen. Om informatie over de textuur van de bovengrond te extraheren uit diepte-gewogen EM38DD gegevens werden artificiële neurale netwerken (ANN) aangewend. Het effect van verschillende inputlagen op de voorspellingsnauwkeurigheid van klei in de bovengrond werd onderzocht en geëvalueerd. Om de reactie van de bovengrond te identificeren werden zowel ANNen met een enkelvoudig signaal als met beide EM38DD signalen getraind. Om de invloed van de lokale pixelomgeving te onder-

zoeken werden ANNen opgesteld met als input contextuele EC_a informatie. Hier toe werd er een venster van EC_a informatie rond een centrale pixel meegenomen in de input. Het beste ANN model was gebaseerd op beide oriëntaties zonder rekening te houden met contextuele informatie: de MSEE was $2.83\%^2$ en het ANN model verklaarde 65.5% van de kleivariabiliteit in de bovengrond met een variantie van $0.052\%^2$. Dus door het samenbrengen van het tweevoudig EM38DD signaal in een ANN werd de voorspelling van het kleigehalte in de bovengrond duidelijk geoptimaliseerd.

Tenslotte werd besloten dat bodem EC_a gegevens zeer waardevol zijn om bodeminventarisaties te verbeteren waarbij het doel van de inventarisatie een cruciale factor is die de verwerkingsaanpak van EM38DD gegevens bepaalt. De resultaten toonden het belang aan van het tweevoudig signaal van de sensor voor nauwkeurige laterale en verticale bodeminventarisaties.

Summary

The characterization of soil properties is a key to sustainable management of land resources. Detailed soil inventory however is complicated by the complex and multivariate nature of soil. Typically, soil inventories are based on a number of soil samples. Nevertheless, the cost and time needed to collect and analyze soil samples limits the detail of soil information. With an increasing need for detailed soil inventories, the use of ancillary data to improve the accuracy of soil information gained interest. In this context, recent technological developments enabled the collection of high-resolution soil information through the introduction of proximal geophysical soil sensors.

This thesis comprises the exploration of a new type of soil sensor, with emphasis on the processing techniques of high resolution dual soil sensor signals. The EM38DD soil sensor measures the soil apparent electrical conductivity (EC_a) based on the principle of electromagnetic induction (EMI). More specifically this sensor consists of two single EM38 sensors, each with its own depth sensitivity. As such dual EC_a measurements are obtained simultaneously: one in the horizontal orientation (EC_{a-H}) and one in the vertical orientation (EC_{a-V}). The possibilities of an EM38DD sensor data to inventory and characterize soil properties were examined.

First, it was investigated whether the EM38DD soil sensor offers insight into the within-field nitrate dynamics of a pasture field. On a 1 ha field in the sandy region of Flanders the within-field soil textural variability was revealed by the EC_a maps. Both sensor orientations yielded similar EC_a patterns. Using a fuzzy- k -means classification two EC_a zones were delineated in the field: one zone was characterized by low EC_a values and a homogeneous sandy profile while in the other zone higher EC_a values were observed and a reduction of the sand content in depth (60–90 cm) was found. The nitrate content (0–60 cm depth) was measured in November, January and March and a stratification of these data according to the EC_a zones indicated a differential nitrate behaviour in these zones over winter. Therefore, stratified simple kriging was used to map the nitrate content at the three sampling times. This mapping technique was evaluated to be better in terms of prediction accuracy (MSEE) compared to ordinary kriging of the nitrate data. Moreover, large nitrate losses were observed in the period from November till January in the zone with lower EC_a values, while in the EC_a zone characterized by more clay and silt content in the subsoil less nitrate losses were observed. As such, the EC_a zones represented zones with a distinct evolution of the residual soil nitrate content during winter. The three ordinary kriged nitrate maps were classified

and a substantial agreement was observed between the EC_a zones and the nitrate zones as expressed by a κ of 0.73. These results emphasized the potential of the EM38DD soil sensor for site specific nitrogen management.

A next study aimed at the spatial characterization of earthworms occurrence using an EM38DD soil sensor. To optimize soil ecosystem management knowledge about the spatial variability of earthworms is essential. Both biotic and abiotic factor drive the occurrence of earthworms, but elucidating them is not straightforward. Based on an intensive earthworm sampling campaign and an EC_a survey, the spatial variability of earthworm communities and populations was described and similarities between their spatial patterns and the EC_a variability as the abiotic component were assessed. Using geostatistics, it was observed that the endogeic species occurred in overlapping clusters, while the anecic species occurred in spatially separated zones. Furthermore, the measured EC_a variability was only related to occurrence of the deep-burrowing earthworm species. An elaborated categorical map comparison between the earthworm and EC_a maps confirmed these findings. A high degree of similarity was found between the classified maps of the anecics and the soil EC_a indicating that their spatial variability was determined by the subsoil variability. These results proved that an EM38DD based soil EC_a survey is a promising approach for identifying the spatial variability of deep burrowing earthworm species.

Frost-wedge pseudomorphs are important paleoclimate indicators but only limited methods exist to detect the wedge casts. In the following study, a new approach to detect the polygonal pattern of frost-wedge pseudomorphs based on the EM38DD sensor was proposed and evaluated. Temporarily, a polygonal pattern of sand-filled frost-wedges was exposed with polygonal cells ranging from 0.2 m to 10 m diameter. A detailed drawing of the polygonal network of these frost-wedge pseudomorphs was made allowing the creation of a validation image which only considered wedges wider than 0.25 m. Manually, an EM38DD soil survey was performed at a resolution of 0.5 by 0.5 m. Surprisingly, neither on the EC_a -H map, neither on the EC_a -V map the polygonal pattern was visible. Nevertheless, soil EC_a differences were expected since the wedges were characterized by a different soil texture compared to the host material. The wedges were filled with wind-blown sands while the host material consists of marine Fe-rich sand and clay. Calculating the profile ratio (PR) by rationalising EC_a -H to EC_a -V however was very informative: the PR map clearly revealed the polygonal pattern. It was concluded that the textural variability of the host material was larger than the textural differences between wedges and host material and a single EM38DD signal was dominated by the host soil textural variability. The dual dipole configuration of the EM38DD soil sensor appeared to be essential to detect the network of frost-wedge pseudomorphs. The accuracy of the EM38DD based polygonal network detection was confirmed by an overall accuracy of 81% when comparing the classified PR map with the validation image.

Additionally to lateral soil variability, also information concerning the vertical soil variability is required for most soil management decisions. Therefore, the potential of the EM38DD soil sensor to delineate soil textural discontinuities was inves-

tigated. The study site was part of a forest plantation in the Campine region where in the typically homogeneous sandy soils a clay substratum occurred at a variable depth. Based on an EM38DD soil survey, only the PR map seemed to be informative for the delineation of the clay substratum: a circular pattern with decreased PR values (<1) was identified as the clay substratum. The variation in a single sensor orientation was mainly driven by the lateral textural variability of the site and the presence of a clay substratum could not be deduced from a single signal. To delineate the extent of the clay substratum, two numerical methods were used: (i) a fuzzy- k -means classification of the PR map focusing on the lowest centroid class, and (ii) a probability approach through indicator kriging. Using a validation image obtained from directed auguring, cell-by-cell comparisons were made for these two methods complemented with spatial accuracy measures. Although both methods tended to underestimate the spatial extent of the clay substratum, the indicator kriging method was the most accurate, with an overall accuracy of 0.838, a proportion of error due to locational errors (κ_{loc}) of 0.864, and S_a was 0.841. These results emphasized that the dual signals of an EM38DD soil sensor are primordial to characterize soil textural discontinuities.

The last part of the thesis focused on the optimal integration of the dual signals of the EM38DD to improve topsoil textural predictions. In a 10.5 ha polder field, an EM38DD survey was performed with a resolution of 2 by 2 m and 78 samples of top- and subsoil were analyzed for texture. Due to the presence of former water channels in the subsoil, the coefficient of variation of the subsoil clay content (45%) was much larger compared with the topsoil (13%). The EC_a -H and EC_a -V measurements displayed a similar pattern, indicating a dominant influence of the subsoil features on both signals. To extract topsoil textural information from the depth-weighted EM38DD signals we turned to artificial neural networks (ANNs). We evaluated the effect of different input layers on the ability to predict the topsoil clay content. To identify the response of the topsoil, both EM38DD orientations were used. To examine the influence of the local neighbourhood, contextual EC_a information by means of a window around each soil sample was added to the input. The best ANN model used both EM38DD signals but no contextual information: a mean squared estimation error (MSEE) of $2.83\%^2$ was achieved and the model explained 65.5% of the topsoil clay variability with a variance of $0.052\%^2$. So, by fusing the two EM38DD signals through ANNs, the prediction of the topsoil clay content was optimized.

Finally, it was concluded that soil EC_a data are valuable to improve soil inventories with the objective of the soil inventory as crucial factor determining the approach to EM38DD data processing. The dual signals of the EM38DD soil sensor were stressed to be essential for adequate lateral and vertical soil inventories.

1

Introduction

1.1 Framing soil sensors within a soil inventory context

We know more about the movement of celestial bodies than about the soil underfoot - Leonardo DaVinci, circa 1500s

Although the characterization of soil properties is a key to the sustainable management of land resources, soil is not straightforward to inventory. From the beginning of soil science, man struggled to capture and map information of soil properties. Soil is inherently highly variable due the complex interaction of the soil-forming factors climate, organisms, relief and parent material acting and interacting through space and time (Jenny, 1941). The significance of soil spatial variability is related a.o. to controlling the hydrological cycle, optimizing the crop yield potential and limiting environmental pollution risks. This variation is usually considered to be problematic in relation to sampling effort, soil data quality and uncertainty for optimal soil management decisions (McBratney, 1992). Moreover, our information-based era forced soil survey (or soil mapping) methods to evolve and adapt to the demand for quantitative information at finer and finer resolutions (McBratney et al., 2003).

Traditional soil survey concepts are based on qualitative recognition of soil properties in relation to landscape and environmental variables (Scull et al., 2003), yielding the delineation of homogeneous soil map units. As an answer to the criticisms on the qualitative character of the soil maps, quantitative methods have been developed the last thirty years, recognizing soil variability as a key soil attribute (Cook et al., 1996; Burrough et al., 1994). To account for the uncertainty and complexity associated with a soil survey a wide range of mathematical, (geo)statistical and numerical methods were developed (McBratney et al., 2000). Besides, classical soil survey and analyses in the lab are expensive and time-consuming, limiting the scale and accuracy of the soil survey information.

With an increasing awareness of environmental, ecological, agricultural and natural resources issues, a precise characterization of the soil and its variability becomes essential. This has prompted the development of more time- and cost-efficient quantitative methodologies for soil characterization. The last decade the use of ancillary data has gained attention to improve the accuracy of predicting soil properties (Finke et al., 2004; Minasny and McBratney, 2007; Peng et al., 2003). Moreover, recent technological developments enabled high resolution data acquisition through remote and proximal sensors. The term ‘proximal sensor’ is used for terrestrial based sensors collecting data from a distance of less than two metres above the soil surface, while the term ‘remote sensor’ is reserved for air- or space-borne sensors. Soil properties suited to be characterized by remote sensing should have a clear surface expression (e.g. soil moisture, salts). Proximal soil sensors are suited best for small areas and detailed investigations (Rossiter, 2005), as such

these sensors fit perfectly in the framework of soil mapping at a fine resolution. In this context, a high spatial resolution is defined as 10 m or less (Viscarra Rossel et al., 2008). Advantages of using these sensors over classical soil sampling include their (possible) non-invasive nature, the collection of high resolution soil information in real time, the relative low cost of data acquisition and the possibility for a mobile survey configuration (Viscarra Rossel et al., 2007). In fact, proximal soil sensors supply a window to the surface and are a highly valuable source of soil information. Over the last years, the most widely used soil sensors are electrical and electromagnetic induction (EMI) sensors (Adamchuk et al., 2004).

This study concentrates on one type of EMI sensor being commercially available: the EM38DD soil sensor (Geonics Ltd, Canada). An EMI sensor measures the apparent electrical conductivity (EC_a) of the soil, which generally varies with salinity, clay content and water content (McNeill, 1980a). The EM38DD soil sensor consist of two single EM38 sensors. The depth of influence of the EM38 depends on its operation mode: it is designed to be used in horizontal or vertical orientation, and this orientation determines the depth-weighted response curve of the sensor (Lesch et al., 2005). The horizontal EC_a reflects topsoil properties (up to 0.75 m depth), while in the vertical orientation the depth of influence increases up to 1.6 m depth. As such, the orientation of the sensor is crucial to obtain soil information from the depth of interest.



FIGURE 1.1 – The EM38DD soil sensor.

In 2001, an adapted version of the EM38 was developed, being the dual dipole EM38 or the EM38DD soil sensor (Fig. 1.1), which consists of two EM38 units perpendicular to each other, allowing the simultaneous measurement of both horizontal and vertical soil EC_a . The dual dipole character of the EM38DD sensor has a lot of potential for applications in soil management since both top- and sub-soil properties are addressed simultaneously. The department of Soil Management

from the Ghent University was the first in Europe to buy an EM38DD sensor in 2001. It made them a pioneer in exploring the possibilities of two co-located soil EC_a signals.

1.2 Aim and research questions

The rationale of this research goes back to the late 20th century when non-invasive soil sensors were first introduced into soil science. At that time, soils were characterized by taking a limited number of soil samples, which were analyzed in the lab and used in combination with knowledge on soil-landscape relations to draw soil map units. When the concept of spatial variability within these units was launched, the limitations of our soil information sources became a general awareness. Methods to bridge the gap between qualitative and quantitative soil inventory techniques and the increasing demand for highly detailed soil information occupied soil scientists over the last few decades.

The general aim of this dissertation is to demonstrate the possibilities as well as the shortcomings of a modern sensor-based approach to inventory and characterize soil properties at a high resolution. In fact, this dissertation introduces the in 2001 developed EM38DD soil sensor in soil surveying for different soil management applications. The novelty of this sensor is its dual character by which two signals are obtained simultaneously: the strength of processing two signals instead of a single one was assessed.

To reach the general aim, the following research questions were answered:

- Which soil management applications can benefit from an EMI sensor-based approach?
- Does the within-field soil variation of Flanders (Belgium) require high resolution soil data?
- Lateral and vertical soil heterogeneity: a single or dual EM38 problem?
- Is there a difference in characterizing top- or subsoil properties?
- Does the dual dipole character supply complementary soil information?
- Are soil samples and lab analyses now redundant?
- Which processing techniques are suitable to analyse EM38DD data?
- What is the optimal procedure to increase information extraction from dual signals?
- What are the limitations of the EM38DD soil sensor?

These questions are related to the objective of the inventory, to the characteristics of the EM38DD sensor or to the methods and techniques needed for sensor-based soil mapping.

1.3 Outline of the research

The outline of this research is schematized in Fig. 1.2.

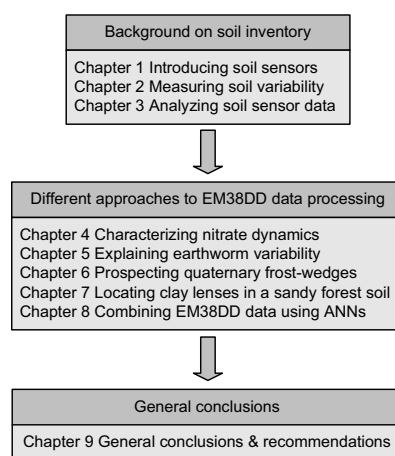


FIGURE 1.2 – Outline of the research.

In the first three chapters a general background on the topic is given. An introduction on soil sensors and the state of the art is given in Chapter 1, Chapter 2 contains a literature review related to measuring soil variability and Chapter 3 addresses the methodology to analyze soil sensor data. Then the results were structured according to the objective of soil inventory and the approach followed to analyze the EM38DD data.

Chapter 4 deals with the within-field variability of a pasture soil in the sandy region of East-Flanders, Belgium. A literature review on precision agriculture highlights the need for site-specific nitrogen management. The nitrate dynamics of the pasture field were characterized and a nitrate mapping approach based on electromagnetic induction was proposed. Finally, the use of the sensor to delineate zones with similar nitrogen dynamics was evaluated.

In Chapter 5, within-field soil classes were characterized in terms of the soil biology. The study site was an arable field in the central loess belt of Belgium. Related to the soil EC_a, differences in the abundance of earthworm species were investigated and similarities in the spatial patterns of soil EC_a and earthworm abundances were quantified through map comparison statistics. Soil quality assessment

using an EMI-based soil sensor was discussed.

Chapter 6 and 7 both aimed at defining a strategy to locate soil textural discontinuities. In Chapter 6 paleo-patterns from quaternary times were the topic under investigation. A field survey was performed and a combination of the simultaneously measured horizontal and vertical EC_a was established to reveal the polygonal network of frost-wedge pseudomorphs.

In Chapter 7 a detailed inventory of the lateral and vertical soil heterogeneity of a sandy forest soil was the objective. The extent of the vertical heterogeneity was delineated using two data analyzing techniques: (i) a classification approach and (ii) a probability approach through indicator kriging. Moreover, several advanced map comparison statistics were used to quantify the accuracy of the delineation.

Chapter 8 evaluated the use of artificial neural networks for topsoil textural predictions. The aim was to filter out the effect of the subsoil on the horizontal soil EC_a . As a data fusion method both orientations of the EM38DD sensor were used in the input layer of the network. The sensitivity of the network to changes in the input layer (either including single or dual sensor information, either including contextual EC_a information) was assessed and compared in terms of the accuracy of the prediction.

The last chapter summarizes the results from the different case studies and links these results to the originally defined research questions. General conclusions are drawn concerning the use of an EM38DD sensor for high resolution soil inventories.

2

Quantifying soil variability

2.1 Soil variability: the scale issue

Soils have the key characteristic to vary in space and with time (Sommer, 2006). This variation in space holds true for a range of spatial scales: from within-field soil variability (the micro-scale) over variability between fields (the meso-scale) and landscapes (the macro-scale) to variation between regions (the mega-scale) (Bouma and Finke, 1993). With the term scale is referred to the geographical or observation scale which is defined as the spatial extent of a study or the spatial environment within which a question is posed (Jenerette and Wu, 2000; Lam and Quattrochi, 1992). There are two categories of variability - systematic or functional variations and random fluctuations or noise (Hall and Olson, 1991; Logsdon et al., 2008). According to Wilding and Drees (1978), the systematic variations are those related to the five soil-forming factors, with topography (relief) being the most addressed factor (Demas and Rabenhorst, 2001). In other words, the systematic soil patterns are described and analyzed in terms of pedogenesis. However, the relative importance of these factors is scale dependent (Benayas et al., 2004). With an increasing spatial scale, the predominance of topography to predict the variability of soil properties would gradually be replaced by parent material and climatic factors (Park and Vlek, 2002). Focusing on two spatial scales, the factors influencing the spatial variation of the soil chemical composition was put in a model by Benayas et al. (2004) (Fig. 2.1). Variation in climate and parent material are most relevant at the landscape scale, while at a field scale topography, soil texture, soil moisture and plant community composition affect gradients of the soil chemical composition. The arrows indicate an (in)direct effect from one factor to another.

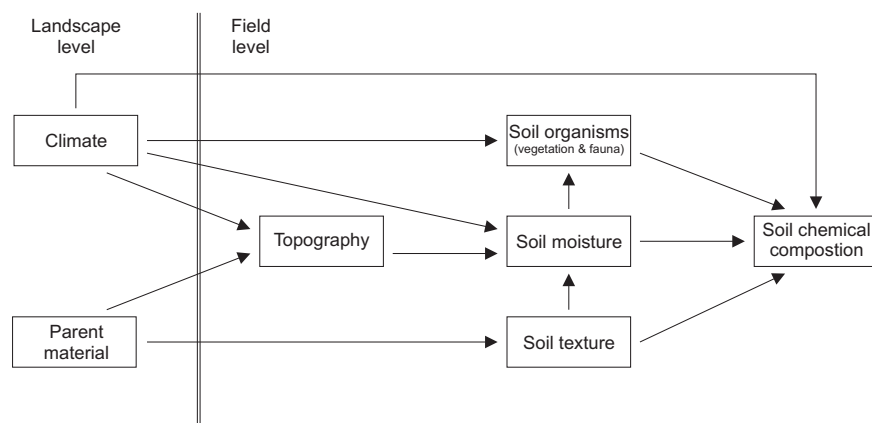


FIGURE 2.1 – Hypothesized model of spatial variation of soil properties, adapted from Benayas et al. (2004).

Random differences, on the other hand are said to compose those parts of the observed variation that are not related to the soil-forming factors or cannot be directly analysed at the chosen observation scale (Burrough, 1983). Soil management induced variability also belongs to this category of variation (Bouma and Finke, 1993). Upchurch and Edmonds (1991) state that the division between random and systematic soil variation is based on the source of error produced by the variation. Systematic variability can be attributed to a known cause, understood and predicted. Variability that cannot be related to a certain cause is called random variability. In general, natural or undisturbed soils are expected to have larger systematic than random variability.

Nevertheless, the distinction between systematic and random components is also scale dependent since an increase of the scale of observation almost always reveals structure in the noise (Burrough, 1983). Geostatistical tools (see section 3.1), like variography, allow differentiation between structured variation and micro-scale and/or random variation (Cressie, 1991). Each scale level has different implications for the soil information required (McBratney et al., 2000). At the within-field level, information is needed for agricultural or natural resources management. As the scale increases, precise soil knowledge is needed for environmental monitoring and for studying changes brought about by a disruption of ecosystems. At the mega-scale, soil information is required for the study of the global climate, food production and supply and for advising policy-makers. Besides, soil spatial variability also affects processes such as crop growth and leaching of pollutants and the key factors causing variation in a process may differ according to the spatial scale (Lark, 2005). Within-field clay variability might determine variations in the organic carbon content, while differences in clay content between fields might be relatively unimportant compared to topography or land use differences. Already in 1992 it was recognised that scaling issues should be considered as a primary focus of research efforts. Levin (1992) stated that 'The description of pattern is the description of variation, and the quantification of variation requires the determination of scales'.

2.2 Quantifying soil variability

The characterization of the spatial variability of soil properties is essential in environmental prediction, precision agriculture and natural resources management (Lin et al., 2005). Considerable research has been done characterizing soil variability at a particular scale. Physical and chemical properties do not only vary across a field, substantial variability can be found at a within-field scale (Corwin et al., 2003). Under Flemish field conditions however, chemical properties were shown to display dominantly a micro-scale (< 5 m) and random variation, while the physical and biological properties displayed a strong structured variability (Van Meir-

venne, 2003). Lin et al. (2005) made an inventory of the soil spatial variability at four scales and found that most of the variability of soil pH and depth to the A-horizon was within a 12 m distance. All these studies characterized the soil variability at a micro-scale using one of the methods explained below.

2.2.1 Direct information

Direct measurement of soil variability involves soil augering according to a predefined sampling design, followed by a laboratory analysis of the required soil properties. At the origin of the sampling strategy in the 1930s, the aim of the sampling was to estimate the mean, and to a lesser extent the variance of the soil property of interest. With the introduction of geostatistics in soil science, the expression of the spatial variation became essential since it uses a model of spatial dependence, the variogram (Webster and Oliver, 2001) (see section 3.1). Quantifying soil variability requires soil sampling at an intensity that will allow the variability to be mapped with some degree of confidence (Wollenhaupt et al., 1997). As such, the degree of spatial soil variability influences the intensity with which a soil must be sampled to characterize the properties of an area of interest (Corwin et al., 2003; Petersen and Calvin, 1986; Oliver, 1999). The more heterogeneous the soil, the more intense the sample rate must be to attain a given precision. To quantify the spatial variation, the sampling interval should fall within the scale of variation. However, few data are available from which the magnitude of the different categories of variability may be estimated. A variogram provides an unbiased description of the scale and pattern of spatial variation (Oliver, 1999). Nevertheless, to obtain a reliable variogram Webster and Oliver (2001) recommended at least 100 sampling locations. Commercial soil surveys based on auger sampling are typically conducted based on grid of 100 m by 100 m (1 sample per hectare) (Earl et al., 2003). Frogbrook (1999) demonstrated that datasets with this sample density are too sparse to identify close-spaced spatial variation in soil characteristics. However, mostly the affordable number of observations is limited from an economic and practical point of view due to the time-consuming nature of hand augering and expensive laboratory analyses. The development of alternative methods to obtain information of the soil variability is crucial. Two approaches can be followed: (i) the use of ancillary, easy-to-obtain data to guide the sample intensity (Godwin and Miller, 2003; Kerry and Oliver, 2003) or (ii) the use of sensor technology to obtain a direct or indirect measurement of the soil property (Viscarra Rossel and McBratney, 1998). The latter method is not not as accurate as conventional laboratory analysis, but they allow the collection of larger amounts of data using simpler, cheaper and less laborious techniques. Furthermore, the information is produced in a timely manner (Viscarra Rossel et al., 2008).

2.2.2 Sensed information

To provide a cheap and fast measurement of a soil property, different types of soil sensors have been developed. The type of soil sensor can be defined according to:

- the measurement method;
- the distance from the sensor to the soil: proximal or remote;
- the invasiveness into the soil: invasive or non-invasive.

Non-invasive sensors provide estimates of a soil property from above ground observations, while invasive sensors disrupt the soil surface and penetrate the soil to either sense directly, or collect soil for an external detection. Several soil sensor systems are being developed and most of them involve one of the following measurement methods (Adamchuk et al., 2004):

- electrical and electromagnetic sensors that measure the electrical resistivity/conductivity or capacitance affected by the composition of the soil
- optical and radiometric sensors that use electromagnetic waves to detect the level of energy reflected/absorbed by soil particles;
- mechanical sensors that measure forces resulting from a tool engaged with the soil;
- acoustic sensors that quantify the sound produced by a tool interacting with the soil;
- pneumatic sensors that assess the ability to inject air into the soil;
- electrochemical sensors that use ion-selective elements producing a voltage output in response to the activity of selected ions.

Sensors based on the application of geophysical methods to investigate near-surface physico-chemical properties are also called geophysical sensors. Geophysical methods respond to the physical properties of the sub-surface media: geophysics is the application of physical quantity measurement techniques to provide information on conditions or features beneath the Earth's surface (Allred et al., 2008). Geophysical methods can be classified into passive methods which detect variations within the natural fields associated with the earth and active methods which artificially generate signals that are transmitted into the soil (Reynolds, 1997). Electrical, electromagnetic and radiometric sensors belong to this type of sensor. An ideal soil sensor responds to the variability of a single soil property. In reality however, the signal of most sensors is affected by more than one soil property. An overview of the type of soil sensors and the targeted soil properties

is given in Table 2.1. Mostly the best correlation between the sensor output and a particular soil property was found for a specific soil type or when the variation of the interfering properties was negligible.

TABLE 2.1 – Type of soil sensor and its targeted soil properties, adapted from Adamchuk et al. (2004). (P = Proximal, R = Remote, I = Invasive, NI = Non-invasive)

	Soil sensors				
	Electrical and electromagnetic	Optical and radiometric	Mechanical	Acoustic and pneumatic	Electrochemical
<i>Type</i>					
Distance	P	R/P	P	P	P
Invasiveness	I/NI	I/NI	I	I/NI	I
<i>Soil property</i>					
Soil texture	X	X		X	
Soil organic matter	X	X			
Soil moisture	X	X			
Soil salinity	X				X
Soil compaction			X	X	
Depth variability	X		X	X	
Soil pH		X			X
Nitrate	X	X			X
Other macronutrients					X
CEC	X	X			

A detailed description of how measurements done by these sensors relate to the targeted soil properties is given below.

2.2.2.1 Remote sensing

Remote sensing can be defined as ‘the science and art of obtaining information about an object, area, or phenomenon through the analysis of data acquired by a device that is not in contact with the object, area, or phenomenon under investigation’ (Lillesand and Kiefer, 1995). The data are collected from a distance that varies from a few metres to hundreds of kilometers. Remote sensing occurs through aerial or satellite imagery and is mostly based on the electromagnetic (EM) spectrum. Sensors are designed to detect and record the reflectance or emission of radiation from an object. This radiation may range from the Visible-UV wavelengths through the Near-, Mid- and Far-Infrared (NIR, MIR and FIR), microwave and gamma ray regions of the EM spectrum (Fig. 2.2). Soil reflectance is a function of soil’s chemical and physical composition. The response in different parts of the spectral range may be affected by a different combination of soil properties. Most of the spectral responses can be related to differences in organic matter content, iron content, texture, soil moisture and soluble salts as is shown by several studies e.g. Sudduth and Hummel (1991); Barnes and Baker (2000); Chen et al. (2000); D’Urso and Minacapilli (2006).

One of the earliest applications of remote sensing was the use of aerial imagery to aid soil mapping (Bushnell, 1932). Nowadays, a variety of aircraft and satellite

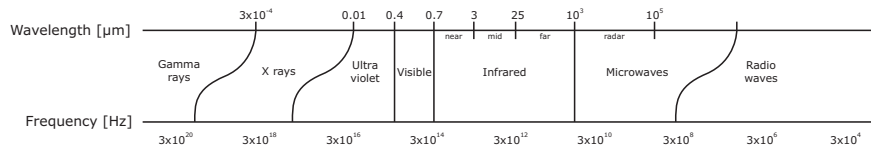


FIGURE 2.2 – The electromagnetic spectrum.

based remote sensing data such as photographs, videographs, hyperspectral and multispectral images have become available at a high spatial resolution. For example, the resolution of multispectral data increased from 250 m (MODIS MS), to 30 m (Landsat 7 MS), 10 m (SPOT 5MS), 4 m (IKONOS MS) and 1 m (Quickbird MS) (Dalsted et al., 2003). As such, the spatial resolution of remote sensing data is sufficient for using them as an ancillary data source in high resolution soil inventories (Sullivan et al., 2005; Lopez-Granados et al., 2005).

However, the use of remote sensing data is limited due to the presence of growing crops, crop residues or other land cover, as mostly the reflectance of bare soil is used to characterize the soil. It is difficult to obtain bare soil images since there is always some degree of mixture with vegetation or dry plant components. Besides, changes in surface tillage condition, rain compaction and moisture also induces changes in the soil's reflectance that might approach or exceed spectral responses due to physical soil properties (Barnes and Baker, 2000). It should be noted that remote sensing of crops (mostly aerial photography) also has been used to investigate soil, especially for archeological prospection. However, its success is less stable since growth differences must be sharply developed within one crop. Differentiation in growth patterns is influenced by a combination of growth factors such as the crop type, climate, soil moisture deficit and a range of physical and chemical soil properties (Evans, 1990).

2.2.2.2 Proximal soil sensors

The term 'proximal sensing' was first introduced in 1998 to describe devices that collect data from a distance that is in close proximity (1 m or less) to the object of interest (Viscarra Rossel and McBratney, 1998). Proximal soil sensors are able to gather soil data intensively, as such soil properties may be determined almost continuously. A number of proximal soil sensors have been developed for use in a mobile configuration, allowing on-the-go soil measurements.

Electrical and electromagnetic sensors

Electrical and electromagnetic sensors use electrical circuits to determine the ability of soil particles to conduct or accumulate electrical charge. Since the soil becomes part of the circuit, its physical and chemical properties can affect the

measured electric parameters. The ability of the soil to conduct electricity is usually quantified by electrical resistivity (ER) or electrical conductivity (EC). Both values are related to the voltage and electric current ratio for a known configuration of transmitting and receiving electrodes.

In case of measurement of ER, an invasive galvanic or direct-current (DC) electrode method consisting of coulter electrodes that are inserted into the soil measures the resistance to the flow of an electric current through the soil. Figure 2.3(a) is a schematic diagram showing the basic principle of resistivity measurements. An electrical current (I) is introduced into the soil through coulter electrodes (A and B) at the soil surface, and the difference in current flow potential (or voltage U) due to the resistance of the soil is measured by two additional receiver electrodes (M and N). The electrode configuration is referred to as a Wenner array when the four electrodes are equally spaced in a straight line at the soil surface, with the outer electrodes being the current transmitters and the inner electrodes the receiving electrodes (Corwin and Lesch, 2003). The depth of investigation is a function of the electrode spacing. The greater the spacing between the outer current electrodes, the deeper the electrical currents will flow into the soil, hence the greater the depth of exploration. Therefore, if more than two electrodes are used multiple depths can be explored simultaneously. For example, Panissod et al. (1997) developed the Multidepth Continuous Electrical Profiling (MUCEP) system which is able to explore three depths of investigation simultaneously (Fig. 2.3(b)).

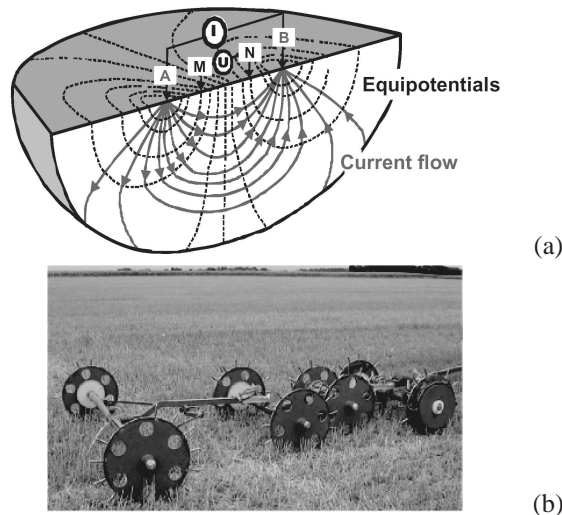


FIGURE 2.3 – (a) Basic principle of galvanic electrical resistivity measurements and (b) illustration of the MUCEP system.

The measured electrical resistance (R) is the ratio of the voltage difference and the flow of the current between the two electrodes. The resistivity, ρ [Ohm.m], is the resistance standardized over the spacing between the electrodes (a):

$$\rho = 2\pi a R \quad (2.1)$$

Alternatively, this can be expressed as conductivity since soil conductivity (σ) is the reciprocal of its resistivity:

$$\sigma = \frac{1}{\rho}. \quad (2.2)$$

Recently an overview of the use of ER in soil science was given by Samouelian et al. (2005). DC soil sensors are produced by VERIS technologies, Inc. (Salina, Kansas, USA), Geocarta (Paris, France) and Crop Technology, Inc. (Bandera, Texas, USA).

Non-contact EC measurements on the other hand are based on electromagnetic induction (EMI). The basis concept of EMI is that any time-varying electromagnetic field will cause current to flow in any (semi)conductive object that it encounters. An EMI soil sensor uses wound coils to transmit and receive electromagnetic fields and the principle is illustrated in Fig. 2.4(a). These coils can be treated as magnetic dipoles since the distance between the transmitter and receiver is more than several coil diameters. A transmitter coil (Tx) energized with an alternating current irradiates a low (radio) frequency magnetic field (Ampere's law): the primary field (H_p). As described by Faraday, this magnetic field H_p induces eddy current loops in the soil, proportional to the conductivity of the soil in the vicinity of the loops. Each current loop generates a secondary magnetic field (H_s) that is proportional to the current flowing through that loop. The primary and secondary magnetic field are superimposed on each other and the resulting field is measured by the receiver coil (Rx). The amplitude and phase of the secondary field differs from the primary as a result of soil properties, spacing of the transmitter and receiver coils and their orientation, the frequency and the distance from the soil surface (Hendrickx and Kachanoski, 2002). The distance between the two coils (intercoil spacing) and their orientation also determines the effective measurement depth. The resulting secondary field can be decomposed into two components: one part is in phase with the primary field and the other is 90° out-of-phase (quadrature) with the primary field.

An important issue for EMI based soil sensors is the electrical skin depth (δ). The skin depth is the depth at which the primary magnetic field has been attenuated to e^{-1} (37%) of its original strength at the surface of the soil. The induction number (β) is deduced from this skin depth and defined as the ratio of the intercoil spacing

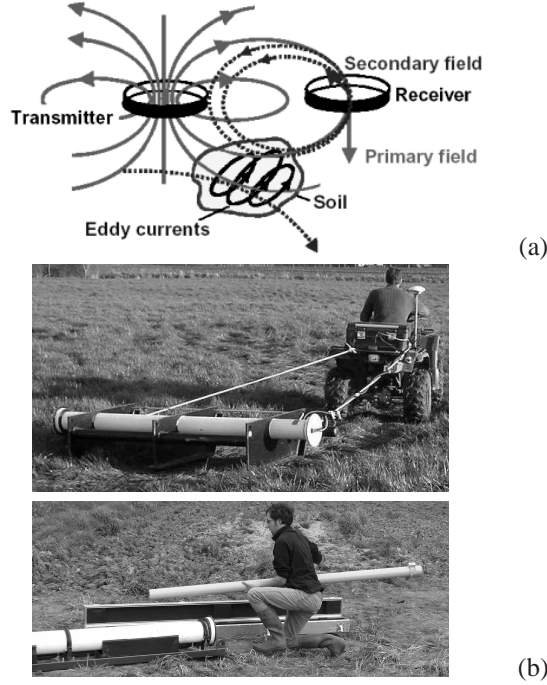


FIGURE 2.4 – (a) Basic principle of electromagnetic induction and (b) illustration of the mobile DUALEM-21S sensor configuration with below a photograph of the DUALEM-21S sensor outside its sled.

to the skin depth (McNeill, 1980b):

$$\beta = \frac{s}{\delta} = \frac{s}{\sqrt{\frac{2}{\sigma\omega\mu_0}}} \quad (2.3)$$

where s = intercoil spacing [m]

$$\omega = 2\pi f$$

$$f = \text{frequency [Hz]}$$

$$\mu_0 = \text{permeability of free space} = 4\pi \cdot 10^{-7} \text{ H m}^{-1}$$

$$\sigma = \text{soil conductivity [mS m}^{-1}\text{]}.$$

Soil conductivity sensors are designed to work with a skin depth that is much larger than the intercoil spacing, resulting in an induction number less than unity. For sensors with a fixed f and s , the constraint of $\beta \leq 1$ holds true for soils where $\sigma \leq 100 \text{ mS m}^{-1}$. Under this condition (known as operating at low induction numbers), the conductivity measured by the soil sensor is proportional to the ratio of the out-of-phase component of the secondary field to the primary field (McNeill,

1980b):

$$\sigma = \frac{4}{\omega \mu_0 s^2} \frac{H_s}{H_p} \quad (2.4)$$

where H_s = secondary magnetic field at the receiver coil [H m^{-1}]

H_p = primary magnetic field at the receiver coil [H m^{-1}].

The primary magnetic field also interacts with the soil magnetic properties, which is measured by the in-phase component of the secondary field. The ratio of the in-phase component of the secondary field to the primary field is related to the magnetic susceptibility (MS, expressed in ppt) of the soil. The MS of the soil is related to the concentration of Fe-oxides, mainly magnetite and maghemite, in the soil, which can be altered through biological decay and burning, enhancing the magnetic susceptibility of the soil. Mostly MS is used for archaeological prospecting by the detection of magnetic anomalies produced by archaeological structures (Simpson et al., 2008). Commercially available EMI sensors are produced by Geonics Ltd. (Mississauga, Ontario, Canada) and DUALEM, Inc. (Milton, Ontario, Canada) (Fig. 2.4(b)).

In both cases, the measured EC or ER is called ‘apparent’, since the soil is rarely uniform and the exact sample size cannot be determined. It is the electrical conductivity or resistivity of a homogeneous half-space, averaged to the depth of investigation, that would yield the observed instrument response. The measured EC or ER values are affected by more than one physical or chemical soil property: soil salinity, clay content and mineralogy, cation exchange capacity (CEC), soil pore size and distribution, soil moisture content and temperature (McNeill, 1980a; Mueller et al., 2003; Samouelian et al., 2005). Since both type of measurements relate to the same range of properties, the selection of an EC_a sensing system should be based on both practical issues and the intended use of the data (Sudduth et al., 2003a).

Alternatively, capacitively coupled resistivity (CCR) systems are used to measure ER (Allred et al., 2006). They employ capacitive coupling through coaxial cables to introduce electric current into the ground. The metal shield of the coaxial cable acts as one capacitor plate, the soil surface is the other capacitor plate and the outer insulation of the coaxial cable is the dielectric material separating the two plates. Like with DC systems a dipole-dipole electrode array is used with two transmitter coaxial cables incorporated into the transmitter dipole and two receiver coaxial cables into the receiver dipole. At the transmitter dipole, an alternating current passes from the cable to the soil through capacitance, while at the receiver dipole the current starts in the soil, charges up the capacitance of the coaxial cable where the voltage generated by the current flowing in the soil is measured. Again the spacing between the two dipoles determines the investigation depth, given that

the dipole lengths are constant. A commercialised CCR system is the OhmMapper produced by Geometrics, Inc. (San Jose, California) (Fig. 2.5).

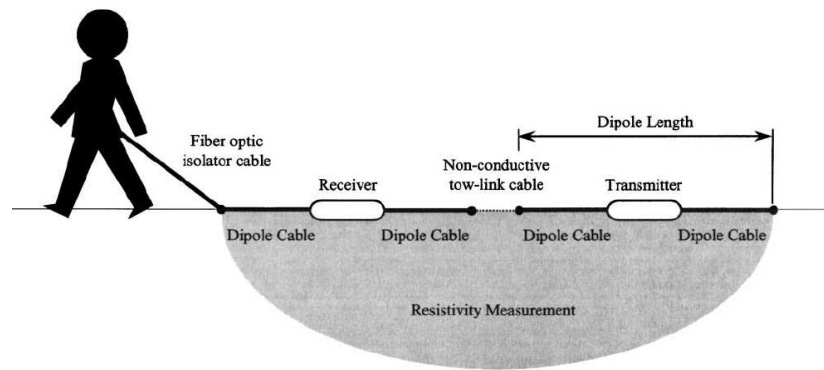


FIGURE 2.5 – Schematic of the Geometrics, Inc. Ohmmapper CCR system (Walker and Houser, 2002).

Optical and radiometric sensors

Spectral reflectance imagery is a very promising technology for rapid measurement of many physical and chemical soil properties. These sensors measure the absorbed, reflected and transmitted radiation of soil particles actively illuminated by a source with known spectral characteristics (in contrast to the passive sensors used in remote sensing). The NIR (0.7 - 3 μm) and MIR (3 - 25 μm) spectral regions seem to be the most informative since most soil constituents have their molecular frequencies in this range. The results are complex absorption patterns and the information should be mathematically processed to be correlated with soil properties. Until now, research focused on spectral calibration and prediction of soil properties using multivariate statistics like principal component analysis, partial least squares, neural networks, etc.

A review of the soil properties predicted by a multivariate analysis of the spectral response in different regions of the EM spectrum is given in Viscara Rossel et al. (2006). Soil organic matter (OM), soil moisture, CEC and the soil textural fractions have been predicted using spectral reflectances. A major advantage of diffuse reflectance spectroscopy for soil analysis is that from a single spectrum many properties may be (accurately) determined. However, NIR-MIR spectrometry methods are invasive and measure only at a specified operational soil depth (typically in the plough layer). The potential of this type of sensor is largest in precision agriculture. Besides, most spectral reflectance analyses are based on measurements in the laboratory. Only recently results concerning mobile measurements have been published (Mouazen et al., 2007) and a commercial mobile VIS-

NIR spectrophotometer has been launched by Veris Technologies, Inc. (Salina, Kansas, USA) (Fig. 2.6(a)). Some problems concerning portable spectrometers involve instability due to variable distances between sensor and soil, dust on the probe, shakiness of the vehicle and changing environmental factors like soil water content. Also the calibrations are not straightforward, limiting at the moment the wide use of this technique (Adamchuk et al., 2004).

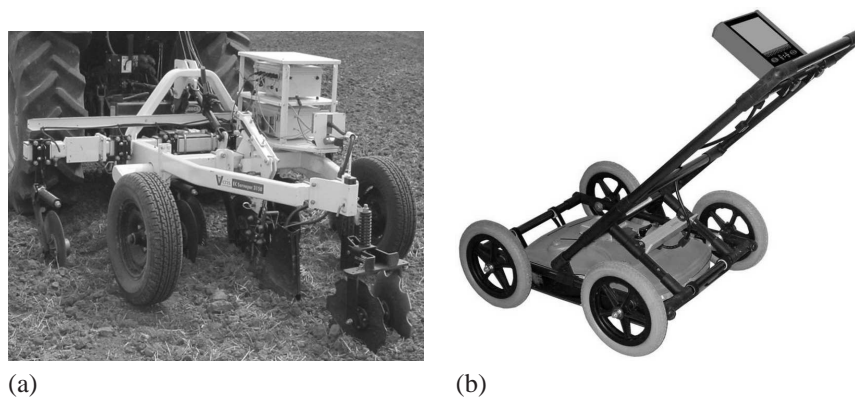


FIGURE 2.6 – Examples of (a) the VERIS VIS-NIR sensor and (b) a GPR

Radio EM waves are also applied in soil science; ground penetrating radar (GPR) is a non-invasive geophysical sensor used to predict soil texture and organic matter, differences in soil compaction, depth to soil horizons and lithologic discontinuities (Fig. 2.6(b)). The term radar is an acronym for RAdio Detection And Ranging. GPR usually consists of two antennas: a transmitter radiates high frequency pulses of radio waves and a receiver detects the reflected EM waves as a function of time (Davis and Annan, 2002). The reflected energy is recorded as a pattern on a radargram which should be processed and analyzed using specialized software. The depth range of GPR is limited by the electrical conductivity of the ground and the transmitting frequency. As conductivity increases, the penetration depth also decreases. This is because the radar energy is more quickly dissipated into heat, restricting the investigation depth. Higher frequencies do not penetrate as far as lower frequencies, but give better resolution. Generally, depending on the antenna used and the chemistry of the soil material, observation depths are 5 to 30 m in sandy soils, 1 to 5 m in loamy soils and less than 0.5 m in clayey soils (Doolittle and Collins, 1998). As a consequence EMI sensors have been used as a precursory tool to guide the more time-consuming GPR measurements (Inman et al., 2002).

At the other end of the EM spectrum, gamma ray spectrometry (GRS) detects the natural radioactive decay of isotopes of several elements, mainly potassium, thorium and uranium since they have long half-lives, are abundant in the environment and produce γ -rays of sufficient energy and intensity to be measured by γ -ray spectrometry (Viscarra Rossel et al., 2007). The emitted γ radiometric signal can be related to soil mineralogy and texture, making it very suitable to measure the clay content after a careful calibration with respect to the local geology and geomorphology. Due to a strong attenuation by soil components the operational depth of GRS is limited to the upper 0.5 m. Both remote and proximal GRS have been used, but better relationships with soil properties were found for proximal sensors.

Mechanical sensors

Mechanical sensors can be used to estimate soil mechanical resistance. High mechanical resistance in the soil may arise naturally, be caused by compaction from heavy machinery or by the formation of plow pans. In all cases, soil particles are positioned closer to each other, a process called compaction. The most common technique to measure soil compaction is the field measurement of bulk density and penetration resistance under static conditions, using core sampling methods and penetrometers respectively (Mouazen and Ramon, 2006). Indirectly, soil compaction is referred to as on-line measured soil mechanical resistance (or draught) to different penetration tools using load cells or strain gauges (Adamchuk et al., 2001; Hemmat and Adamchuk, 2008; Verschoore et al., 2003). These mechanical sensors use a mechanism that penetrates or cuts through the soil, and records the force measured by strain gauges or load cells (Fig. 2.7). Despite a lot of research, no commercial available system has been developed yet.

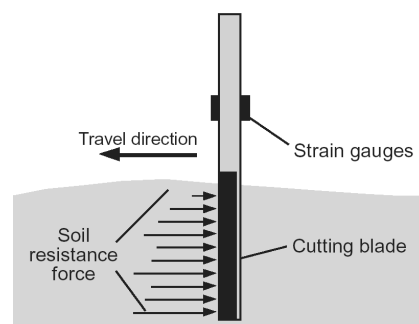


FIGURE 2.7 – Principle of a mechanical soil sensor

Acoustic and pneumatic sensors

Acoustic sensors are based on reflection and transmission in the low frequency (audio) range to determine surface air porosity, air permeability and pore structure. The change in the level of noise caused by an interaction of soil particles is measured. Pneumatic sensors measure the pressure required to force a given volume of air into the soil and compare it with air permeability (Clement and Stombaugh, 2000). Based on this principle changes in soil structure and compaction can be detected. Both types serve as an alternative for mechanical sensors to measure soil compaction (Sabatier et al., 1990). Nevertheless, since the relation between the sensor output and the physical state of the soil is still poorly understood, additional research is needed.

Electrochemical sensors

Electrochemical sensors are capable of directly measuring the soil nutrient availability and pH through either an ion-selective electrode (ISE) or an ion-selective field effective transistor (ISFET). In both cases, the measured voltage (potential difference) between the sensed and referenced parts of the system is related to the concentration of specific ions (H^+ , K^+ , NO_3^- , etc.). The potential of using ISE is highest for pH, followed by K^+ and then NO_3^- (Adamchuk et al., 2005). On-the-go measurements of these ions always involve a soil sampling mechanism followed by chemical extraction and measurement of the ion concentration through interfacing with the sensor. Before a new sample is obtained, the electrode surfaces are rinsed with water. A major drawback is that a significant amount of stabilization time to reach equilibrium with the measured soil or soil solution is required. Besides there is a time lag between sample collection and sensor output. Nevertheless a mobile pH system using two ISEs has been developed, tested (Adamchuk et al., 1999) and commercialized (Veris Technologies Inc., 2003). Direct measurement of the ion activity is still the topic of considerable research (Adamchuk et al., 2007; Birrell and Hummel, 2001; Sethuramasamyraja et al., 2008).

Of all these types of sensors, the most intensively used sensors in soil science are based on DC or EMI methods, measuring the soil EC_a . Quite some research studies have compared both methods. Mostly similar EC_a patterns were found (Allred et al., 2006; Buchleiter and Farahani, 2002; Doolittle et al., 2002; Sudduth et al., 2001). Nevertheless each method has its advantages and drawbacks making them suitable for surveys under certain conditions (Dabas and Tabbagh, 2003). Concerning EMI sensors, the major advantage is the easiness of collecting mobile non-invasive measurements regardless the conditions of the soil (dry or stony soils, crop growth on the field). DC sensors on the other hand can easily change the depth and volume of measurement by changing the spacing between the electrodes.

2.3 Factors influencing the soil apparent electrical conductivity

Three pathways of current-flow contribute to the EC_a of the soil (Fig. 2.8): (i) a solid-liquid phase pathway through alternating layers of soil particles and soil solution, this conductance is primarily achieved via the exchangeable cations associated with clay minerals, (ii) a liquid phase pathway traveling via dissolved solids contained in the soil water occupying the large pores, and (iii) a solid pathway traveling through or along the surface of soil particles in direct and continuous contact (Rhoades et al., 1989; Lesch and Corwin, 2003). Because of these three pathways involving ion displacement through the volume of water in the pores or at the surface of particles (cations), the soil EC_a is influenced by several chemical and physical properties: soil salinity, saturation percentage, water content and bulk density (Corwin, 2005; Corwin and Lesch, 2005a). Saturation percentage and bulk density are directly related to clay content and organic matter. Furthermore, the negatively charged surface sites on clay minerals and organic matter provide the solid-liquid phase primarily via exchangeable cations.

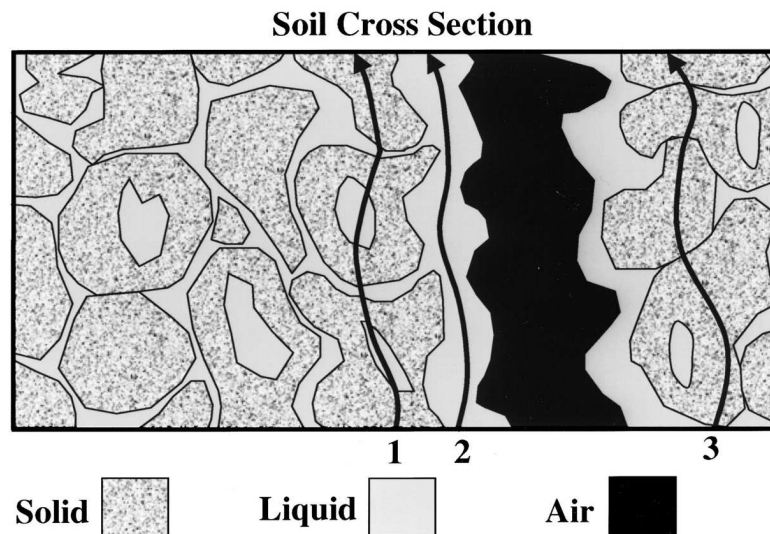


FIGURE 2.8 – Electrical conductivity pathways: (1) solid-liquid, (2) liquid and (3) solid. Modified from Rhoades et al. (1989).

Besides, also soil temperature is an EC_a influencing factor since the electrical conductivity increases at a rate of approximately 1.9% per °C increase in temperature. For comparison reasons, the measured EC_a can be expressed at a reference temperature of 25 °C using following equation (Sheets and Hendrickx, 1995):

$$EC_{25} = EC_a \left\{ 0.24470 + 1.43034e^{-\frac{t}{26.615}} \right\} \quad (2.5)$$

where EC_{25} = temperature standardized EC_a

t = soil temperature in °Celsius.

For a given soil temperature, it can be concluded that the soil EC_a is influenced by the soil pore configuration, the pore water content and its ionic composition. In absence of salts this means that clay and organic matter will have the major influence.

Nevertheless when only within-field measurements are concerned, the absolute EC_a values are not of major interest. It are the relative changes in the soil EC_a that indicate the soil's spatial variability.

In general, the magnitude and spatial heterogeneity of EC_a in a field is dominated by one (or two) of these factors. Its within-field changes are large enough with respect to the variation of the other EC_a influencing factors that the EC_a can be calibrated as a direct measurement of that dominant factor. In case where dynamic soil properties (salinity, water content and temperature) dominate the EC_a measurements, the spatial patterns exhibit more temporal changes, than systems dominated by static factors (texture) (Corwin, 2005). In the latter case, the spatial patterns remain consistent since the dynamic properties only affect the magnitude of the measured EC_a .

2.4 The EM38DD soil sensor

2.4.1 Background on the EM38 sensor

The EM38 sensor manufactured by Geonics Ltd. has a spacing of 1 m between the transmitting coil (Tx) at one end of the instrument and the receiver coil (Rx) at the other end. In Figure 2.9 the EMI principle (section 2.2.2.2) is schematically illustrated. The strength of the current loops generated by the H_p depends on the electrical conductivity of the soil.

The sensor operates at a frequency of 14.6 kHz and measurements can be made with the instrument dipoles either in the horizontal or vertical orientation. The measured EC_a will be denoted as EC_a -H, EC_a -V respectively. In the horizontal orientation (or dipole mode) both transmitter and receiver coils are oriented parallel to the earth's surface, while in the vertical orientation both coils are perpendicular to the surface. Each orientation has its own depth-response profile, as illustrated in Fig. 2.10. The relative response curves (Fig. 2.10(a)) describe the relative contribution to the secondary magnetic field arising from a thin horizontal layer at any normalized depth z (with z being the depth divided by the intercoil spacing) (McNeill, 1980b). Since the EM38 sensor has an intercoil spacing of 1 m,

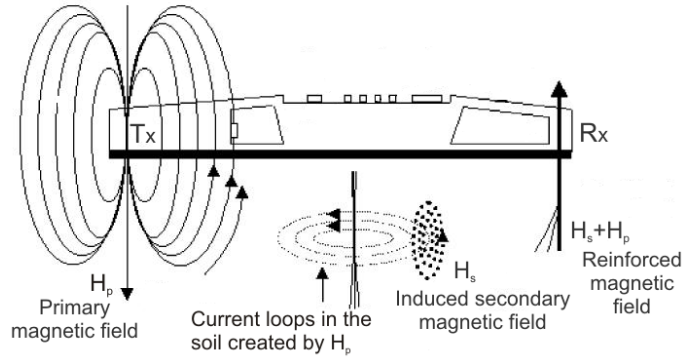


FIGURE 2.9 – Schematic diagram of the EM38 showing the principle of electromagnetic induction. Adapted from Robinson et al. (2004).

the normalized and actual depth are the same. The response peaks between 0.3 and 0.5 m depth for EC_a-V, while the surface soil only makes a very small contribution to the secondary field. At 1.5 m depth the soil still contributes significantly. The response curve of the horizontal orientation shows a completely different behavior with the largest contribution from the near surface soil, monotonically decreasing with depth. For a homogeneous halfspace, the equations governing the relative response $\phi(z)$ of the EM38 are (McNeill, 1980b):

$$\phi_V(z) = \frac{4z}{(4z^2 + 1)^{\frac{3}{2}}} \quad (2.6)$$

$$\phi_H(z) = 2 - \frac{4z}{(4z^2 + 1)^{\frac{1}{2}}} \quad (2.7)$$

By integrating the relative response between the surface and a depth z , the cumulative response $R(z)$ from the soil surface up to a depth z is given by (adapted from McNeill (1980b)):

$$R_V(z) = 1 - \frac{1}{(4z^2 + 1)^{\frac{1}{2}}} \quad (2.8)$$

$$R_H(z) = 1 - (4z^2 + 1)^{\frac{1}{2}} + 2z. \quad (2.9)$$

The cumulative response (expressed as % of the measured signal) is defined as the relative contribution to the secondary magnetic field from all material above a certain depth z (Fig. 2.10(b)). Since the cumulative graphs are quasi exponential, an effective depth of exploration (DOE) was defined as the depth from which 70% of the response comes (or the depth below which the contribution is only 30%) and for the vertical and horizontal orientation respectively the effective DOE is 1.6 m and 0.75 m (Abdu et al., 2007; Corwin and Lesch, 2005b). Nevertheless, for the

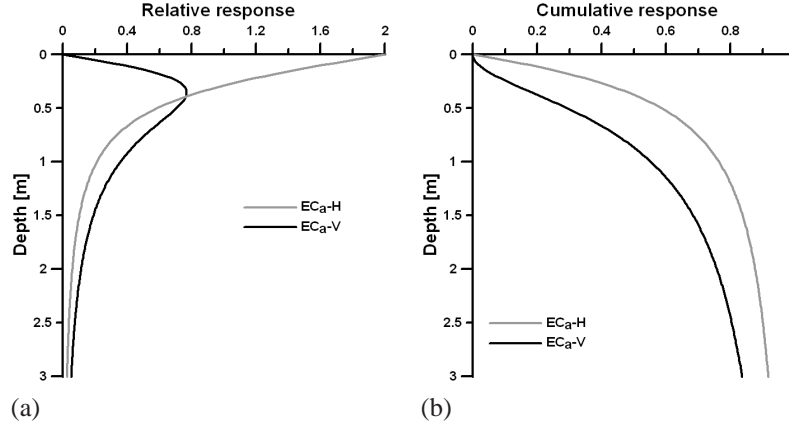


FIGURE 2.10 – The response curves as a function of the depth for the two orientations of the EM38 sensor: (a) the relative response and (b) the cumulative response.

vertical coil orientation the soil layers below a depth of 2 m still contribute 25% to the secondary magnetic field, while in the horizontal orientation this contribution is only 12%.

For a N -layered earth, the reading of the EM38 sensor in vertical (EM_V) and horizontal (EM_H) orientation is (Slavich, 1990):

$$EM_V = \sum_{i=1}^{N_v} EC_{ai}(R_{V(i-1)} - R_{Vi}) \quad (2.10)$$

$$EM_H = \sum_{i=1}^{N_h} EC_{ai}(R_{H(i-1)} - R_{Hi}). \quad (2.11)$$

where EC_{ai} is the mean EC_a of the i^{th} depth layer, R_{Vi} and R_{Hi} are the cumulative response components of the i^{th} depth layer for the horizontal and vertical orientation respectively, N_v is the number of layers till the vertical depth of measurement and N_h the number of layers till the horizontal depth of measurement. Since the EC_a readings represent the average conductivity of the different layers within the depth of investigation, the measurements are referred to as depth-weighted or bulk EC_a .

The operation of the EM38 is illustrated for two cases with a difference in topsoil depth, given the constraint that the topsoil has a lower clay content than the subsoil (Fig. 2.11). The EM38 induces horizontal electric current loops in the soil and the current in each loop is proportional to the electrical conductivity of the soil which is in Fig. 2.11 shown by the thickness of the ellipses. The individual currents are weighed as a function of the depth and summed to generate

the instrument response. The response is larger when more of the soil profile has a higher conductivity as in the left case of Figure 2.11.

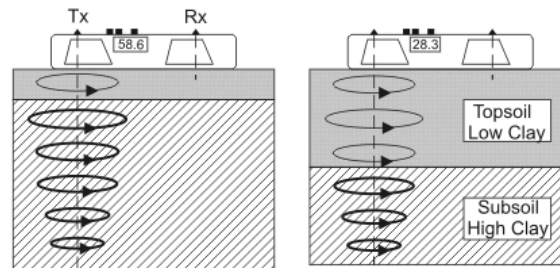


FIGURE 2.11 – Principle of operation of the EM38 (in vertical orientation). Adapted from Sudduth et al. (2001).

2.4.2 Introducing the EM38DD soil sensor

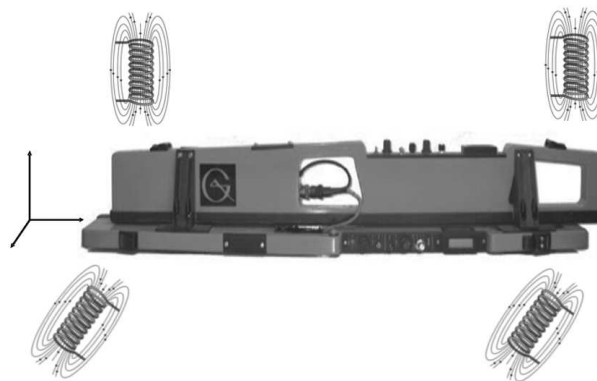


FIGURE 2.12 – The EM38DD with the receiver and transmitter coils of each orientation.

In 2001, a new sensor was launched in addition to a single EM38: the dual dipole EM38 or the EM38DD (Geonics Ltd.). This sensor actually consists of two EM38 sensors perpendicularly to each other which allows for simultaneous measurement of the EC_a in both vertical and horizontal orientation (Fig. 2.12). In the horizontal orientation both transmitter and receiver coils are oriented parallel to the earth's surface, while in the vertical orientation both coils are perpendicular to the surface. In the normal position (as in Fig. 2.12) the vertical unit is called the master, the horizontal unit the slave.

The EM38 soil sensor was used from the 1980's on resulting in a considerable amount of publications: in the beginning soil salinity hazards were addressed, but with time also agricultural applications were explored. In contrast, the EM38DD soil sensor was only recently developed and in 2003, the time this thesis work started, no scientific publications covered research with an EM38DD soil sensor. Obtaining dual EMI based signals of a soil is a next step to meet the needs of an evolving soil information era. New challenges related to the optimal interpretation and/or integration of the dual signals arose herewith.

2.4.3 Calibration of an EM38DD

Before starting a field survey the EM38DD needs to acclimatise due to the temperature effect on the readings and a calibration is required. The calibration was done according to the steps described in the EM38DD Operating Manual (Geonics Limited, 2001):

1. Initial nulling

The transmitter coil is used to generate the time-varying primary magnetic field which induces the eddy currents in the soil, and the receiver coil measures both this strong primary magnetic field and the much smaller secondary field arising from the eddy currents. To facilitate the measurement of this small signal from the eddy currents, an internally generated signal is used to cancel or 'null' the large primary signal so that it does not overload the electronic circuit. To eliminate the response to the soil electrical conductivity, the instrument is calibrated in free space by lifting it to a height of 1.5 m (Fig. 2.13). With the EM38DD master unit in the horizontal position, the I/P mode is nulled first by adjusting manually the control knob. Then the mode is switched to the Q/P mode and this mode is nulled. When returning to the I/P mode the value must stay zero. The nulling phase ends when in both modes a zero is kept.

2. Instrument zeroing

This adjustment is used to set the instrument zero so that if the unit were taken to a great height above the earth it would actually read zero. With the instrument at the height of 1.5 m and the master unit still in the horizontal position, the Q/P mode is adjusted to read approximately $50 \text{ mSm}^{-1}(\text{H})$. After rotating the master unit in vertical orientation the difference of this Q/P reading (V) and the one in horizontal position (H) should be adjusted to the master unit again in horizontal orientation ($\text{H}' = \text{V} - \text{H}$). When turning then the master unit back in vertical orientation (V') the reading must be twice the value of the horizontal orientation ($\text{V}' = 2\text{H}'$). This relationship applies only for a uniform or horizontally layered soil and can be deducted from

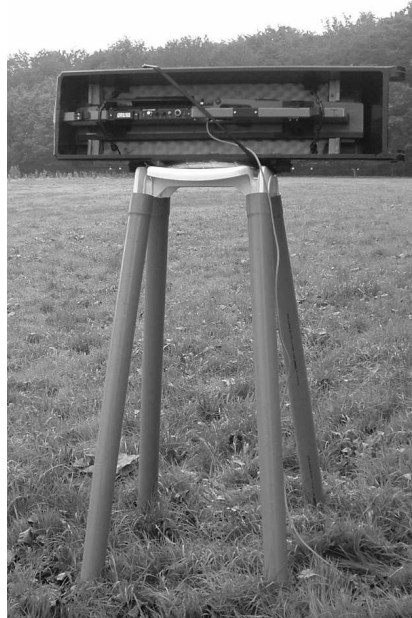


FIGURE 2.13 – Calibration of the EM38DD.

the cumulative response at a depth of 1.5 m (Fig. 2.10(b)). In the vertical orientation material below 1.5 m contributes 32% to the response while the same soil volume contributes 16% to the horizontal reading. If this relation is not met, the zero was not correctly set and the steps should be repeated until V' equals $2H'$.

3. Final check-up

Finally it must be checked if both units (master and slave) give the same reading when they are in the same orientation (horizontal or vertical), with the sensor placed on the ground to read the soil's electrical conductivity. When this holds true, the calibration is complete.

Drift of the EM38DD reading can occur and should be accounted for. The sources of this drift was the topic of several studies. Diurnal changes of the soil temperature did not have a significant effect on soil EC_a readings (Brevik et al., 2004). Ambient conditions, dominantly the air temperature, are known to affect the response of an EM38 sensor, especially the in-phase reading. However, experiments performed by Sudduth et al. (2001) were not able to relate changes in the air temperature to drift of the EM38 reading. Drift occurs rather due to instrument instability. Also Robinson et al. (2004) showed that at operating temperatures lower than 40°C the drift effect is minimal. To compensate for this drift effect, a cali-

bration transect should be used to monitor the drift during the field survey. Also putting the sensor in a well isolated sled will minimize the effect of temperature changes.

2.4.4 Mobile EM38DD soil survey

After calibrating the EM38DD sensor, the actual survey can start. Originally, the sensor was developed as a hand carrying device with an external carrying strap. In this way, the EC_a readings can be logged to a field computer, or recorded on paper. For agricultural purposes however, a mobile survey is preferred since it increases the capacity of operation. At the Research Group of Soil Inventory Techniques a mobile survey system was used, as illustrated in Fig. 2.14.

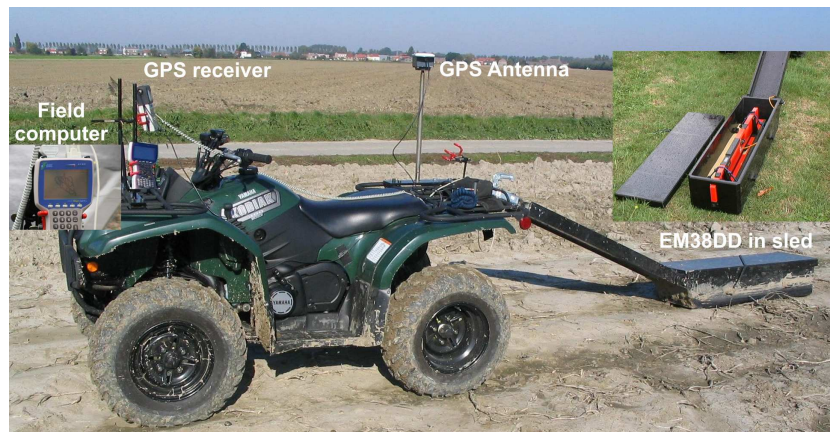


FIGURE 2.14 – Mobile configuration of an EM38DD soil survey.

The EM38DD sensor was put in a polyethylene sled which is pulled by an all terrain vehicle (ATV). A global positioning system (GPS) system (Magellan Pro-mark Systems Corporation, USA) records the geographical position of the measurements with an accuracy of 3 m. The coordinates recorded in WGS84 coordinate system were projected to the Belgian Lambert72 coordinate system. In the field computer (type Allegro CX, Juniper Systems Inc.) both the EC_a readings (horizontal and vertical) and the geographical position of the measurement were logged through specific software (HGIS - Starpal Inc, Colorado, USA). This software allows to specify the frequency of data acquisition: every second or at sub-second intervals. For characterizing the soil variability, fields are surveyed with a driving speed of 15 km h^{-1} along parallel transects separated by a predefined distance in the order of 5 m. At this speed and with a measurement every second, the measurement interval within a line is (approximately) 4 m. This measurement resolution allows geostatistical analyses, as explained in Chapter 3.

3

Analyzing soil sensor data

3.1 Geostatistical methods: from data points to a map

3.1.1 Introduction

In chapter 2 the ways to measure soil variability directly or indirectly were discussed. Nevertheless neither with an augur nor with a soil sensor, a continuous area can be sampled, rather a finite number of locations is sampled. Spatial prediction is the process of estimating the values of a target variable at the unvisited locations. When applied to a whole study area, it is also referred to as spatial interpolation or mapping (Hengl et al., 2004). Geostatistics has been applied widely in soil science to solve the problem of estimating soil properties at unvisited locations from limited sample data (Webster and Oliver, 2001). The term ‘geostatistics’ refers to the statistical analysis of phenomena which vary in a spatial (‘geo-’) continuous way. Georeferenced observations are almost always correlated to some degree in relationship with the distance between the observations. This corresponds to the intuitive feeling that places close to one another tend to have more similar values than the ones further apart. Using geostatistics this spatial dependence can be expressed quantitatively and used for spatial prediction, a procedure called kriging.

3.1.2 Variography: modeling spatial variation

Firstly, the spatial variation between the observations should be characterized. Geostatistics views soil properties as continuous variables and models these as realizations of a random function or a random process (Webster, 2000). To characterize a random function assumptions of stationarity are required and in spatial statistics these assumptions are limited to the intrinsic hypothesis (Matheron, 1973). The intrinsic hypothesis assumes stationarity of the first- and second-order moments of the increments $[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})]$ of the random function, where \mathbf{x} is the location vector and \mathbf{h} is a spatial lag. This implies that the expected value of the increments $[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})]$ is zero and their variance exists and both are independent of the position \mathbf{x} . These two conditions are expressed as follows:

$$E[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] = 0 \quad \forall \mathbf{x} \quad (3.1)$$

$$\text{Var}[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] = E[(Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x}))^2] = 2\gamma(\mathbf{h}) \quad \forall \mathbf{x} \quad (3.2)$$

where $\gamma(h)$ is the semivariance, mostly called the (semi)variogram.

The semivariance measures the average dissimilarity between data separated by vector \mathbf{h} and can be calculated based on a series of observations $z(\mathbf{x}_\alpha)$ (Goovaerts, 1997):

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [z(\mathbf{x}_\alpha + \mathbf{h}) - z(\mathbf{x}_\alpha)]^2 \quad (3.3)$$

where $\hat{\gamma}(\mathbf{h})$ is the calculated semivariance for a lag vector \mathbf{h} between observations $z(\mathbf{x}_\alpha)$ and $z(\mathbf{x}_\alpha + \mathbf{h})$ and $N(\mathbf{h})$ the number of pairs of observations separated by \mathbf{h} .

Ideally pairs of observations closer together should show a smaller semivariance, whereas pairs of observations farther away from each other should display a larger semivariance. A plot of the calculated $\hat{\gamma}(\mathbf{h})$ versus \mathbf{h} yields the experimental variogram to which a theoretical model is fitted (Fig 3.1). Four of the most common variogram models are the spherical, the exponential, the gaussian and the linear model (Burrough, 1993). Unlike the first three model types, the linear model describes an unbound variogram meaning that the variogram increases with increasing lag distance. Bounded variograms are characterized by three parameters which describe the spatial variance across the study area: the nugget variance (C_0), the sill variance ($C_0 + C_1$) and the range (a) (Fig 3.1). In theory, the semivariance at $\mathbf{h} = 0$ is zero, but it is often found that as the lag distance approaches zero, the semivariance remains a positive value, called the nugget. The nugget is the intercept of the variogram with the Y-axis and represents unexplained spatially dependent variation (micro variability at distances closer than the smallest sampling lag) or purely random variance (like measurement or sampling error). The semivariance increases with increasing lag until it stabilizes to a maximum, the sill variance. The lag distance at which the sill is reached is called the range. The sill is the a priori sample variance σ^2 and the range represents the limit of the spatial dependence since at that distance the autocorrelation becomes zero. Beyond the range, the expected difference between two observations is maximum (equalling the sill) and independent of the lag distance between them. At distances smaller than the range there exist a spatial dependence between two observations which increases with an increasing lag distance. So, the variogram describes the pattern of spatial variability in terms of its magnitude, scale and general form (Oliver, 1987).

The different bounded variogram models are illustrated in Figure 3.2 and they are described by the following equations:

- Spherical model

$$\begin{cases} \gamma(h) = C_0 + C_1 \left[\frac{3h}{2a} - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right] & \text{for } h \leq a, \\ \gamma(h) = C_0 + C_1 & \text{for } h > a. \end{cases} \quad (3.4)$$

A spherical model has a linear start up to about 2/3 of the range, beyond which it levels off abruptly to the sill value. Spherical models describe variables with abrupt boundaries at discrete and regular spacings but without a well defined distance between the abrupt changes.

- Exponential model

$$\gamma(h) = C_0 + C_1 \left[1 - \exp\left(\frac{-3h}{a}\right) \right] \quad (3.5)$$

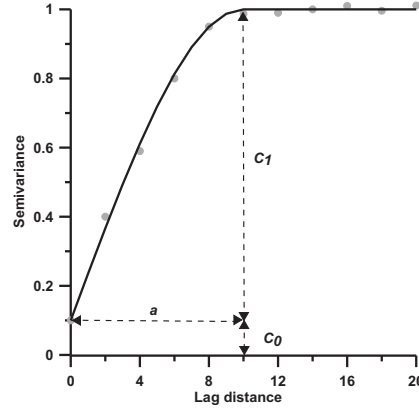


FIGURE 3.1 – An experimental variogram (grey dots) fitted by a spherical model (black line) with indication of the three variogram parameters ($C_0 = 0.1, C_1 = 0.9, a = 10$).

Variables characterized by abrupt changes at all distances are described by an exponential model. Exponential models have a linear start up to $1/3$ of the range, but they start steeper than a spherical model. Since the sill is reached only asymptotically there is no finite range. For practical purposes, an effective range is defined as the h value at which 95% of the sill is reached.

- Gaussian model

$$\gamma(h) = C_0 + C_1 \left[1 - \exp\left(\frac{-3h^2}{a^2}\right) \right] \quad (3.6)$$

A Gaussian model describes continuous, gradually varying variables. Gaussian models show an almost horizontal behaviour at the start (parabolic start) after which they increase steeply to bend over to the sill. Also here the same definition of the effective range is valid. The initial flat behaviour represent homogeneities over very short distances, like can be found in geological deposits or digital elevation models.

Sometimes variograms appear to be complex and more complex functions are needed to describe them. The best way to do this is by combining two or more simple models not necessarily of the same type, generating a nested model. A nested variogram comprises more than one variogram structure and the individual variogram models are added to generate the nested model. Nested models are used to describe complex patterns acting over multiple spatial scales.

Throughout the chapters, the fitting of the variograms was done in VARIOWIN 2.2. The best fit was obtained with the 'Indicator Goodness of Fit' parameter as a guide (Pannatier, 1996).

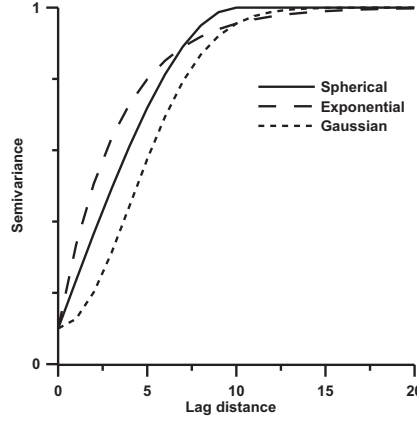


FIGURE 3.2 – Bounded variogram models with the same variogram parameters ($C_0 = 0.1, C_1 = 0.9, a = 10$).

3.1.3 Ordinary kriging

Kriging aims to estimate the value of a variable Z at any unsampled location \mathbf{x}_0 based on a set of observations of $z(\mathbf{x}_\alpha)$ ($\alpha = 1, \dots, N$) using a weighted linear combination of these observations. Ordinary kriging (OK) is the most common type of kriging and assumes the mean of the observations to be unknown but locally stationary. The ordinary kriging estimator $Z^*(\mathbf{x}_0)$ can be written as:

$$Z^*(\mathbf{x}_0) = \sum_{\alpha=1}^{n(\mathbf{x}_0)} \lambda_\alpha Z(\mathbf{x}_\alpha) \quad \text{with} \quad \sum_{\alpha=1}^{n(\mathbf{x}_0)} \lambda_\alpha = 1 \quad (3.7)$$

where $n(\mathbf{x}_0)$ is the number of observations in the local neighbourhood around \mathbf{x}_0 and λ_i are the weights assigned to each of these observations.

The interpolation problem is thus reduced to finding the appropriate weights λ_i . In geostatistics, the variogram model with its associated parameters is an essential tool in determining the weights. Also the best estimator must be unbiased and must have minimum variance which is expressed in two conditions:

- Condition of unbiasedness

$$E[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] = 0 \quad (3.8)$$

- Condition of minimum variance of estimation (kriging variance)

$$\sigma^2(\mathbf{x}_0) = E[\{Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)\}^2] = \text{minimum}. \quad (3.9)$$

Based on the these two conditions the OK system becomes (Goovaerts, 1997):

$$\begin{cases} \sum_{\beta=1}^{n(\mathbf{x}_0)} \lambda_{\beta} \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) + \varphi = \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_0) & \alpha = 1, \dots, n(\mathbf{x}_0) \\ \sum_{\beta=1}^{n(\mathbf{x}_0)} \lambda_{\beta} = 1 \end{cases} \quad (3.10)$$

where $\gamma(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta})$ is the semivariance between the observation locations \mathbf{x}_{α} and \mathbf{x}_{β} , $\gamma(\mathbf{x}_{\alpha} - \mathbf{x}_0)$ is the semivariance between the observation locations \mathbf{x}_{α} and the unsampled location \mathbf{x}_0 and φ is a Lagrange multiplier used to satisfy the restriction that the sum of the weights must equal one.

The kriging system, consisting of $(n+1)$ linear equations and $(n+1)$ unknowns (one Lagrange multiplier φ and $n(\mathbf{x}_0)$ weights), can be solved through matrix algebra. All variogram values are known through the modelled variogram. In matrix notation, the OK system becomes:

$$[\mathbf{A}][\lambda] = [\mathbf{B}] \quad (3.11)$$

where

$[\mathbf{A}]$ is the matrix of the variogram values between the observations:

$$[\mathbf{A}] = \begin{bmatrix} \gamma(\mathbf{x}_1 - \mathbf{x}_1) & \dots & \gamma(\mathbf{x}_1 - \mathbf{x}_{n(\mathbf{x}_0)}) & 1 \\ \gamma(\mathbf{x}_2 - \mathbf{x}_2) & \dots & \gamma(\mathbf{x}_2 - \mathbf{x}_{n(\mathbf{x}_0)}) & 1 \\ \vdots & \dots & \vdots & \vdots \\ \gamma(\mathbf{x}_{n(\mathbf{x}_0)} - \mathbf{x}_1) & \dots & \gamma(\mathbf{x}_{n(\mathbf{x}_0)} - \mathbf{x}_{n(\mathbf{x}_0)}) & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix}, \quad (3.12)$$

$[\lambda]$ is the matrix of the unknown weights and the Lagrange multiplier:

$$[\lambda] = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{n(\mathbf{x}_0)} \\ \varphi \end{bmatrix} \quad (3.13)$$

and $[\mathbf{B}]$ is the matrix of the variogram values between the observations and \mathbf{x}_0 :

$$[\mathbf{B}] = \begin{bmatrix} \gamma(\mathbf{x}_1 - \mathbf{x}_0) \\ \gamma(\mathbf{x}_2 - \mathbf{x}_0) \\ \vdots \\ \gamma(\mathbf{x}_{n(\mathbf{x}_0)} - \mathbf{x}_0) \\ 1 \end{bmatrix}. \quad (3.14)$$

The OK weights are obtained by multiplying the inverted $[\mathbf{A}]$ matrix with $[\mathbf{B}]$:

$$[\lambda] = [\mathbf{A}^{-1}][\mathbf{B}]. \quad (3.15)$$

This solution provides the weights λ_β and the Lagrange multiplier from which the estimator $Z^*(\mathbf{x}_0)$ can be found using Eq. (3.7). The procedure can then be repeated for every location until the study site is mapped. The endproduct of the mapping process is called a map. In this context, a map refers to a raster map or pixelplot and is defined as a regular array of pixel (grid cell) values. The number of locations to be estimated equals the number of pixels of the map and depends on the pixel resolution.

3.1.4 Validation of the prediction

To determine the accuracy of the predicted value a validation should be applied. Two common ways of validating are validation with an independent dataset and crossvalidation. The former, as the name suggests, requires an independent test data set, while the latter is based on the original dataset (size N) from which each time the observation point to be estimated is left out and the remaining observations ($N - 1$) are used to estimate that value. Once the interpolation is obtained the ignored observation point returns to the data set and the next observation is removed; this continues until all observations are estimated. This method is also known as the leave-one-out validation. Both methods generate a paired data set of size N with N being the size of the independent dataset or for cross validation it is the number of observations. Each pair consists of an observation $z(\mathbf{x}_\alpha)$ and an estimation $z^*(\mathbf{x}_\alpha)$ at the same location. To quantify the performance of the interpolation method several validation indices can be used:

- The mean estimation error (MEE)

$$\text{MEE} = \frac{1}{n} \sum_{\alpha=1}^N [z^*(\mathbf{x}_\alpha) - z(\mathbf{x}_\alpha)] \quad (3.16)$$

The MEE is a measure of the bias of the interpolation, so ideally the MEE should be zero. It can be both positive and negative, but to have an unbiased estimation it should be close to zero. This value should be evaluated relatively to the mean of the observations.

- The root mean squared estimation error (RMSEE)

$$\text{RMSEE} = \sqrt{\left(\frac{1}{n} \sum_{\alpha=1}^N [z^*(\mathbf{x}_\alpha) - z(\mathbf{x}_\alpha)]^2\right)} \quad (3.17)$$

The RMSEE should be as low as possible to have an accurate estimation since it is an indication of the overall error of the estimation.

- The Pearson correlation coefficient (r) between $z(\mathbf{x}_\alpha)$ and $z^*(\mathbf{x}_\alpha)$

The Pearson correlation coefficient indicates the extent to which the observation and the predicted value are linearly related. It should always be considered in combination with the MEE. So ideally, r should be close to one with a small MEE.

These indices can be used to quantify the error of the prediction, to select the best variogram model or to compare several interpolation methods.

3.2 Classification and map comparison

3.2.1 Fuzzy- k -means classification

The purpose of classification is to enable a concise description of the spatial variation of soil as a multivariate system (Cline, 1949). Classification is a means of data reduction whereby sets of observations or maps are made understandable and transparent. Since information is always lost, the optimal classification aims to reduce information, while identifying groups of individuals or pixels that have common properties. To facilitate the interpretation of an EC_a map, it is common practice to classify the pixels into naturally occurring clusters based on their EC_a value. Two classification approaches exist: (i) hard classification into mutually exclusive classes with rigidly defined boundaries allowing only binary membership and (ii) fuzzy classification which admits the possibility of a partial membership and where the class boundaries can not be defined exactly. The complexity of soil variability acting at different spatial and temporal scales with many interacting attributes calls for the latter approach to soil classification. Soil boundaries are characterized by a transitional nature, hence crisp classification would not be accurate and lead to unnatural classes. Dealing with a fuzzy or continuous classification is based on the theory of fuzzy sets, a mathematically meaningful method to group individuals into classes that do not have sharply defined boundaries (Zadeh, 1965). The best known fuzzy classification is the fuzzy- k -means (Bezdek, 1981). The fuzzy- k -means classification is an unsupervised classification used to identify ‘natural clusters’ in a dataset \mathbf{X} (with elements $z(\mathbf{x}_{ig}); i = 1, \dots, n; g = 1, \dots, p$) having n individuals (pixels) with p attributes (Bezdek, 1981). Each individual is allocated a membership m_{ij} to each of the k clusters and the following conditions should be satisfied:

$$\sum_{j=1}^k m_{ij} = 1 \quad 1 \leq i \leq n; \quad (3.18)$$

$$\sum_{i=1}^n m_{ij} > 0 \quad 1 \leq j \leq k; \quad (3.19)$$

$$m_{ij} \in [0, 1] \quad 1 \leq i \leq n; 1 \leq j \leq k. \quad (3.20)$$

It is an iterative algorithm starting with a random set of cluster centroids. Each individual is then assigned to the closest of these centroids (in attribute space) and new centroids are calculated based the attribute values of the individuals in the cluster. Based on the distance in attribute space between the individual and the new cluster centroid the membership values are recalculated. This is repeated until a specified convergence criterion is met. The aim is to identify cluster centroids that minimize the generalized objective function $J(\mathbf{M}, \mathbf{C})$. $J(\mathbf{M}, \mathbf{C})$ is the sum of squared errors (expressed as distances in the attribute space) due to the representation of an individual by the center of its class. $J(\mathbf{M}, \mathbf{C})$ is defined as:

$$J(\mathbf{M}, \mathbf{C}) = \sum_{i=1}^n \sum_{j=1}^k m_{ij}^{\phi} d^2(x_i, c_j) \quad (3.21)$$

where \mathbf{M} is the matrix with membership values m_{ij} of an individual (with $i = 1, \dots, n; j = 1, \dots, k$), \mathbf{C} is the matrix of class centers c_{jg} (with $j = 1, \dots, k; g = 1, \dots, p$), x_i is the vector representing the p attribute values at location \mathbf{x}_i ($z(\mathbf{x}_{i1}), \dots, z(\mathbf{x}_{ip})$), c_j is the vector representing the p centroid values for class j (c_{j1}, \dots, c_{jp}), d^2 is the squared distance between x_i and c_j according to a chosen definition of distance and ϕ is the fuzzy exponent. The distance measures to be chosen from are Euclidean, diagonal or Mahalanobis (Odeh et al., 1992).

The Euclidian distance gives equal weight to all attributes regardless their average and standard deviation and it is insensitive to statistically dependent attributes:

$$d_{ij}^2 = (x_i - c_j)^T (x_i - c_j). \quad (3.22)$$

The diagonal distance is also insensitive to statistically dependent attributes but compensates for non-equivalent variances among the attributes by weighing the distances with the variance of the attributes:

$$d_{ij}^2 = (x_i - c_j)^T \mathbf{A}_D (x_i - c_j) \quad (3.23)$$

with \mathbf{A}_D being a $p \times p$ diagonal matrix with $(\frac{1}{\sigma_p})^2$ as diagonal elements.

The Mahalanobis distance does not only compensate for disparities in the variance, but also accounts for statistically dependent measured variables:

$$d_{ij}^2 = (x_i - c_j)^T \Sigma^{-1} (x_i - c_j) \quad (3.24)$$

with Σ being the within-class variance-covariance matrix.

In case of a strong discrepancy in the variance of the attributes, the Euclidean distance is to be avoided since the result will be strongly influenced by those attributes with a larger variance than others. The choice between the latter two distance measures depends on the correlation among the measured attributes.

The degree of fuzziness, expressed by the fuzzy exponent ϕ , determines the extent to which the groups are compact and separated. The fuzziness exponent ϕ varies between 1 and ∞ : when ϕ is 1 the solution is a hard partition, as

ϕ approaches infinity the solution approaches its maximum degree of fuzziness (McBratney and de Gruijter, 1992). For soil data, a fuzzy exponent between 1.12 and 1.50 is advised (Odeh et al., 1992).

To determine the optimum number of clusters two indices are minimized: the fuzziness performance index (FPI), which measures the degree of fuzziness, and the normalized classification entropy (NCE), which indicates the degree of fuzzification. A balance should be found between structure and continuity.

The FPI is defined as follows (Roubens, 1982):

$$FPI = 1 - \frac{kF - 1}{k - 1} \quad 0 \leq FPI \leq 1 \quad (3.25)$$

where F is the partition coefficient:

$$F = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k m_{ij}^2. \quad (3.26)$$

The FPI describes the membership sharing between any pair of fuzzy classes. As FPI approaches one membership sharing increases, while classes become more distinct with less membership sharing when FPI approaches zero. Classes are non-fuzzy or crisp when FPI = 0. Minimizing the FPI indicates an optimal number of fuzzy classes that best reflect the substructures present in the dataset **X**.

The NCE is defined as:

$$NCE = \frac{H}{\log k} \quad 0 \leq NCE \leq 1 \quad (3.27)$$

where H is the entropy function:

$$H = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k m_{ij} \log(m_{ij}). \quad (3.28)$$

The NCE is an estimate of the amount of desorganization created by the of fuzzy-*k*-partition of elements of dataset **X**. As NCE approaches one, disorganization dominates, while with a value of zero excellent organization is achieved. Minimizing NCE is consistent with maximizing the amount of information about the substructures in **X** generated by the fuzzy-*k*-means classification.

Overall, the fuzzy-*k*-means classification aimed at minimizing $J(\mathbf{M}, \mathbf{C})$ using the following algorithm:

1. Choose the number of classes k with $1 < k < n$;
2. Choose the fuzziness exponent ϕ , with $\phi > 1$;
3. Choose a distance measure in the attribute space, $d^2(x_i, c_j)$;
4. Choose a value for the stopping criterion, ϵ (usually $\epsilon = 0.001$);

5. Initialize $\mathbf{M} = \mathbf{M}_0$, with random memberships;
6. At iteration $it = 1, 2, 3, \dots$ (re)calculate $\mathbf{C} = \mathbf{C}_{it}$;
7. Re-calculate $\mathbf{M} = \mathbf{M}_{it}$, constraining m_{ij} to 0 if $d_{ij}^2 \approx 0$;
8. Stop on convergence: if $\|\mathbf{M}_{it} - \mathbf{M}_{it-1}\| \leq \epsilon$ then stop, otherwise go to step 6.

The fuzzy- k -means classification software used in this work was FuzME (Minasny and McBratney, 2002).

3.2.2 Comparing categorical maps

Reasons to compare maps can be of different nature: (i) accuracy assesment to characterize the similarity between a reference map ('truth') and a prediction or classification (Foody, 2002); (ii) similarity assesment between two maps of related properties; (iii) change detection to detect temporal differences in maps and (iv) model comparison where model outputs are compared with observations or other model outputs (Hagen, 2003). Only the first two reasons fitted into this research.

Mostly a fuzzy or soft classification is hardened based on the maximum membership value of each pixel and by doing so categorical maps are created. The uncertainty of this classification can be expressed by calculating for all pixels i the confusion index (CI_i) (Burrough et al., 1997) or the entropy index (H_i) (Maselli et al., 1994). The confusion of placing a pixel into one of the k classes can be determined by:

$$CI_i = [1 - (m_{imax} - m_{i(max-1)})] \quad 1 \leq i \leq n \quad (3.29)$$

where m_{imax} is the membership value of the class with maximum membership for pixel i and $m_{i(max-1)}$ is the second largest membership for the same pixel. The confusion index determines the degree of class overlap in attribute space for each pixel. In case of small CI values, it is clear to which class the pixel belongs, for large CI values there exist confusion about in which class the pixel should be placed. Analogously, the entropy of a classification shows the degree to which a classification output is fuzzy (i.e. uncertain or soft) or crisp. To calculate the entropy for a pixel i , equation 3.28 is slightly modified as defined by Maselli et al. (1994):

$$H_i = - \sum_{j=1}^k [m_{ij} \log_2(m_{ij})] \quad 1 \leq i \leq n. \quad (3.30)$$

The entropy for a pixel is maximized when the pixel has equal memberships for all classes, while a minimum value is obtained when the pixel is entirely allocated to one class.

Categorical comparisons are generally based on a confusion matrix (also called error matrix or contingency table) containing categorical similarities obtained from a pixel-by-pixel comparison. Table 3.1 shows a confusion matrix of k classes or categories. The columns represent the true or reference classes, while the rows represent the predicted or modeled classes. The elements n_{ij} ($1 \leq i \leq k, 1 \leq j \leq k$) are the number of pixels that fall into each categorical combination and n is the total number of pixels. The same symbology is used in Eq. (3.31) to (3.33b). Diagonal elements represent an agreement between the reference and the prediction, and off-diagonal elements represent misclassifications (Congalton and Green, 1999).

TABLE 3.1 – Confusion matrix.

		True				Total
		Class ₁	Class ₂	...	Class _k	
Classified	Class ₁	n_{11}	n_{12}	...	n_{1k}	n_{1+}
	Class ₂	n_{21}	n_{22}	...	n_{2k}	n_{2+}

	Class _k	n_{k1}	n_{k2}	...	n_{kk}	n_{k+}
	Total	n_{+1}	n_{+2}	...	n_{+k}	n

A first measure is the overall accuracy or proportion observed agreement (P_o), defined as the number of correctly classified pixels divided by the total number of pixels. So it is a relatively simple measure of agreement only using the main diagonal elements (n_{ii}) of the confusion matrix. On the other hand, it does not take into account the possibility that a large number of cells can be classified correctly due to chance, causing an overestimation of the classification accuracy. The Kappa statistic (κ) (Cohen, 1960) eliminates classification agreement by chance, since it also takes into account the non-diagonal elements (see Eq. (3.31)). It indicates proportionally how much better the results (proportion observed agreement - P_o) are compared with a purely random classification (proportion chance agreement - P_c). To calculate kappa the following formulae should be used:

$$\kappa = \frac{P_o - P_c}{1 - P_c} \quad (3.31)$$

$$\text{with } P_o = \sum_{i=1}^k P_{ii} = \frac{1}{n} \sum_{i=1}^k n_{ii}$$

$$P_c = \sum_{i=1}^k P_{+i} P_{i+} = \frac{1}{n^2} \sum_{i=1}^k (n_{+i} n_{i+})$$

P_{+i} , P_{i+} , P_{ii} were defined as the column proportions, the row proportions and the diagonal proportions respectively. Kappa varies between -1 and 1 and belongs to

a family of indices that have the following properties (Landis and Koch, 1977):

- if the classification is perfect, $\kappa = 1$;
- if the observed proportion correct is greater than the expected proportion correct due to chance, then $\kappa > 0$;
- if the observed proportion correct equals the expected proportion correct due to chance, then $\kappa = 0$;
- if the observed proportion correct is less than the expected proportion correct due to chance, then $\kappa < 0$.

Landis and Koch (1977) even proposed a scale to describe the degree of agreement according to the κ value (Table 3.2).

TABLE 3.2 – Scale of Landis and Koch (1977) to interpret κ .

κ	Interpretation
≤ 0	No agreement
0.01 - 0.20	Slight agreement
0.21 - 0.40	Fair agreement
0.41 - 0.60	Moderate agreement
0.61 - 0.80	Substantial agreement
0.81 - 1.00	Almost perfect agreement

Variants of the standard κ were developed to correct κ with random chance agreement (κ^*) (Foody, 1992) and to quantify how much of the error is due to categorical differences (κ_{histo}) and locational errors (κ_{loc}) (Pontius, 2000).

Foody (1992) showed that the standard κ overestimates the agreement by chance and therefore underestimates the accuracy. He modified κ to κ^* by giving each category an equal membership probability being $\frac{1}{k}$ (Brennan and Prediger, 1981). As the number of classes k increases κ^* approaches P_o . The size of the classes on the other hand does not influence the classification accuracy. The κ^* coefficient is calculated as follows:

$$\kappa^* = \frac{P_o - \frac{1}{k}}{1 - \frac{1}{k}}. \quad (3.32)$$

The second criticism to the standard κ is that it confounds quantification error with location error. Therefore, Pontius (2000) defined κ as the product of κ_{histo} and κ_{loc} . κ_{histo} is a measure for the quantitative similarity of two maps and depends only on the total number of cells in each category. κ_{loc} , on the other hand, indicates the extent to which the similarity is a result of the spatial distribution of cells and depends on the spatial distribution of the categories on the map. Areal extents in

the classification and the reference image may match, although the locations may be quite different. The differentiation between the two sources of error is quantified as follows:

$$\kappa_{\text{histo}} = \frac{P_{\text{max}} - P_c}{1 - P_c} \quad \text{with} \quad P_{\text{max}} = \frac{1}{n} \sum_{i=1}^k \min(n_{i+}, n_{+i}); \quad (3.33a)$$

$$\kappa_{\text{loc}} = \frac{P_o - P_c}{P_{\text{max}} - P_c}. \quad (3.33b)$$

P_{max} defines the maximum possible agreement, given that the total number of cells per category does not change.

4

Characterizing soil nitrate dynamics using the EM38DD soil sensor

The content of this chapter was published as:

Cockx, L., Van Meirvenne, M., and G. Hofman. (2005). Characterization of nitrogen dynamics in a pasture soil by electromagnetic induction. *Biology and Fertility of Soils*, 42: 24-30.

4.1 Introduction

Quantification of the nature and extent of variability in soil properties is particularly important from the viewpoint of precision agriculture. Precision agriculture has been defined in several ways. To my opinion the definition as stated by Pierce and Nowak (1999) is the most complete one: 'precision agriculture is the application of technologies and principles to manage spatial and temporal variability associated with all aspects of agricultural production for the purpose of improving crop performance and environmental quality'. Precision agriculture intends to match agricultural inputs and practices to localized conditions within a field, also called site-specific management, a practice summarized as 'doing the right thing, in the right place, at the right time'. The fine-scale management of precision agriculture is in contrast to whole-field management strategies, where management decisions and practices are uniformly applied throughout a field. It is clear that a crucial aspect of precision agriculture is variability.

Nitrate (NO_3^-) leaching is a worldwide concern, evoking an appropriate soil nitrogen (N) management in future agricultural practices (Meinardi et al., 1994). To reduce the risk for leaching during the winter period, efforts should be made to minimize the soil mineral N in late autumn. In this context, site specific N management becomes important since it adjusts management practises to the different nitrate dynamics within one field, with the aim to decrease the amount of NO_3^- -N leached. This implies that methods are needed to assess the soil NO_3^- -N variability both in space and time.

It is known that the NO_3^- -N content of pastures is characterized by a high spatial within field variability (Bogaert et al., 2000). This NO_3^- -N content may originate from the N applied in the form of fertilizers, or from the mineralization of soil organic N (Addiscott, 1996). Within a grazed pasture a large proportion (80%) of the N ingested by the grazing animal returns to the soil as excreta. One urine patch represents a local N application ranging from 400 to 1200 kg ha⁻¹ (Addiscott et al., 1991). Much of the NO_3^- -N leached from pastures during the winter originates from these high concentrations of N deposited in urine patches during the grazing season, since this N cannot be fully utilized by the grass crop at the end of the growing season (Ryden et al. 1984). This source of NO_3^- -N variability alters each time the field is grazed. Besides, the amount of NO_3^- -N lost also depends on soil and climatic conditions and management practises. Variability of soil properties within a field affects the magnitude of soil processes such as nutrient accumulation and solute transport.

The classical approach to characterize within field soil variability is based on an intensive sampling scheme, which is very laborious and costly. A modern approach to map soil properties uses a mobile, non invasive soil sensor. EMI soil sensors have been applied widely to describe field variability for application to

precision agriculture (Rhoades et al., 1999). It has been found that EMI based measurements of the apparent electrical conductivity are related to a number of soil properties, like soil texture (McNeill, 1980b), soil water content (Sheets and Hendrickx, 1995) and CEC and exchangeable Ca and Mg (McBride et al., 1990). Relations between soil NO_3^- -N and electrical conductivity (EC) have been established as well. In the laboratory, measurements of EC as 1:1 soil to water mixtures have been used to estimate the soil NO_3^- -N content (Smith and Doran, 1996; Zhang and Wienhold, 2002). The latter also demonstrated that, in non saline soils with no free carbonates present, portable EC meters are able to detect changes in the soil's inorganic N status in situ. Using an EMI sensor, specific mobile ions (Cl^- , NH_4^+ and NO_3^-) that are associated with animal waste were detected (Eigenberg et al., 1998). Besides, high correlations were obtained with profile weighted EC_a measurements of an EM38 sensor in horizontal orientation (EC_a to about 75 cm) and soil NO_3^- -N in the surface 0–23 cm and 23–46 cm soil layers throughout the growing season in a corn field (Eigenberg et al., 2002).

This study investigated how electromagnetic induction can be used to improve the characterization of NO_3^- -N dynamics within a pasture field.

4.2 Material and Methods

4.2.1 Study site

The study area was a flat cattle pasture of 1 ha in Lovendegem, Belgium (with central coordinates: 51°37'09" N, 3°07'22"E). According to the Belgian soil map (with a scale of 1:20 000) the majority of the study site belongs to the 'Sdh' soil series which represents silty sand topsoil texture (textural class 'S'), moderately wet conditions (drainage class 'd') and a crushed humus and/or iron podzol B horizon (profile development 'h') (Fig. 4.1 (a)). A small part of the study site is mapped as 'Zch', which differs by its sandy topsoil texture and its moderately dry drainage class. The study sites belongs to the sandy region of Flanders ('Vlaamse Zandstreek') (Fig. 4.1(b)).

The pasture field has been treated with two types of fertilizers: liquid cow slurry applied early March 2002 at a rate equivalent to 120 kg N ha⁻¹, and a mineral fertilization applied as NH_4NO_3 (27% N) in April, May and June 2002, resulting in a total of 216 kg N ha⁻¹. During the period november 2002 till the middle of March 2003 there was no cattle on the pasture.

4.2.2 Data collection

The field was sampled according to a clustered sampling design at 116 locations: 58 locations were obtained on a 12.5 by 12.5 m grid basis and another 58 locations were chosen randomly (Fig. 4.2(a)). Soil samples were taken at three depths

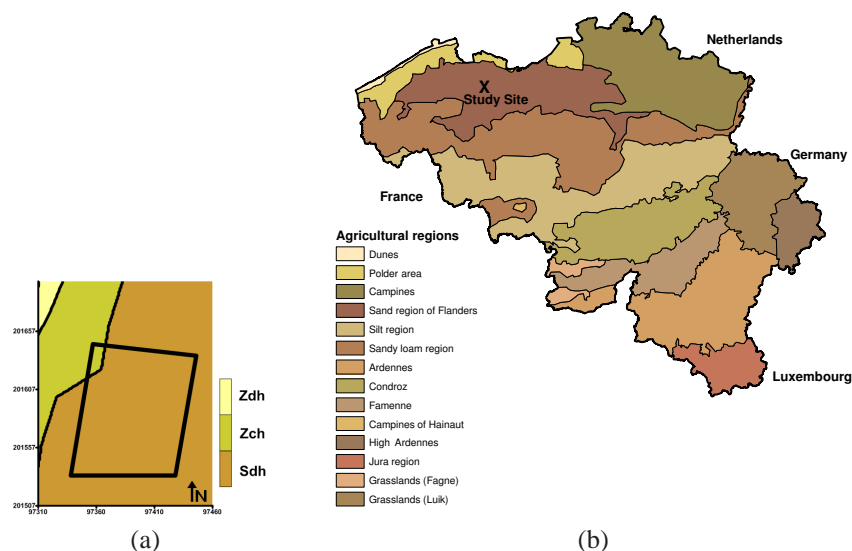


FIGURE 4.1 – (a) Belgian soil map with indication of the study site and (b) Location of the study site in the sandy region of Flanders (Belgium).

(0–10 cm, 10–30 cm and 30–60 cm) on 4 November 2002, 27 January 2003 and 12 March 2003. All these samples were analyzed for the NO_3^- -N content. In the top 30 cm the soil was analysed by a continuous flow auto analyzer (Kalra and Maynard, 1991). The subsoil was analysed with a NO_3^- specific electrode (Orion Research, 1991). Here we have considered only the total NO_3^- -N content of 0–60 cm topsoil. In 10 locations, soil samples were taken up to 90 cm depth and analyzed for the textural fractions using the pipette method (Gee and Bauder, 1986).

In November 2003, the EC_a of the pasture was measured by the EM38DD sensor. The EM38DD measures simultaneously the EC_a -H and EC_a -V. Each EM38 orientation has a different depth response profile, so with one field survey EC_a data of top- and subsoil were collected. The EM38DD sensor was attached to an ATV and connected with a field computer and GPS. The EC_a data and their locations were merged in the field computer using HGIS software (Starpal Inc, Colorado, USA). The ATV was driven along parallel lines with an interval of 5 m and the EC_a -H and EC_a -V measurements were taken simultaneously every second (Fig. 4.2(b)). At an average speed of 10 km h^{-1} , within the lines the measurement interval was around 3 m.

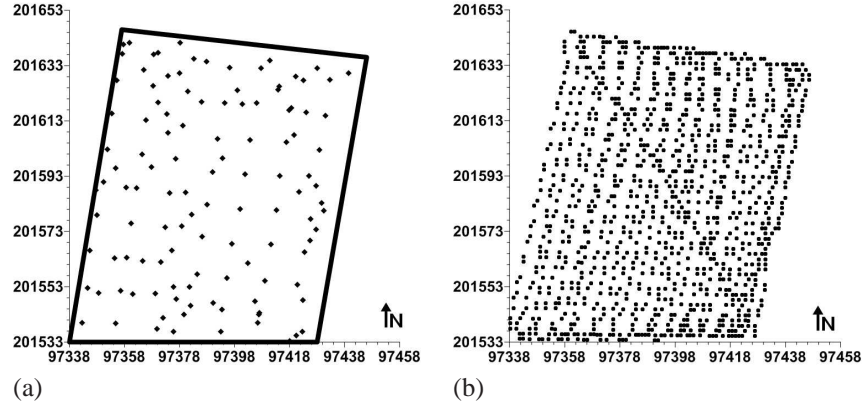


FIGURE 4.2 – (a) Indication of the sampling sites and (b) the EC_a measurements. The coordinates are expressed in meters, conform the Lambert72 projections used in Belgium.

4.2.3 Mapping EC_a

Continuous maps of both EC_{a-H} and EC_{a-V} were obtained using ordinary kriging (OK), based on the spatial autocorrelation among data whereby a variogram is used to model the spatial variability of these data. Estimates of the variable at unsampled locations $Z^*(\mathbf{x}_0)$ were calculated using the basic linear regression equation (Goovaerts, 1997):

$$Z^*(\mathbf{x}_0) - m(\mathbf{x}_0) = \sum_{\alpha=1}^{n(\mathbf{x}_0)} \lambda_{\alpha} [Z(\mathbf{x}_{\alpha}) - m(\mathbf{x}_{\alpha})] \quad (4.1)$$

where λ_{α} is the unknown weight assigned to the observation $z(\mathbf{x}_{\alpha})$, $m(\mathbf{x}_0)$ and $m(\mathbf{x}_{\alpha})$ are the expected values of $Z(\mathbf{x}_0)$ and $Z(\mathbf{x}_{\alpha})$, respectively and the number of data used in the interpolation is $n(\mathbf{x}_0)$, being the data located within a given neighbourhood around (\mathbf{x}_0) . The weight λ_{α} was obtained by solving the kriging algorithm, in which the variogram model was used to convert the spatial distances into variogram values. Ordinary kriging requires the mean $m(\mathbf{x}_0)$ to be stationary within the neighbourhood around \mathbf{x}_0 . The interpolated EC_{a-H} and EC_{a-V} values were used as input for a fuzzy- k -means classification to create management zones. A fuzzy- k -means classification allows allocating each individual as a member of each class (Triantafilis et al. 2003). The method minimizes the multivariate within class variance and consequently individuals classified to the same class have similar attributes (McBratney and de Gruijter 1992). For more details see section 3.2.1. Therefore EC_a -zones reflected differences in soil condition like soil texture and moisture, the most important factors affecting the EC_a of non saline soils.

4.2.4 Mapping NO_3^- -N

In addition, the NO_3^- -N measurements were interpolated using stratified simple kriging (SK), using the EC_a -zones as strata. Simple kriging requires the global mean to be known and stationary over the entire study area. Therefore, for SK, Eq. (4.1) becomes:

$$Z^*(\mathbf{x}_0) - m = \sum_{\alpha=1}^{n(\mathbf{x}_0)} \lambda_{\alpha} [Z(\mathbf{x}_{\alpha}) - m] \quad (4.2)$$

where m is the stationary mean. In the situation where two or more zones (strata) exist within a study area, each with a different mean, the observations can be transformed to residuals with zero mean by subtracting their local means. The residuals can then be interpolated using a variogram of the residuals, after which the local means are added. This procedure was followed to include the EC_a -zones into the mapping of NO_3^- -N at the three sampling times. Comparing these maps, the nitrate dynamics of the pasture could be interpreted by considering two periods: from 3 November till 27 January and from 28 January till 12 March. To get an indication of the NO_3^- -N dynamics during such a period, a ratio was calculated by dividing the NO_3^- -N content of the first sampling time by the NO_3^- -N content of the consecutive sampling time. If this ratio was larger than one, NO_3^- -N was lost during that period, otherwise it accumulated. Finally we interpolated the NO_3^- -N content at the 3 sampling times independently from the EC_a measurements, by OK. These three NO_3^- -N maps were used in a fuzzy- k -means classification to create nitrate zones.

4.2.5 Validation

To validate the accuracy of the interpolation methods a cross validation (leave-one-out method) was performed. The Mean Square Estimation Error (MSEE) was used to compare the two interpolation methods:

$$\text{MSEE} = \left(\frac{1}{n} \sum_{\alpha=1}^N [z^*(\mathbf{x}_{\alpha}) - z(\mathbf{x}_{\alpha})]^2 \right) \quad (4.3)$$

where the number of observed values $z^*(\mathbf{x}_{\alpha})$ ($\alpha = 1, \dots, n$) is n and $z^*(\mathbf{x}_0)$ are the estimated values at the same locations. The relative improvement (RI) of precision was calculated:

$$\text{RI} = \frac{(\text{MSEE}_{\text{OK}} - \text{MSEE}_{\text{SK}})}{\text{MSEE}_{\text{OK}}} \times 100 \quad (4.4)$$

where MSEE_{OK} and MSEE_{SK} are the MSEE for the OK using NO_3^- -N data only and the MSEE for SK of NO_3^- -N data involving the EC_a -zones, respectively. The RI is expressed in %. If RI is positive, the precision of the SK method is superior to OK. Vice versa for negative RI.

4.3 Results and Discussion

An exploratory data analysis of the EC_a data was done (Table 4.1), showing that both orientations had a positively skewed distribution. Compared to EC_a -V, the data range of the measured EC_a -H data was larger and their mean was smaller. A relative measure of the variability is the coefficient of variation (CV), which is defined as the ratio of the standard deviation to the mean. Due to the smaller mean, the CV of the EC_a -H data was larger. The two orientations were well correlated with a linear correlation coefficient of 0.77.

TABLE 4.1 – Descriptive EC_a statistics ($n = 833$).

	Min	Max	Mean	Median	CV [%]	Skewness
EC_a -V [$mS\ m^{-1}$]	8.1	20.4	11.9	11.4	19.0	0.78
EC_a -H [$mS\ m^{-1}$]	6.0	25.1	9.9	9.0	30.0	1.7

Both EC_a datasets were best modeled with an exponential variogram: the EC_a -V data showed a range of 37.8 m with an RNE of 4.2%, the EC_a -H data had a somewhat smaller range (21 m), but a similar very low relative nugget effect (RNE) of 3.8%. The RNE is the ratio of the nugget variance to the total variance (the sill), mostly expressed as a percentage. It is an index of the unstructured variance or the variance not attributable to spatial dependence. Here, almost all variance was spatially structured. The OK interpolated EC_a -V and EC_a -H maps with a resolution of 1 by 1 m are given in Fig. 4.3.

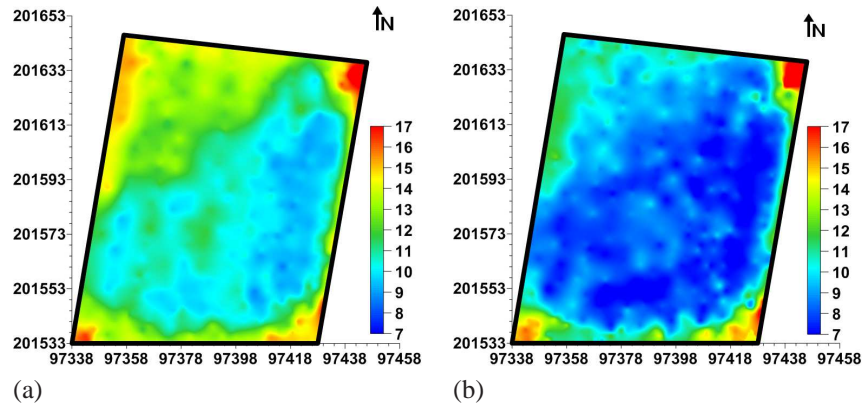


FIGURE 4.3 – (a) Interpolated map of EC_a -V and (b) EC_a -H.

Three corners of the field showed increased EC_a values and these areas corresponded to muddy and stony parts representing the drinking place and the entrances/exits to the field. Since these areas represent human-induced artifacts they

were not considered further. The central and lower part of the field showed low EC_a values, while the upper left part was mainly characterized by intermediate EC_a values. EC_a -V and EC_a -H displayed similar patterns except for the upper left corner where EC_a -V was larger than EC_a -H, indicating a more heterogeneous soil profile in that part of the field. A fuzzy- k -means classification confirmed our observations; the field was optimally divided into 2 zones: zone 1 with low EC_a values and zone 2 with higher EC_a values (Fig. 4.4).

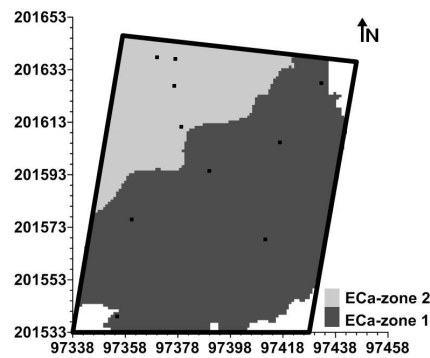


FIGURE 4.4 – Fuzzy- k -means classification based on EC_a with indication of the texture samples.

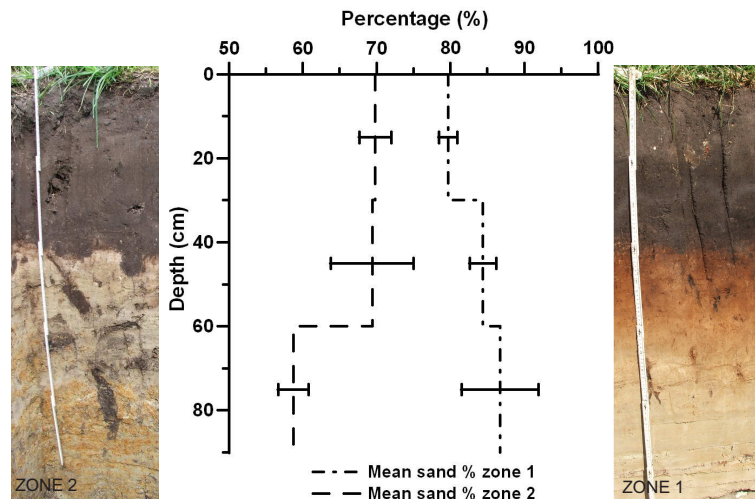


FIGURE 4.5 – Mean sand fraction (in %) in both EC_a zones, with to the left of the graph a soil profile located in EC_a zone 1 and to the right a soil profile located in EC_a zone 2.

A texture analysis explained the differences in EC_a : Fig. 4.4 shows the locations of the samples and a graph of the mean sand fraction per depth interval for

each zone is given in Fig. 4.5. The mean profile of zone 1 had a homogeneous sandy texture, with a sand content ranging from 79% to 87% down to 90 cm. The mean profile within zone 2 also showed a sandy topsoil (70% sand fraction) but with a clear reduction in sand content with depth (to about 60%). The observed textural differences of the field were not captured by the Belgian soil map (Fig. 4.1 (a)) due to the limitations of the soil map legend. The presence of a substrate can be indicated by placing the small letter before the textural class. The 'w' substrate symbol however refers to clay-sand substrates of Tertiary origin, while the clayey-sand enrichment in this field has a fluvial Quaternary origin which is not captured by a legend symbol. Figure 4.5 also shows a picture of the soil profile in each zone, clearly indicating an increase in clay content at a depth of 80 cm in zone 2. With the increase in clay and silt contents in the subsoil; zone 2 was noticeable wetter during field work.

Through a series of EM38DD surveys of this field, it was observed that the absolute EC_a values fluctuate over time. Figure 4.6 shows the EC_a -V measurements obtained in November 2003, 2004 and 2005. Besides a change in absolute values, also the measurement resolution differs among these surveys, but the spatial patterns of EC_a remained constant. At each survey, the northwest corner has an increased electrical conductivity. Like Sudduth et al. (2001), it can be concluded that an EC_a based classification results in stable management zone delineations.

The NO_3^- -N data at the three sampling times were stratified according to the two EC_a -zones; the relative statistical analysis is reported in Table 4.2. On average zone 1 contained higher NO_3^- -N concentrations with smaller coefficients of variation than zone 2. One would expect a lower NO_3^- -N content in EC_a -zone 1 causing the lower conductivities. However differences in soil texture often influence the EC_a much more than small differences in nutrient concentrations (Heiniger et al., 2003). If EC_a changes due to one soil property are much larger than those to other factors than EC_a is mainly determined by that dominant factor (Sudduth et al., 2003b). This means that the texture difference between the zones masked the effect of a different NO_3^- -N content in these zones. The difference in NO_3^- -N content between the zones was largest in November. As the winter period continued, a clear reduction in nitrate content and variability was observed. In November the NO_3^- -N content of both zones was highly variable (CV in the order of 100%), in January NO_3^- -N content of zone 2 (CV = 115%) was much more variable than of zone 1 (CV = 62%) and in March both zones had similar CVs (in the order of 45%).

After subtracting the stratified mean values from the observations, the variograms of the pooled residuals were calculated and modeled (Fig. 4.7). For the three sampling times the semivariance $\gamma(h)$ at lag distances h was defined by a nugget effect C_0 and a spherical model (Eq. 3.4). At the end of the winter pe-

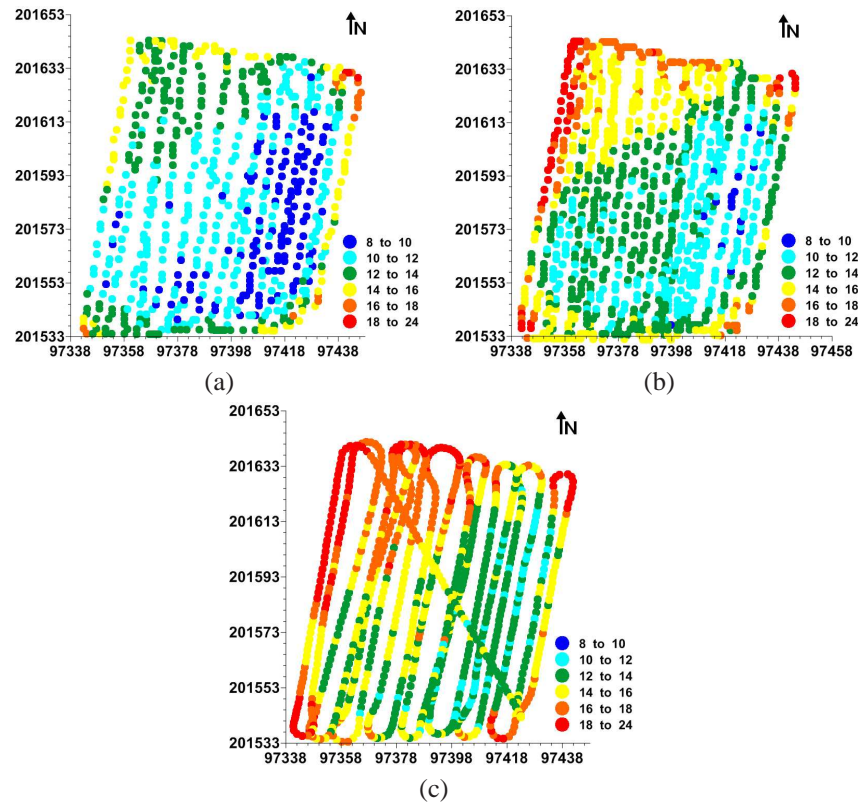


FIGURE 4.6 – EC_a-V measurements obtained in (a) November 2003, (b) November 2004 and (c) November 2005.

TABLE 4.2 – Descriptive nitrate statistics, stratified according to the EC_a-zones.

		4 November 2002		27 January 2003		12 March 2003	
		Zone 1	Zone 2	Zone 1	Zone 2	Zone 1	Zone 2
Mean	[kg ha ⁻¹]	65	24	29	16	29	17
Min	[kg ha ⁻¹]	11	4	5	4	12	2
Max	[kg ha ⁻¹]	290	128	97	109	61	39
CV	[%]	90	107	62	115	40	47

riod, the variograms of the NO₃⁻-N residuals displayed a trend to shorter ranges and a larger relative nugget effect. The range decreased from 23.7 m November to 20.4 m in January and only 11 m in March. The RNE increased from 17.1% in November, to 23.5% in January and 34.4% in March. With time not only the NO₃⁻-N content decreased, also the nitrate variability was less structured and occurred over shorter distances. This implies that gradually the deterministic mean

became a more reliable predictor.

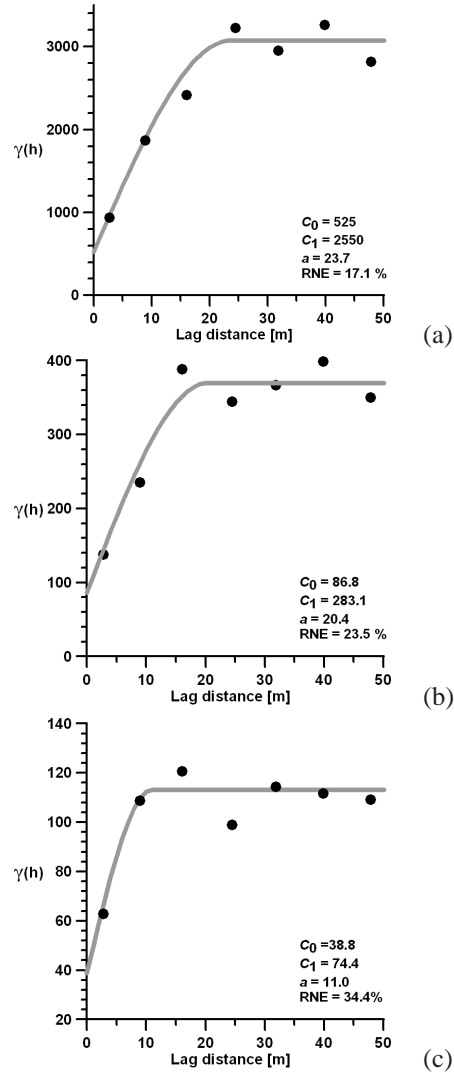


FIGURE 4.7 – Variograms of the NO_3^- -N residuals at the three sampling times (a) 4 November 2002, (b) 27 January 2003 and (c) 12 March 2003. On the X-axis the lag distance is expressed in m, on the Y axis the semivariogram values are expressed in $(\text{kg NO}_3^- \text{-N ha}^{-1})^2$.

Then SK of the NO_3^- -N residuals was performed, resulting in NO_3^- -N maps for the three sampling times (Fig. 4.8). A clear separation line between both zones appeared due to the stratification. In November the highest values were found

from the bottom left (field entrance) to the top right (drinking place) of the field, corresponding to a preferential walking path of the cows (Fig. 4.8(a)).

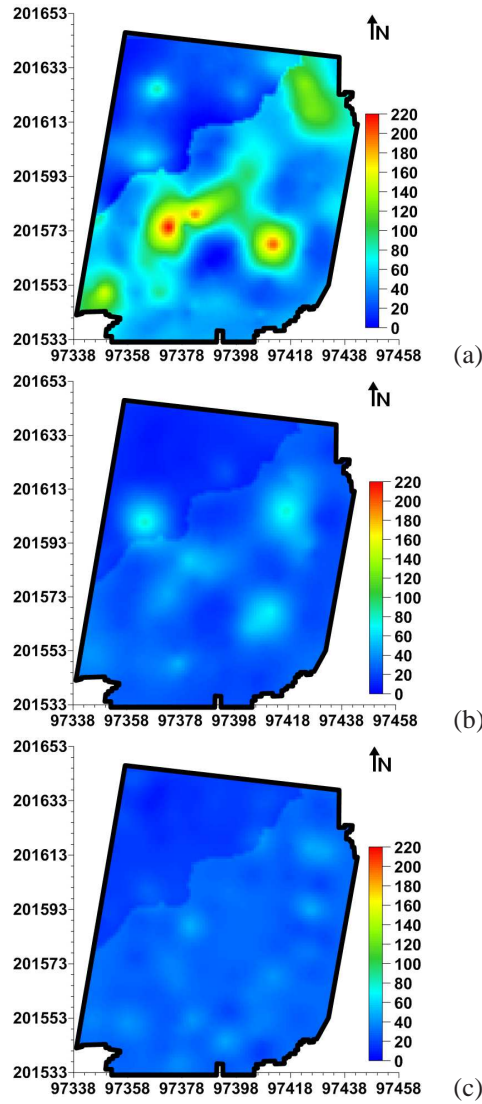


FIGURE 4.8 – Maps of the NO_3^- -N content [kg ha^{-1}] in the top 60 cm at the three sampling times (a) 4 November 2002, (b) 27 January 2003 and (c) 12 March 2003.

This can be related to the fact that cattle frequently visits certain parts of the field as in this case, water sources (Barrow, 1967). In November the NO_3^- -N content of zone 2 was much lower than in zone 1. This indicated that the textural

differences between the EC_a-zones also influenced the NO₃⁻-N content. The homogeneous sandy soils of zone 1 produced less grass biomass than the more clayey zone 2. Consequently less fertilizer was used by the crop, resulting in higher NO₃⁻-N residues. However it could also be that the cows visited this wetter area less frequently or that denitrification was favoured under the more clayey and wet conditions of zone 2 (Garrett et al., 1992). Fine textured soils have slower drainage and a greater potential for denitrification. Nevertheless it was assumed that denitrification was limited by the low winter temperatures. In January the NO₃⁻-N content and variability decreased, but the difference between the zones remained. In March the map was more homogeneous because the NO₃⁻-N content decreased and the variability smoothened. It is clear that the decrease in NO₃⁻-N content between November and January was larger than the decrease between January and March. The excessive rainfall between the first two sampling periods resulted in increased nitrate losses by leaching.

Fig. 4.9 shows the histograms of the ratio of the NO₃⁻-N content in November and the NO₃⁻-N content of January, split for the two zones. In zone 1 this ratio was generally higher than 1, indicating large nitrate losses in most parts of this zone. In zone 2 the graph is bimodal; although a considerable part of this zone also showed values higher than 1, the highest frequencies were observed at values smaller than 1. It may be concluded that zone 1 has a higher risk for nitrate losses than zone 2. This difference is clearly related to the processes of mineralization, (de)nitrification, leaching, immobilization that define the soil inorganic N content. They are controlled by soil characteristics as water content, texture, biological activity, cropping and organic matter (Stevenson, 1982). Gaines and Gaines (1994) stated that soils with a higher silt and clay content retain more NO₃⁻-N than pure sandy soils. Obviously, the homogeneous sandy zone within this field will require a different N management than the more silty-clayey parts.

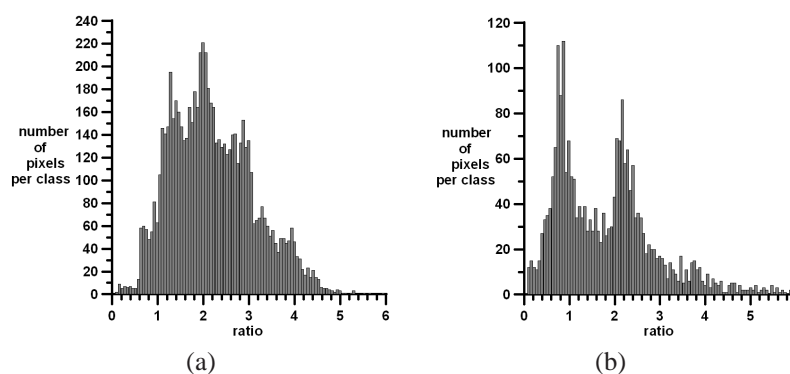


FIGURE 4.9 – Histogram of the ratio of the NO₃⁻-N content on November 4th 2003 to the NO₃⁻-N content on January 27th 2003 for (a) EC_a-zone 1 and (b) EC_a-zone 2.

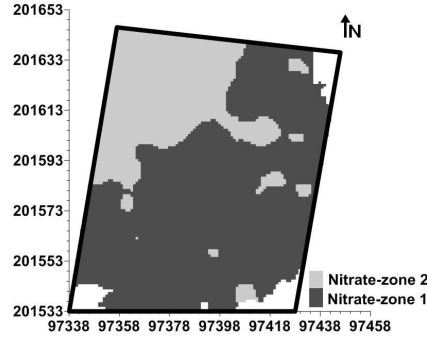


FIGURE 4.10 – Fuzzy- k -means classification based on OK interpolated NO_3^- -N measurements at three sampling times.

To evaluate the added value of including the EC_a -zones in the interpolation of NO_3^- -N the RI (Eq. 4.4) was calculated based on their MSEE. In November the RI was - 1% indicating that there was no difference between the two interpolation methods. For January and March, the RI was 12% and 5% respectively. For those two sampling times, taking into account the EC_a -zones resulted in small to medium improvements in the interpolation of NO_3^- -N. When the NO_3^- -N maps of the three sampling times obtained by OK were subjected to a fuzzy- k -means classification with 2 classes taking the Mahalanobis distance as distance measure, the resulting pattern (Fig. 4.10) was very similar to the classified EC_a map (Fig. 4.4). These NO_3^- -N-zones were compared with the EC_a -zones by cross tabulation. Individual pixels of the two classifications were compared with each other in a confusion matrix (Table 4.3): 90% of the pixels of EC_a -zone 1 pixels were classified as nitrate-zone 1. And 87% of the pixels of EC_a -zone 2 were classified as nitrate-zone 2. The kappa stastic (see section 3.2.2) is 0.73, which means there is substantial agreement between the two classified maps. Since the NO_3^- -N-zones were formed based on the three sampling times they show places with similar nitrate dynamics. Consequently, the EC_a -zones represent zones of very similar nitrate behaviour as such electrical conductivity can be used as an indirect measure of the risk of nitrate losses.

TABLE 4.3 – Confusion matrix obtained after comparison of the NO_3^- -N-zones and the EC_a -zones.

	EC_a zone 1	EC_a zone 2	Total
Nitrate zone 1	6177	321	6498
Nitrate zone 2	715	2130	2835
Total	6892	2451	9343

4.4 Conclusions

This study illustrated that electromagnetic induction measurements are able to delineate zones with different risks of nitrate losses. As an alternative for intensive soil sampling to characterize the nitrate dynamics of a grazed pasture field, a survey with the EM38DD soil sensor characterized the within field soil EC_a variability. In the field two EC_a zones with different textural characteristics were delineated: one zone has a homogeneous sandy profile while the other zone has an increased clay and silt content at the depth 60–90 cm. Based on stratified simple kriging taking the EC_a zones as strata, maps of the residual nitrate content in November, January and March were made. The accuracy of the nitrate maps, taking into account the EC_a zones was slightly better compared to ordinary kriging, based on the NO_3^- -N data only. Also a clear differentiation in nitrate dynamics, in terms of losses throughout the winter period, was observed between the two EC_a zones, an effect attributed to the different textural properties of these zones.

Furthermore, zones with a similar nitrate behaviour were obtained by classifying the nitrate maps of the three sampling times. The resulting nitrate zones were very similar to the EC_a zones, with a kappa coefficient of 0.73.

These results showed that an EC_a survey can be used to characterize nitrate dynamics. Consequently in such conditions, a site specific N-management based on EC_a -zones can be recommended, even for the small fields of Flanders.

5

Explaining the within-field variability of earthworm populations using the EM38DD

Parts of this chapter were published as:

Valckx, J., Cockx, L. Wauters, J., Van Meirvenne, M., Govers, G., Hermy, H. and B. Muys. (2009). Within-field spatial distribution of earthworm populations related to species interactions and soil apparent electrical conductivity. *Applied Soil Ecology*, 41: 315-328.

5.1 Introduction

Ecosystem engineers such as earthworms play a major role in the regulation of ecosystem processes and, ultimately, in the services delivered by the ecosystem (Lavelle et al., 2006). The spatial variability of earthworms will likewise influence spatial patterns of important ecosystem processes such as litter decomposition and nutrient cycling (Ettema and Wardle, 2002). The patchy spatial distribution of earthworms at the field scale in a variety of ecosystems is now well accepted (e.g. Poier and Richter (1992); Rossi et al. (1997); Cannavacciuolo et al. (1998); Nuutinen et al. (1998); Hernández et al. (2007)), but the factors that determine this variability are not sufficiently understood yet. Understanding these driving factors may provide vital information about colonization rates, target earthworm densities for optimized soil ecosystem management and for the design of adequate sampling schemes for earthworm population monitoring.

The spatial variability of earthworms at the field scale can be attributed to two types of factors: biotic and abiotic (Whalen, 2004; Aubert et al., 2005). Biotic interactions of facilitation, competition or coexistence occur both at intra- and interspecific level, and may be responsible for the spatial structuring within and between earthworm populations. From the biotic point of view, it can be expected that individuals/species exhibiting similar ecological strategies may try to minimize resource competition by occurring in spatially differing locations (Jiménez et al., 2006). Abiotic conditions at the field scale are the result of the interaction between local topography, soil texture and related soil hydrology. Also management practices, especially tillage, alter the spatial variability of these abiotic factors (Govers et al., 1994). Given pre-existing spatial variability in soil properties at the field scale, earthworms are expected to occur in higher densities where the soil is intrinsically more favourable. This may be either because individuals move to the best soil patches and/or reproduction is faster in favourable patches (Barot et al., 2007). Relationships between earthworm spatial distribution and soil properties have been studied before (e.g. organic carbon content and soil hydrology in Poier and Richter (1992); Cannavacciuolo et al. (1998)), but rarely these factors were able to satisfactorily explain the observed spatial variability. Earthworm activities in turn can affect soil structure and soil chemical properties, hence the importance of investigating the effect of stable soil variables such as texture on the spatial variability of earthworm populations.

Characterizing the within-field soil textural variability is time, labour and cost intensive, since it relies on extensive soil sampling and subsequent soil analyses. Recently, the use of mobile, sensor-based measurements of soil EC_a has proved to be a quick, easy and reliable method for establishing within-field spatial variability of soil texture (Cockx et al., 2005).

In this study, a geostatistical analysis of the density of the most abundant anecic and endogeic earthworm species and of the soil EC_a was performed for a temperate, arable loess soil. The aims were (i) to describe the observed within-field spatial variability of earthworms and soil EC_a and (ii) to discuss the (dis)similarities between these spatial patterns in terms of community (within and between ecological categories), population (juvenile vs. adult) and abiotic (EC_a) interactions.

5.2 Material and Methods

5.2.1 Study Site

The study was conducted in a 4-ha tilled arable field in Sart-Messire-Guillaume, located in the loess belt of central Belgium (50°36'51"N, 4°34'20"E) (Fig. 5.1). The soil is a Luvisol formed in aeolian loess that was deposited during the last glaciation, and typically its soil series on the Belgian soil map is 'Aba(b)'. The soil has a silt topsoil (textural class 'A') with well drained conditions (drainage class 'b') and is characterized by a mottled argillic B-horizon (profile development 'a(b)'). Topography is rolling with elevations between 135 and 145 m above sea level. The studied field was converted from a pasture to arable land in 1966 and a rotation scheme with sugar beet and winter wheat was grown ever since. Every year, the field is conventionally ploughed in spring to a depth of 20 cm. Before sugar beet planting, 60 t of farm yard manure per hectare was applied. In 2002, as an exception to the regular rotation scheme, chicory (*Cichorium intybus* cv. "sativum") was grown instead of sugar beet. In November 2003, when the field sampling for this study was carried out, the wheat crop had been harvested while its stubble was still left undisturbed. Field sampling for this study was conducted in a representative subarea of 105 by 75 m representing the EC_a range of the field. The sampled area was buffered against edge effects to the north, east and west by a strip of the same arable field of at least 100 m in width. The southern side was buffered by a narrower strip of at least 20 m. Beyond this buffer a pasture was located.

5.2.2 Data collection

5.2.2.1 Earthworm sampling

Earthworms were sampled at 100 locations within the study area between late October and mid November 2003. Temperature and moisture conditions were favourable for earthworm sampling throughout the period. 48 sampling locations were laid out on a regular grid (15 x 15 m) while the remaining 52 sampling plots were randomly distributed over the area (Fig. 5.2). Earthworms were sampled by mustard extraction in 0.5 m² plots. Collected earthworms were preserved in 5%

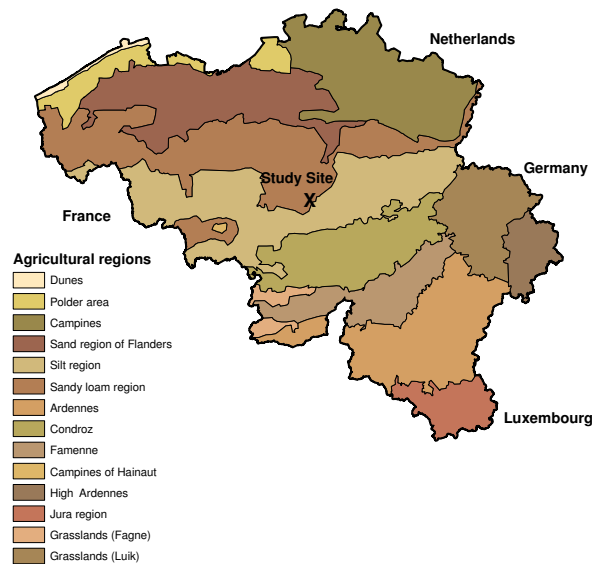


FIGURE 5.1 – Indication of the study site in the Belgian sandy loam region

formalin prior to identification. All individuals were identified to species level following the nomenclature of Sims and Gerard (1999) and their development stage ((sub)adult vs. juvenile) was recorded based on the presence of a clitellum. The ecological category as defined by Bouché (1977) was also recorded. Epigeic species (litter dwelling) are small-sized and live and feed in the litter layer. Endogeic species (shallow dwelling) are medium-sized and while eating soil they build lateral burrow systems through the mineral topsoil layers (mostly within the top 10 cm). Anecic (deep burrowing) species are large-sized and build permanent vertical burrows often extending beyond 2 m. They feed on surface residues and pull these into the burrows. Densities per development stage were expressed as individuals m^{-2} .

5.2.2.2 EC_a measurements

In autumn 2003, when the soil moisture content was near field capacity, the soil was characterized by a mobile EC_a survey. The EM38DD sensor was pulled by an all terrain vehicle in parallel lines 5 m apart at a steady speed of 10 $km\ h^{-1}$. Georeferenced EC_a measurements were recorded every second, resulting in 720 measurements ha^{-1} . The EM38DD sensor consists of two perpendicular EM38 devices, each with its own depth response profile (McNeill, 1980a). The device with a horizontally oriented dipole is most sensitive to the topsoil properties (0–

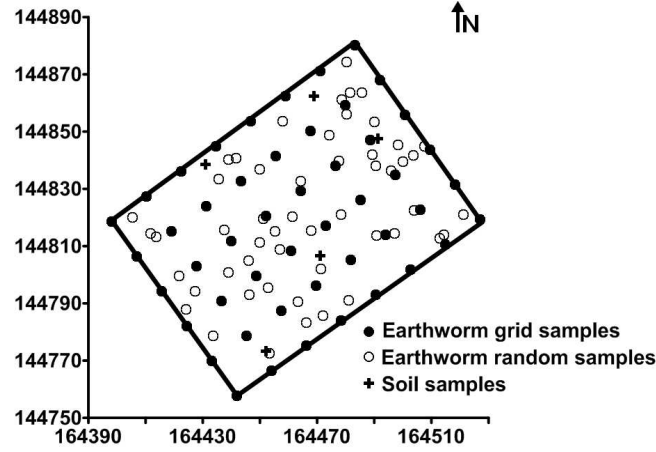


FIGURE 5.2 – Location of the earthworm samples (grid and random) and the soil samples. The coordinates are expressed in meters, conform the Belgian Lambert72 projection used in Belgium.

0.30 m), while characteristics of deeper soil layers (0.30–0.80 m) mainly influence the signal received by the vertical dipole. The EM38DD thus simultaneously measures the soil bulk electrical conductivity of shallow and deeper layers. When the soil profile is heterogeneous these two measurements will show a different spatial pattern. In the case of vertically homogenous soils, the two orientations yield a similar pattern and the spatial variability can be represented by their average.

5.2.2.3 Soil sampling

Based on the observed EC_a variability, soil samples at three depth intervals (0–0.30 m, 0.30–0.60 m and 0.60–0.90 m) were collected in the field at five locations, representing the range of EC_a values (Fig. 5.2). These samples were used to document the variability of the soil properties. Samples were analyzed for texture, pH_{KCl} , gravimetric moisture and organic carbon content for each depth interval.

5.2.3 Data analysis

Information on the spatial variability of the earthworm species density as extracted by the mustard method and the EC_a was obtained through geostatistical analysis. Omni-directional experimental variograms were computed for total populations, for adult and juvenile individuals separately and for the soil EC_a . All experimental variograms were best fitted with a spherical model (Eq. (3.4)). Based on the variogram parameters range, nugget and sill, the RNE was calculated to quantify the strength of the spatial structure since it is an index of variance not attributable to

spatial dependence (Isaaks and Srivastava, 1989). Strengths vary from weak ($RNE > 0.75$), to medium ($0.25 < RNE < 0.75$) and strong ($RNE > 0.25$) (Cambardella et al., 1994). Modeled variograms of a variable (e.g. earthworm density) were then used to estimate the variable's values at unsampled locations through ordinary kriging (OK). These interpolated earthworm and EC_a values were used as an individual input for a fuzzy- k -means classification procedure to identify natural clusters in the dataset (Minasny et al., 2006) (see section 3.2.1).

5.2.4 Measures of (dis)association

The influence of both biotic and abiotic conditions on the spatial variability of the species was investigated following two approaches: (i) a point-by-point correlation analysis of the 100 samples and (ii) a pixel-by-pixel comparison of their categorical classes as defined by the fuzzy- k -means classification procedure. In the point-by-point approach correlations were calculated as the Pearson correlation coefficient (r). The pixel-by-pixel method was based on a contingency table: the rows are formed by the classes of one categorical map while the columns are formed by the classes of the second map. Each cell in this table represents the number of pixels belonging to a certain combination of categories. The degree of agreement between pixel classifications was quantified using the κ and κ^* statistics (see section 3.2.2). However, the kappa statistics derived from a contingency table do not account for spatial relations, being similarities of spatial patterns (Power et al., 2001). Hagen (2003) introduced a new approach based on fuzzy set techniques to obtain a spatial analysis of the similarity of two maps. Fuzziness is introduced in the pixel-by-pixel approach by taking into account the categories of neighbouring cells when comparing two maps (also called fuzziness of location). Fuzziness of location is set by a distance decay membership function that defines to which level neighbouring pixels influence the fuzzy representation of the central cell. At distance 0 the function returns 1, indicating that every pixel fully belongs to its own category. As the distance increases, the function returns values between 0 and 1. Besides a distance decay membership function, also the size of the neighbourhood and the distance from every pixel in that neighbourhood to its central pixel should be determined. Therefore the concept of a neighbourhood ring is introduced. A neighbourhood ring is formed by all pixels located at the same distance from a central cell (Fig. 5.3).

An illustration of ring characteristics for a two pixel radius is given in Table 5.1. The radius defines the size of the neighbourhood, pixels beyond this radius are not considered part of the neighbourhood. To determine the radius of the neighbourhood, the range of the variograms was considered.

In this study, fuzziness of location was set with an exponential decay member-

			3			
		2	1	2		
	3	1	0	1	3	
		2	1	2		
			3			

FIGURE 5.3 – Numbered rings within a two pixel radius.**TABLE 5.1** – Ring characteristics for a two pixel radius.

	Ring			
	0	1	2	3
Number of pixels in ring	1	4	4	4
Number of pixels in neighbourhood	0	4	8	12
Distance from central pixel (expressed in pixels)	0	1	$\sqrt{2}$	2
Distance based membership value	1	0.71	0.61	0.5

ship function with a decay rate of two pixels:

$$f(d_i) = e^{\ln(\frac{1}{2})x\frac{d_i}{2}} = 2^{\frac{-d_i}{2}} \quad (5.1)$$

where $f(d_i)$ is the distance decay membership functions with distance d_i for the i^{th} ring.

A fuzzy approach to comparing two maps results in a fuzzy similarity map indicating for each cell the degree of similarity on a scale of 0 to 1 and allows the calculation of the average fuzzy similarity (S_a). When comparing two maps A and B, S_a is calculated following these steps:

1. Each pixel is represented by a fuzzy membership vector

To account for fuzziness of location, two types of membership vectors are defined: the Crisp Vector (\mathbf{V}_{crisp}) and the Fuzzy Neighbourhood Vector (\mathbf{V}_{nbh}). The size of the vectors correspond to the number of categories.

The membership values of the Crisp Vector are given by Eq. (5.2):

$$\mathbf{V}_{crisp} = \begin{pmatrix} \mu_{crisp,1} \\ \mu_{crisp,2} \\ \vdots \\ \mu_{crisp,c} \end{pmatrix} \quad (5.2)$$

with $\mu_{crisp,i} = 1$ if the original category i is found, otherwise $\mu_{crisp,i} = 0$ and c the number of categories.

The membership values of the Fuzzy Neighbourhood Vector are given by Eq. (5.3):

$$\mathbf{V}_{nbh} = \begin{pmatrix} \mu_{nbh,1} \\ \mu_{nbh,2} \\ \vdots \\ \mu_{nbh,c} \end{pmatrix} \quad (5.3)$$

$$\mu_{nbh,i} = |\mu_{crisp,i,1} * f(d_1), \mu_{crisp,i,2} * f(d_2), \dots, \mu_{crisp,i,m} * f(d_m)|_{Max}$$

with $\mu_{nbh,i}$ = the degree of membership of category i for all neighbouring pixels, set by the highest contribution of each category, $\mu_{crisp,i,j}$ = membership of category i for neighbouring pixel j in \mathbf{V}_{crisp} , $f(d_j)$ = distance based membership of neighbouring pixel j and m = the number of pixels in the neighbourhood.

2. Comparison of two fuzzy pixels

The similarity of the two maps A and B can be assessed by a pixel-by-pixel comparison of the fuzzy vectors assigned to all pixels. Eq.(5.4) expresses the similarity between a pixel in map A and one at the same location in map B. First, the minimum of a pairwise comparison of the vector elements for the same pixel on both maps is performed. Secondly, the maximum of these minima (maximum over all categories) is assigned to the pixel. The similarity values range from 0 for two completely dissimilar neighbourhoods to 1 for neighbourhoods with matching central pixels.

$$S(\mathbf{V}_{nbh,A}, \mathbf{V}_{nbh,B}) = [|\mu_{A,1}, \mu_{B,1}|_{Min}, |\mu_{A,2}, \mu_{B,2}|_{Min}, \dots, |\mu_{A,c}, \mu_{B,c}|_{Min}]_{Max} \quad (5.4)$$

3. Two-way comparison

To avoid an overpowering influence of the similarities between neighbourhoods (excluding the central pixel), the two-way comparison is introduced. First, the Fuzzy Neighbourhood Vector of pixel l on map A is compared to the Crisp Vector of the same location on map B. Next, the Crisp Vector of pixel l on map A is compared to the Fuzzy Neighbourhood Vector of the same location on map B. Finally, the minimum of the two comparison results forms the similarity at pixel l (Eq. (5.5)).

$$S_l = S_{l,TwoWay}(A, B) = |S(\mathbf{V}_{nbh,A}, \mathbf{V}_{crisp,B}), S(\mathbf{V}_{crisp,A}, \mathbf{V}_{nbh,B})|_{Min} \quad (5.5)$$

4. Average fuzzy similarity (S_a)

The fuzzy map similarity is calculated as the average similarity over all pixels (n):

$$S_a = \frac{\sum_{l=1}^n S_l}{n}. \quad (5.6)$$

S_a ranges from 0 (fully distinct maps) to 1 (identical maps).

Besides fuzziness of location, fuzziness in the membership of the categories can also be introduced. Fuzziness of category means that the distinction between categories is not always sharp and the vagueness among categories is set by a fuzzy category matrix. This type of fuzziness is useful for maps with qualitative (ordinal) legends where some pairs of categories are more similar than others. To account for the spatial context, fuzziness of category was not considered.

All map comparisons were carried out by the Map Comparison Kit software (Visser and de Nijs, 2006).

5.3 Results

5.3.1 Earthworm variability

In total seven earthworm species were observed of which the anecic *Lumbricus terrestris* L. was the most abundant species (Table 5.2). It was followed by the endogeic *Aporrectodea caliginosa* (SAVIGNY) and the anecic *A. longa* (UDE). The other endogeic species *Allolobophora chlorotica* (SAVIGNY), *Aporrectodea rosea* (SAVIGNY) and *Octolasion cyaneum* (SAVIGNY) were less frequently encountered. *Dendrodrilus rubidus* (SAVIGNY), the only epigeic species, was found only at two locations near the southern border of the sampled area. The same relative rankings were observed for adults and juveniles. As an exception to this general pattern, adult *A. caliginosa* individuals occurred in higher densities than *L. terrestris* adults. Anecic populations were dominated by juvenile individuals (adult/juvenile ratio of 0.06 and 0.23 respectively for *L. terrestris* and *A. longa*) while the adult/juvenile ratios of the endogeic species were much higher (0.42, 0.55, 0.64 and 2.52 respectively for *O. cyaneum*, *A. caliginosa*, *A. rosea* and *A. chlorotica*).

Table 5.3 shows the variogram parameters based on earthworm densities for the anecic species and two representative endogeic species (*A. chlorotica* and *A. caliginosa*). In general, the fitted variogram models showed medium to strong spatial structures in earthworm species with the RNE varying between 17% and 58% (e.g. Fig. 5.4(a) and (b)). However, the densities of *A. rosea* and *A. longa* adults were only weakly spatially correlated (RNE = 77% and 71%, respectively), while the density of *L. terrestris* adults showed a strong spatial structure (RNE =

TABLE 5.2 – Mean ($n = 100$) earthworm densities as collected by mustard extraction in a conventionally ploughed arable field. Data are given for each developmental stage per earthworm species and per ecological category

Earthworm species/Ecological category	Density [individuals m ⁻²]					
	Adults		Juveniles		Total	
	Mean	SD	Mean	SD	Mean	SD
<i>Allolobophora chlorotica</i>	5.2	7.50	2.0	3.35	7.2	9.26
<i>Aporrectodea caliginosa</i>	14.3	10.23	26.0	16.54	40.3	23.54
<i>Aporrectodea rosea</i>	1.9	2.46	3.0	3.51	4.9	4.89
<i>Octolasion cyaneum</i>	0.2	0.65	0.6	1.42	0.8	1.71
Endogeics total	21.6	13.74	31.6	18.41	53.2	28.65
<i>Aporrectodea longa</i>	1.6	2.12	6.9	5.88	8.5	6.63
<i>Lumbricus terrestris</i>	9.1	5.24	138.5	55.14	147.6	54.53
Anecics total	10.7	5.88	145.3	54.78	156.1	53.93
<i>Dendrodrilus rubidus</i>	0.0	0.28	0.0	0.00	0.0	0.28
Epigeics total	0.0	0.28	0.0	0.00	0.0	0.28
Total	32.4	16.23	177.0	58.89	209.3	63.25

17%). The ranges of autocorrelation varied between 14 and 64 m. For the anecic species, the variogram range was remarkably larger for the juveniles than for the adults (e.g. 14 m vs. 34 m for *L. terrestris* adults and juveniles, respectively). As anecic population structures were dominated by juveniles, the variogram ranges of total *L. terrestris* and *A. longa* populations were similar to the range of juveniles of these species (30 m vs. 34 m for *L. terrestris* and 64 m vs. 63 m for *A. longa*). In contrast, the endogeic populations had a more even distribution of adults versus juveniles and the ranges of the total population were therefore comparable to the ranges of adults or juveniles separately.

TABLE 5.3 – Variogram model parameters for each developmental stage of the representative earthworm species density. All data were best fit by a spherical model

Earthworm species	Developmental stage	Nugget C ₀	Range a	Sill C ₁	RNE [C ₀ /(C ₀ +C ₁)]*100
<i>L. terrestris</i>	-	1010	30	1960	34
	Adult	5	14	24	17
	Juvenile	1062	34	2094	34
<i>A. longa</i>	-	21	64	27	44
	Adult	3.1	40	1.2	71
	Juvenile	18.5	63	18.2	50
<i>A. caliginosa</i>	-	222	29	354	39
	Adult	72	33	41	29
	Juvenile	123	30	143	46
<i>A. rosea</i>	-	12.6	45	11	53
	Adult	4.6	40	1.4	77
	Juvenile	6.8	44	4.9	58

In Figure 5.5 it can be observed that *A. longa* densities were higher in those areas where *L. terrestris* is less abundant and the reverse. On the other hand, the spatial distributions of both endogeic species were largely confined to the same area in the field and overlapped to some degree. Generally, densities of endogeic species did not show a spatial correspondence with those of anecic species.

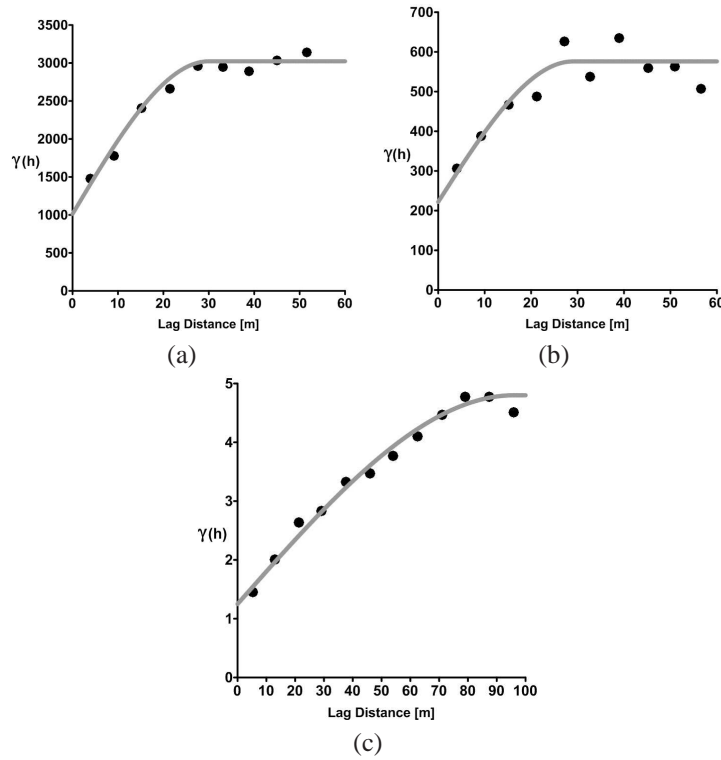


FIGURE 5.4 – Experimental (dots) and fitted (line) variograms of (a) *L. terrestris* densities, (b) *A. caliginosa* densities and (c) EC_{av} .

5.3.2 EC_a variability

The maps of the horizontal and vertical EC_a displayed similar patterns but higher absolute EC_a values were measured in the vertical orientation. The EC_a measured in both orientations was closely correlated ($r = 0.819$), so we decided to perform the further analysis of EC_a variability with the arithmetic average of both orientation, noted as EC_{av} . EC_{av} data ranged between 10.4 and 22.7 $mS\ m^{-1}$, with a mean of 16.5 $mS\ m^{-1}$ and a standard deviation of 2.1 $mS\ m^{-1}$. The EC_{av} variogram (Fig. 5.4(c)) had a strong spatial structure, indicated by its RNE of 22.9%, meaning that only 22.9% of the variability consists of inexplicable or random variations. The range of 96 m indicates the spatial extent of the EC_{av} autocorrelation.

The EC_{av} map shows a variation from low values in the southeast boundary to high values at the upper northwest boundary (Fig. 5.6(a)). This map was subjected to a fuzzy- k -means classification which resulted in an optimum of 2 EC_{av} classes (Fig. 5.6(b)). Class 1 had a centroid EC_{av} value of 14.4 $mS\ m^{-1}$, while class 2 had

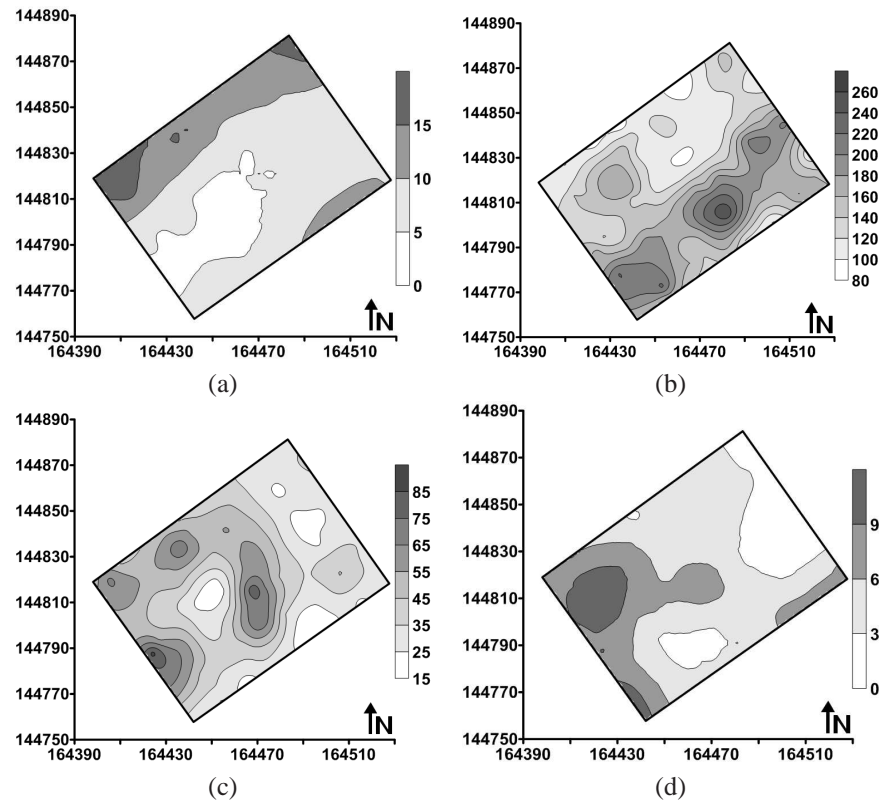


FIGURE 5.5 – Kriged contour maps of the densities [individuals m^{-2}] of (a) *A. longa*, (b) *L. terrestris*, (c) *A. caliginosa* and (d) *A. rosea*.

a centroid value of 17.9 mS m^{-1} . The textural difference between these two classes could clearly be shown by their clay content, averaged over the soil samples per EC_{av} class (Fig. 5.7). EC_{av} class 2 was characterized by a distinctly higher clay content than EC_{av} class 1. In both classes the clay content increased with depth, but for EC_{av} class 2 the increase in clay content was slightly higher than for class 1 (a relative increase of the clay content from top- to subsoil of 54.8 % versus 42.5 %). Similarly the moisture content was higher in class 2, but the difference between the classes was not significant. Also organic carbon and pH were not significantly different between the two EC_{av} classes: their depth weighted mean was 0.56 % and 6.84 respectively for class 1, and 0.51 % and 6.95 respectively for class 2.

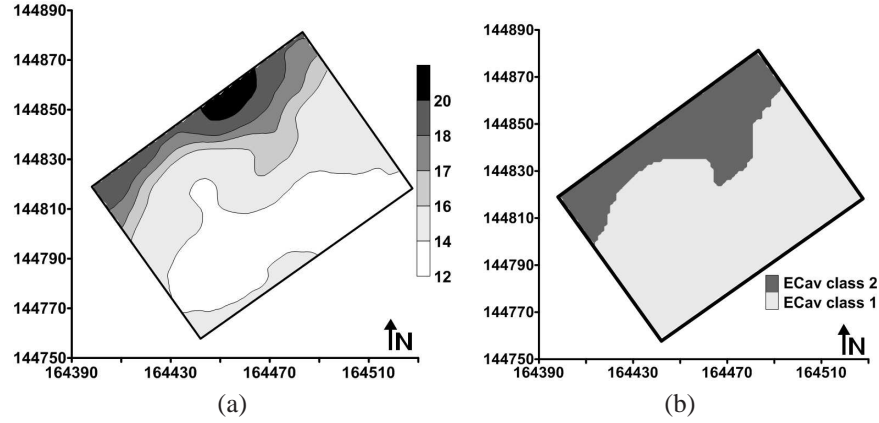


FIGURE 5.6 – (a) Kriged contour map of EC_{av} [$mS\ m^{-1}$] and (b) Fuzzy- k -means classification of EC_{av} .

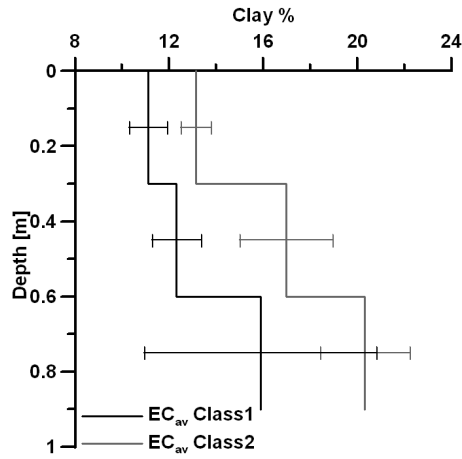


FIGURE 5.7 – Clay content [%] per depth interval in each EC_{av} class, the error bars depict the standard deviations from the mean.

5.3.3 Measures of (dis)association

5.3.3.1 Abiotic-biotic comparisons

Pearson correlations between the abiotic soil conditions, represented by the EC_{av} values, and earthworm densities were substantial for total *L. terrestris* ($r = -0.471$), *L. terrestris* juveniles ($r = -0.472$) and total *A. longa* ($r = 0.481$). Abundance of *L. terrestris* was clearly negatively correlated with EC_{av} , and this correlation was mostly driven by juveniles. *A. longa* abundance had a strong

positive correlation with EC_{av} , indicating a preference for soil with relatively higher clay content. The endogeic species did not show any correlation with EC_{av} ($r < 0.200$).

Since the EC_{av} map was optimally split into two classes ($k = 2$), and a kappa analysis requires the same number of classes in the maps to be compared, this division was maintained for the fuzzy- k -means classification of the biotic maps (Fig. 5.8). The results of the categorical comparison of the earthworm species and EC_{av} confirmed the correlation analysis (Table 5.4). The strong spatial agreement between *A. longa* and EC_{av} was expressed by the S_a of 0.838, indicating the occurrence of *A. longa* in locations in the field with the highest clay content. For *L. terrestris* (total and juvenile densities) the spatial agreement was also high ($S_a = 0.739$), but an opposite behaviour towards soil conditions was observed (negative r). It should be mentioned that in case of negative Pearson correlation coefficient the map comparisons were performed with inverted categories for one variable to avoid negative indices.

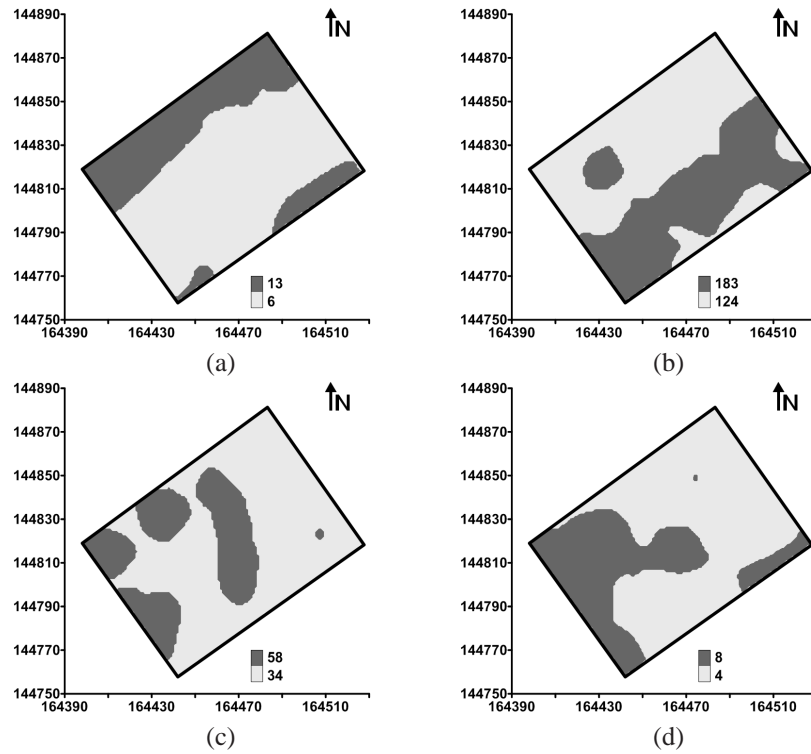


FIGURE 5.8 – Fuzzy- k -means classification (with $k = 2$) of (a) *A. longa*, (b) *L. terrestris*, (c) *A. caliginosa* and (d) *A. rosea*. Centroid values per class are given in the legend [individuals m^{-2}].

TABLE 5.4 – Kappa indices (κ and κ^*) and average fuzzy similarity (S_a) of classified maps related to biotic and abiotic interactions.

		κ	κ^*	S_a
Biotic interactions				
Community level	Anecics: <i>L. terrestris</i> vs <i>A. longa</i>	0.388	0.360	0.699
	Endogeics: <i>A. caliginosa</i> vs <i>A. rosea</i>	0.310	0.398	0.726
Population level	<i>L. terrestris</i> : juvenile vs adult	0.027	0.020	0.554
	<i>A. caliginosa</i> : juvenile vs adult	0.681	0.678	0.859
Abiotic interactions (EC_{av} vs species)				
	<i>L. terrestris</i>	0.478	0.450	0.739
	<i>L. terrestris</i> adult	0.110	0.124	0.462
	<i>L. terrestris</i> juvenile	0.422	0.368	0.706
	<i>A. longa</i>	0.646	0.698	0.838
	<i>A. caliginosa</i>	0.193	0.328	0.683
	<i>A. caliginosa</i> adult	0.040	0.018	0.521
	<i>A. caliginosa</i> juvenile	0.255	0.282	0.634
	<i>A. rosea</i>	0.098	-0.022	0.525

5.3.3.2 Community level comparisons

The occurrences of both anecic species were weakly negatively correlated ($r = -0.151$) while on the other hand, occurrences of the representative endogeic species *A. caliginosa* and *A. rosea* were positively correlated ($r = 0.362$). Pixel-by-pixel comparison of the categorical maps of the anecic species (based on inverted categories for *L. terrestris*) resulted in a κ^* of 0.360, while S_a was 0.699 (Fig. 5.8, Table 5.4). Thus, although the linear correlation between *L. terrestris* and *A. longa* was poor, a similar spatial variability was found, indicating a considerable degree of spatial exclusion. For the endogeic species (Fig. 5.8) the degree of agreement as expressed by S_a was somewhat larger (0.726), indicating a slightly stronger (and positive) spatial association.

5.3.3.3 Population level comparisons

Occurrences of *L. terrestris* development stages were weakly negatively correlated ($r = -0.163$). The degree of agreement between the categorical maps of *L. terrestris* adults (inverted categories) and juveniles was poor ($\kappa^* = 0.020$), indicating also a low degree of map similarity (Fig. 5.9, Table 5.4). The linear correlation between the development stages of both endogeic species was stronger and positive: $r = 0.520$ for *A. caliginosa* and $r = 0.318$ for *A. rosea*. Comparison of the maps of *A. caliginosa* adults and juveniles yielded a high S_a (0.859), confirming a strong spatial agreement (Fig. 5.9).

5.4 Discussion

This study demonstrated the spatial structure of earthworm species occurrence within a tilled agricultural field. The degree of spatial dependence was in general

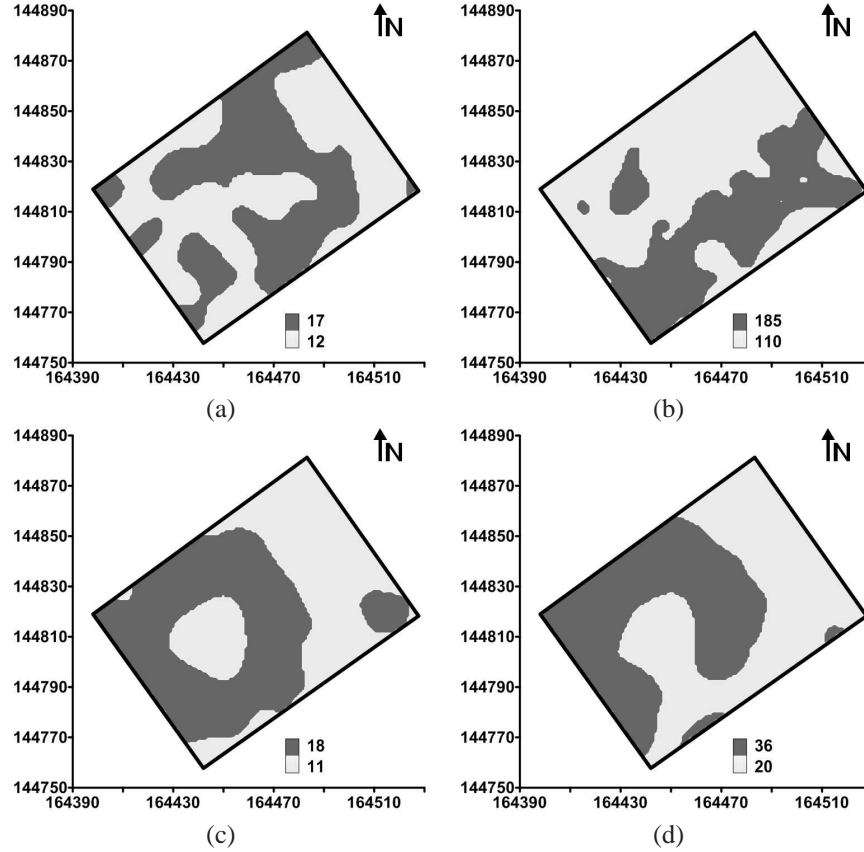


FIGURE 5.9 – Fuzzy- k -means classification (with $k = 2$) of the developmental stages for one representative species per ecological category: (a) *L. terrestris* adults, (b) *L. terrestris* juveniles, (c) *A. caliginosa* adults and (d) *A. caliginosa* juveniles. Centroid values per class are given in the legend [individuals m^{-2}].

medium strong with RNE varying between 17% and 58%. However, the distributions of *A. longa* and *A. rosea* adults had a weak spatial structure with RNEs of 71% and 77% respectively. The methodological count resolution of one individual combined with the low occurrences of both species in the field (on average < 2 individuals m^{-2}) inevitably leads to a high proportion of unstructured variability. It was found that earthworms live in spatially distinct clusters whose size was species- and life stage-specific. Map comparisons showed that the endogeic species *A. caliginosa* and *A. rosea* lived in close association with each other. The anecic species on the other hand had dissimilar spatial patterns as indicated by a fair agreement between the maps of *L. terrestris* (inverted categories) and *A. longa* on the scale of Landis and Koch (1977).

The results also showed that the field distributions of deep-burrowing species were strongly spatially correlated with the soil conditions as measured by the EC_{av} . Occurrences of adult and juvenile *A. longa* were confined to zones with high EC_{av} values while conversely *L. terrestris* (particularly juveniles) appeared in areas with low EC_{av} values. Also Blackshaw et al. (2007) found that juvenile abundance was spatially correlated with soil properties in an arable field, although they did not differentiate between species. The spatial segregation of both deep-burrowing species was driven by a preference for differing soil conditions rather than by negative species interactions, such as competition for food at the soil surface. Field distributions of the shallow dwelling species were not spatially related to soil EC_{av} conditions. Given the homogenizing effect of the long-term applied ploughing regime on soil properties in the upper 20–30 cm of the soil profile, the EC_{av} measurements are likely to represent the non-homogenized subsoil properties (Cockx et al., 2008). As a consequence the EC_{av} map does not represent the topsoil variability well. The practical implication of these observations is that the use of EC_{av} measurements to predict the spatial occurrence of earthworm species is confined to earthworm activity of the deeper soil layers.

Decaens et al. (2003) found that soil type was a driving factor for the spatial distribution of earthworm community structure and diversity at the landscape scale, while at the field scale the agricultural practices had most impact. At our study site with uniform management practices we found that the spatial distribution of deep burrowing earthworm species was related to the within-field spatial soil variability, measured by the EC_a . As a consequence, the rapid, easy and non-destructive geo-referenced soil characterization by means of an EMI sensor can be used to design an optimal earthworm sampling scheme since the soil EC_a variability provides an indication of the spatial distribution of deep burrowing species. Also Johnson et al. (2004) demonstrated the biological relevance of soil EC_a classification and stated that it provides a point of reference through which microbial, within-field-, and field-scale data can be related. However, this novel sensor-based approach to determine and understand the spatial distributions of soil organisms requires further testing in a variety of (agro-)ecosystems.

Concerning the measures of (dis)association, it can be seen from Table 5.4 that the difference between the two kappa statistics was rather limited. In this case κ did not overestimate agreement by chance and κ^* was not more informative than κ . This can be explained by the even distribution in both the row and column totals of the contingency tables: $n_{i+} \approx n_{+i} \approx n/2$ (see Table 3.1). On the other hand, taking into account the local neighbourhood by a fuzzy pixel-by-pixel comparison resulted in an increased agreement index: for all comparisons S_a was higher than the kappa indices. Even for small kappa indices, the S_a was still considerable good. This implies a cautious use of S_a . To interpret the degree of association between two maps it is advised to use a combination of accuracy measures.

5.5 Conclusion

Through a mobile, non-invasive EC_a soil survey clusters of high anecic activity within an arable field were identified. It was also found that both the spatial patterns of the earthworm densities and the degree of association of biotic and abiotic interactions was species and life stage specific and related to the ecological category of the earthworm species concerned.

The endogeic species *A. caliginosa* and *A. rosea* inhabited similarly sized and overlapping patches ($r = 0.362$, $\kappa^* = 0.398$ and $S_a = 0.726$). Clusters of *A. caliginosa* adults and juveniles were situated in the same locations in the field ($S_a = 0.859$). At the same time it was observed that these spatial similarities between and within endogeic populations were not related to the measured soil EC_a variability. In contrast, both deep-burrowing species *A. longa* and *L. terrestris* were found to live in spatially separated zones ($r = -0.151$, $\kappa^* = 0.360$ and $S_a = 0.399$) and their spatial distributions were mainly driven by the abundance of juveniles. A high degree of spatial similarity was found between the maps of the anecics and the EC_a : the S_a was 0.739 for *L. terrestris* and 0.838 for *A. longa*. This association was attributed to the preference of these species to slightly differing subsoil conditions: their spatial distribution was determined by the subsoil properties.

Our results proved that an EM38DD based soil EC_a survey is a promising approach for identifying the spatial variability of deep burrowing earthworm species. This opens perspectives for more efficient earthworm sampling based on an EMI guided sampling strategy.

6

Prospecting quaternary frost-wedges and their polygonal network with the EM38DD soil sensor

The content of this chapter was published as:

Cockx, L., Ghysels, G., Van Meirvenne, M., and I. Heyse. (2006). Prospecting frost-wedge pseudomorphs and their polygonal network using the electromagnetic induction sensor EM38DD. *Permafrost and Periglacial Processes*, 17: 163-168.

6.1 Introduction

Frost-wedge pseudomorphs are sedimentary structures representing the imprint of former frost-fissure wedges, developed under palaeo-periglacial conditions (Harry and Gozdzik, 1988; French, 1996). Frost-fissure wedges are currently formed in present-day cold environments at high latitudes, such as Canada, Greenland, and Siberia. They form by the progressive infilling of thermal contraction cracks with ice, non-ice material or a mixture, resulting in wedge-shaped bodies (Murton and French, 1993). Detection and identification of their casts or pseudomorphs give fundamental information about Pleistocene periglacial environments (Vandenbergh and Pissart, 1993). The term ‘pseudomorph’ is used for frost-wedges containing ice that experienced thaw modifications and usually bear little resemblance to the original form. During thaw modification, material from above and/or the sides of the wedge replaced the original ice volume.

Methods used for detecting wedge casts include: (i) direct observation in trenches and excavations, or (ii) indirect observation of their polygonal networks using aerial photographs (Boike and Yoshikawa, 2003; Svensson, 1972; Walters, 1994) or geophysical methods like ground penetrating radar (Dansart et al., 1999) and electrical resistivity (Greenhouse and Morgan, 1977). Indirect observation techniques are preferred since they have least impact on the soil. However, they give a very fragmentary view of the total distribution of near-surface frost-wedge pseudomorphs, because they are (i) slow or (ii) time dependent or (iii) controlled by a variety of external factors. On aerial photographs, polygonal networks are visible through differential ripening of the crops resulting in ‘crop marks’. An example is given in Fig 6.1.

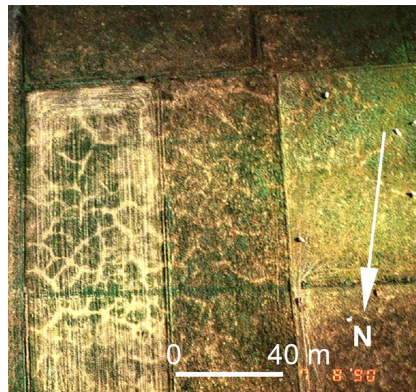


FIGURE 6.1 – Oblique aerial photograph of large scale polygonal crop marks in North-west Belgium (photo: J. Semey, source: Department of Archeology and Ancient History of Europe, Ghent University).

The visibility of these crop marks depends on the crop type, the meteorological conditions before and during the aerial survey, the thickness of the Quaternary cover and the contrast in lithology between host (adjacent) and wedge materials. Less dependent on these external factors are geophysical methods. Nevertheless only little research has focused on these methods. Moreover, the already investigated geophysical methods were based on wedges visible on aerial photographs (Dansart et al., 1999). Small-scale resistivity profiling was used for locating previously identified polygon boundaries, but for large areas this method is slow and tedious (Greenhouse and Morgan, 1977).

The focus of this paper is on locating and mapping frost-wedge pseudomorphs using EMI. The major advantages of EMI are (i) it is non-invasive; (ii) it gives an immediate response; (iii) the signal has a good penetration capacity and (iv) the signal response can be related to soil composition. We used EMI at a test site in Sint-Niklaas (Belgium) to develop a new methodology for locating near-surface frost-wedge pseudomorphs. At this site a polygonal pattern of frost-wedge pseudomorphs was exposed during temporary excavation. Besides the methodological aspects, the advantages and restrictions of the technique are discussed.

6.2 Material and Methods

6.2.1 Study site

The study site is located on a low-lying plateau (elevation about 25 m a.s.l.), in Sint-Niklaas, Belgium (51°08'26"N, 4°10'24"E) and is known as the Waasland Plateau, an adjacent interfluvium of the Flemish Valley. The Waasland plateau is underlain by a more than 1.5 m thick cover of Quaternary fluvio-aeolian and aeolian sediments (coversands) overlaying the Tertiary marine glauconiferous sand and clay. The site is located in the sandy region of Flanders (Fig. 6.2). According to the Belgian soil classification, the soil series is 'Pbm', a dry (drainage class 'b') sandy silt topsoil texture (textural class 'P') with a thick anthropogenic humic A horizon (profile development class 'm').

A polygonal pattern of sand-filled wedges was exposed during excavation works which removed the top 3 m soil (Fig. 6.3 (a)). The wedges were 0.02 – 1.5 m wide and 0.2 – > 2.50 m deep and the pattern comprises polygonal cells ranging from 0.2 m to > 10 m in diameter (Fig. 6.3). The study site was a rectangle 7.5 by 14.5 m, comprising a variety of well developed polygonal structures. Ground wedges with a primary sandy filling are commonly referred to as 'sand-wedges'. A primary filling is the original infilling of thermal contraction cracks, while secondary infillings refer to sediments that replace the original ice content. The sand-wedge structures are interpreted as frost-wedge pseudomorphs suggesting thermal contraction cracking in a periglacial environment during the Pleistocene (Ghysels

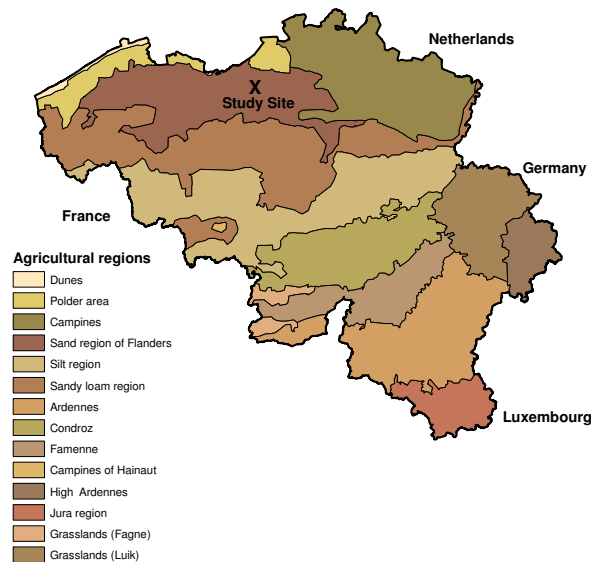


FIGURE 6.2 – Location of the study site in the sandy region of Flanders (Belgium).

and Heyse, 2006; Heyse and Ghysels, 2003). Here we use the term ‘wedges’ to refer to these frost-wedge pseudomorphs.

6.2.2 Methods

6.2.2.1 EM38DD survey

Differences in apparent soil EC_a were measured by electromagnetic induction using a Geonics EM38DD. The EM38DD simultaneously measures the EC_a (expressed in $mS\ m^{-1}$) in the two orientations, each having a different depth response profile. The vertical orientation (EC_a -V) receives a dominant influence from deeper (0.3 - 0.5 m) soil layers, while the horizontal orientation (EC_a -H) receives its major influence from the near-surface soil layers (McNeill, 1980b). The two sensor orientations allow for relative measures of soil EC_a like the profile ratio (PR). The PR is calculated as the ratio of EC_a -H to EC_a -V and reveals differences in the soil profile. A PR greater than unity indicates that the topsoil is more conductive than the subsoil and the reverse is true for a PR less than unity. Values of PR close to unity indicate a homogeneous profile.

The wedges were visibly filled with a different mineral material which could be expected to result in differences in EC_a (see Fig. 6.3 (b)). EC_a measurements were made at 0.5 m intervals by walking along lines spaced 0.5 m apart, creating a 0.5 by 0.5 m grid. EC_a -H and EC_a -V were recorded at each measurement location. Grid

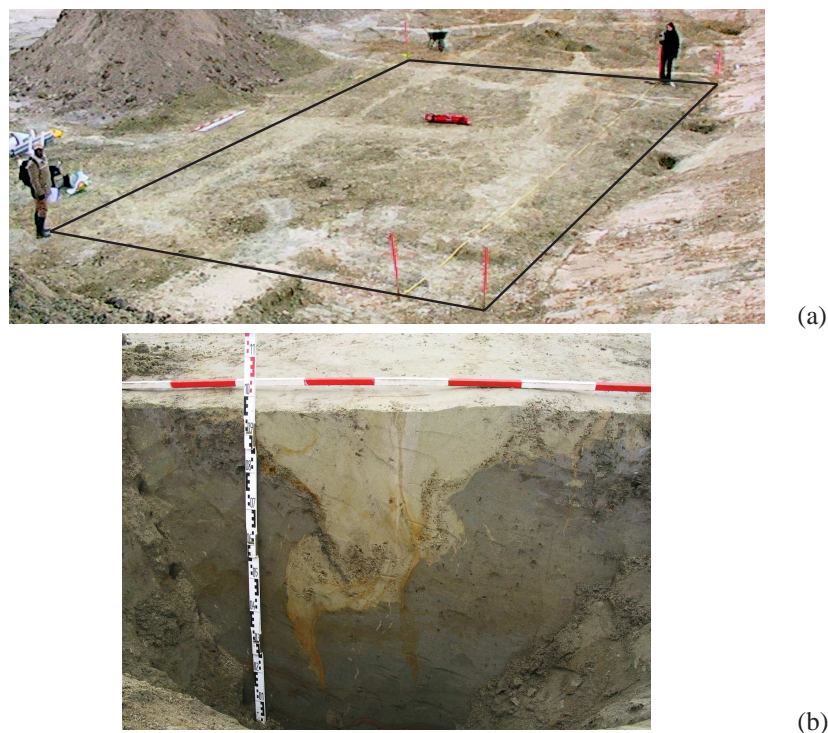


FIGURE 6.3 – (a) Study area showing the polygonal pattern and the selected rectangle with the EM38DD sensor in the centre and (b) a vertical cross-section of a sand-filled wedge (the horizontal scale bar shows intervals of 0.20 m).

coordinates were used to reference the measurements. Because the instrument has a length of 1 m, an overlap of 0.5 m occurred between successive measurements.

6.2.2.2 Validation of the polygonal network

For validation, a detailed drawing of the polygonal network was made after cleaning the soil surface in March 2004. The visible polygonal cells were drawn on graph paper, aided by a grid of 0.10 by 0.10 m made by strings tightened in a wooden frame. This drawing was digitized first and then rasterized in IDRISI with a pixel resolution of 0.25 m by 0.25 m (Fig. 6.4). The resolution was a compromise between the detailed drawing and the sensitivity of the EM38DD sensor. Sensitivity tests of the EM38DD in the field showed that the width of a wedge had to be at least 0.25 m to differentiate it from the host material. To be able to identify these small wedges the resolution was set to 0.25 m. As a consequence, only wedges wider than 0.25 m were rasterized and considered in further analyses.

This raster of the polygonal network was used as the reference or validation image of the polygonal pattern (Fig. 6.4(b)).

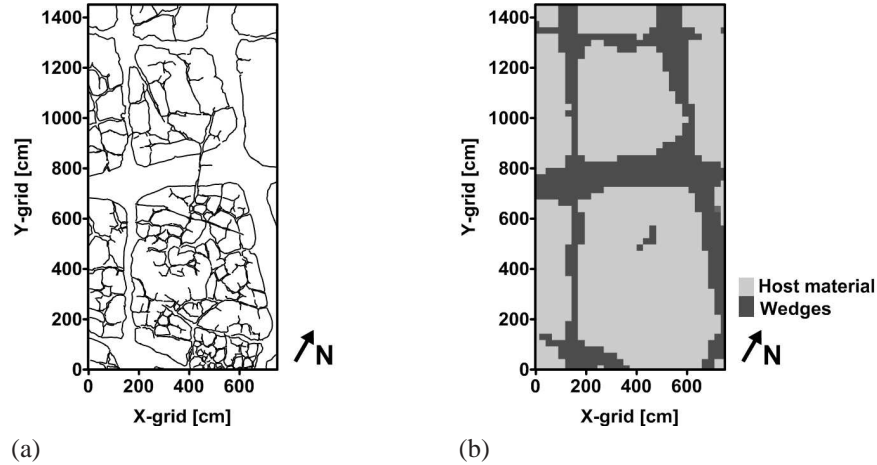


FIGURE 6.4 – (a) Drawing of the polygonal pattern and (b) Validation raster of the polygonal pattern in which only wedges wider than 0.25 m were retained. The axes represent grid coordinates expressed in cm.

6.2.2.3 Accuracy of EM38DD based frost-wedge allocation

The EC_a measurements were interpolated to the same resolution as the validation image using ordinary kriging (Goovaerts, 1997). Since we aimed at locating the polygonal pattern of wedges, a fuzzy- k -means classification was performed after the interpolation to obtain a categorical image (see section 3.2.1). The results from the fuzzy- k -means classification were compared with the validation image by cross-tabulation and the accuracy of the classification was obtained from a confusion matrix (see section 3.2.2).

6.3 Results and Discussion

6.3.1 Soil texture and EC_a measurements

Material inside the wedges was paler (7.5Y7/3 (light yellow), 5Y6/2 (greyish olive), 2.5Y6/6 (bright yellowish brown)) than the host material (7.5Y4/3 (dark olive) and 2.5Y4/6 (olive brown)) determined using a standard soil color chart (Oyama and Takehara, 1970). At two locations, in and outside a wedge, soil samples were taken at 0.30 m intervals and analysed for their textural fractions

and Fe-content (Table 6.1). The textural fractions were determined by the pipette method (Gee and Bauder, 1986), the Fe-content was determined by atomic absorption spectrophotometry (AAS) of ammoniumoxalate extracts (using a Varian Spectra AA-20Plus spectrophotometer). Their sand contents were similar (about 80–90 %), but the clay content was higher in the host material (14%) than in the wedges (9%). The host material was Fe-rich sand (due to the presence of glauconite) explaining the darker colour, whereas the wedges were paler (white sand) (Figure 6.3(b)). The difference in texture results from a differential origin of the sediments: the cracks opened by thermal contraction were filled with wind-blown sands, while the host material consists of marine sand and clay.

TABLE 6.1 – Textural fractions and Fe-content of two samples taken at 0.30 m intervals in the host material and in a wedge.

	Host material			Wedge material		
	Clay [%]	Sand [%]	Fe [mmol/kg soil]	Clay [%]	Sand [%]	Fe [mmol/kg soil]
0-0.30 m	13.0	82.0	3.04	8.3	80.8	1.78
0.30-0.60 m	13.0	82.1	2.65	6.9	90.4	1.78
0.60-0.90 m	16.8	77.3	2.64	11.0	84.9	2.64

A summary of the descriptive statistics of the EC_a measurements is given in Table 6.2. The EC_a -H values were smaller than the EC_a -V measurements but displayed more variability with a CV of 10.9%. The correlation coefficient between the two orientations was low ($r = 0.30$).

TABLE 6.2 – Descriptive EC_a statistics ($n = 480$).

	Min	Max	Mean	Median	CV [%]	Skewness
EC_a -V [$mS\ m^{-1}$]	40	55	46.3	46	5.8	0.36
EC_a -H [$mS\ m^{-1}$]	21	36	27.7	28	10.9	0.03

Both EC_a variograms displayed a nested model structure resulting from two scales of variation. The small scale variability with a range around 1.30 m was attributed to the polygonal pattern of frost-wedge pseudomorphs, while the EC_a variability of the host soil was considered to be the source of variation with a range around 6 m. The two scales of variation are most apparent from an experimental variogram with long lag distances (Fig. 6.5(a) and (c)). Commonly the maximal lag distance is set to half of the maximal sampling dimension. Beyond this distance the quantity of pairs will decrease since only observations situated near the edges are involved. Then the variogram loses its representativity: the randomness of variogram calculating value increases, the precision of model fitting decreases and even possibly the regularity of variability can be distorted (Cambardella et al.,

1994). For this study site this corresponds to a maximal lag distance of 8.16 m. On Figure 6.5(a) it can be seen that at lag distances longer than 7 m, the number of pairs started to decrease. Nevertheless only the beginning part of those variograms was relevant for the interpolation and to this part a double spherical model was fit (Fig. 6.5(b) and (d)). The RNE was 9.6% and 10.6% for EC_a -H, EC_a -V respectively.

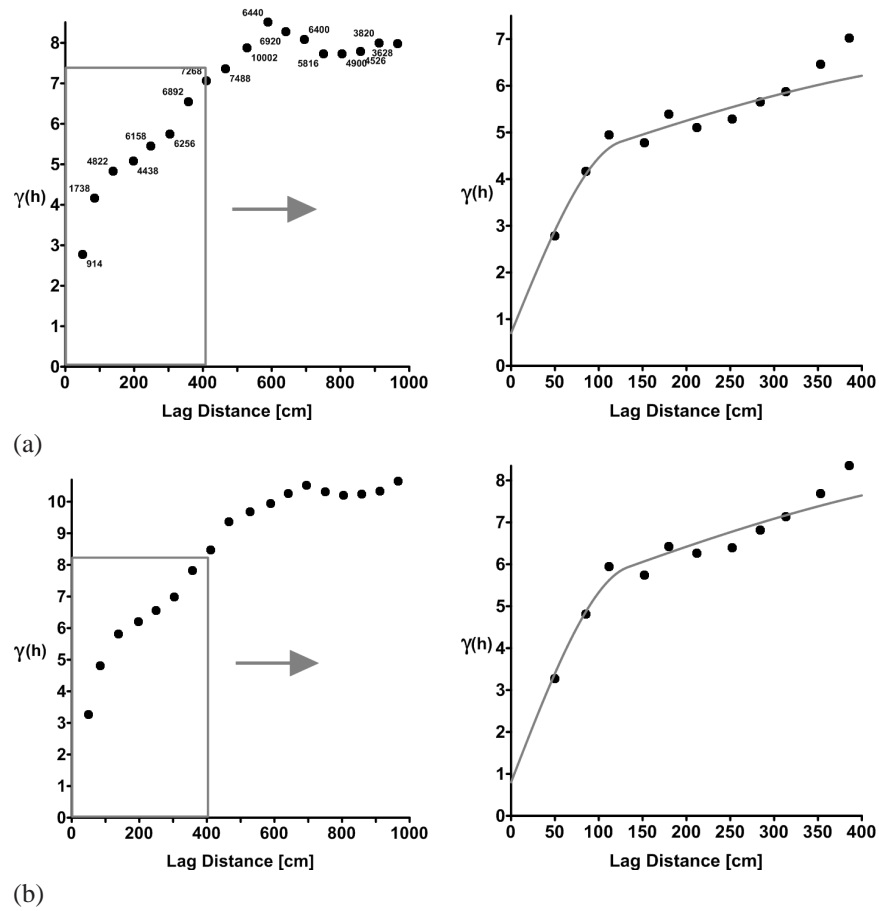


FIGURE 6.5 – Experimental and modeled variograms of (a) EC_a -V and (b) EC_a -H. The left part of the figure shows the experimental variogram till long lag distances, while the right graph has a maximum lag distance of 400 m. Based on the left graph the variogram was modeled.

The interpolated EC_a maps show that for both EM38DD orientations lower EC_a values were found on the right side of the study area (Fig. 6.6), indicating

spatial soil variability in the host material. Since no polygonal pattern was visible in the EC_a -V or EC_a -H maps, the host soil variability was assumed to be larger than the absolute EC_a -V or EC_a -H differences between wedge fillings and host material.

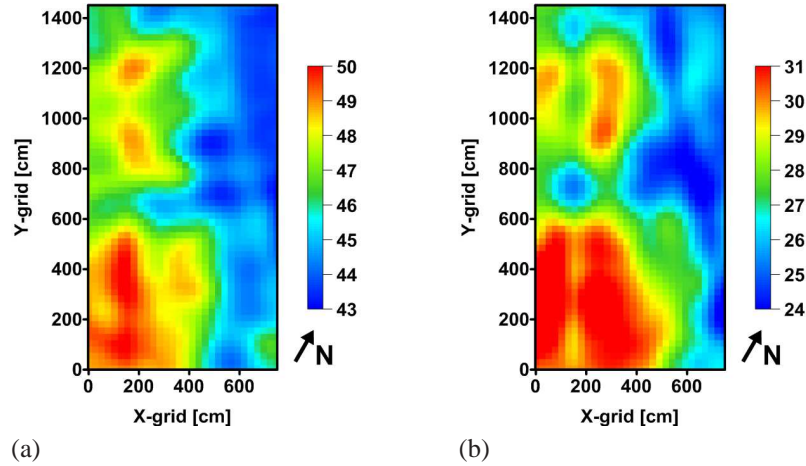


FIGURE 6.6 – Map of (a) EC_a -V and (b) EC_a -H. The EC_a maps are expressed in $mS\ m^{-1}$.

Table 6.3 lists the descriptive statistics of the EC_a variables stratified on basis of the validation image. The differences of the mean values in and outside wedge material were negligible for EC_a -V (0.2%) and small for EC_a -H (6.7%). The lower conductivity in the wedges can be explained by the lower clay content. The EC_a variability was comparable for the wedge and host material. However a difference in correlation between the two orientations was observed: for EC_a measurements located in the host material r increased to 0.58, while in the wedge material the correlation between EC_a -V and EC_a -H decreased to 0.19. This is probably related to the decreasing width of the wedges with depth: in the wedges the subsoil differs more from the topsoil than in the host material.

TABLE 6.3 – Descriptive EC_a statistics, stratified according to the validation image.

	Host material		Wedge material	
	Mean	CV [%]	Mean	CV [%]
EC_a -V	46.4	3.94	46.3	4.14
EC_a -H	28.3	7.08	26.4	7.64
PR	0.61	4.35	0.56	4.63

Looking relatively, the mean PR value of wedge and host material also differed only slightly (6.5%). In contrast, the PR map clearly revealed the polygonal pattern (Fig. 6.7(a)). Wedges were represented by local PR minima. By calculating the PR, the EC_a variability of the host material was eliminated and the PR map emphasized the wedge/host material contrast.

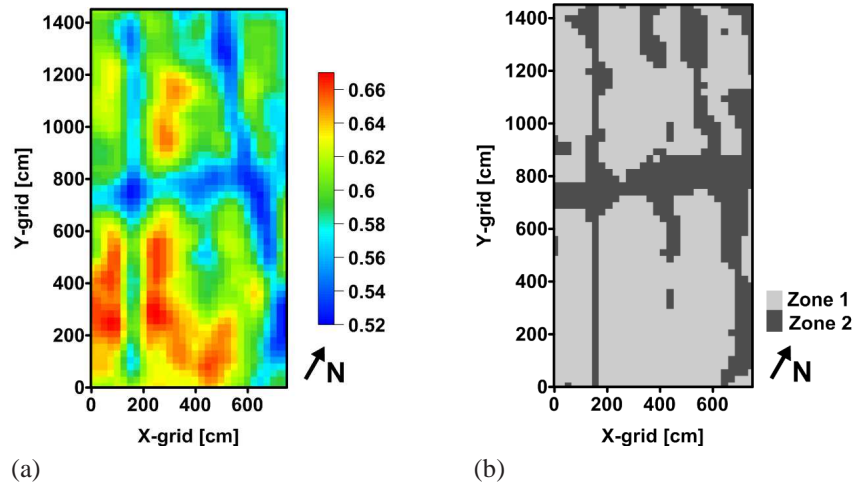


FIGURE 6.7 – (a) Map of PR and (b) fuzzy- k -means classification of PR.

6.3.2 Predictive quality of the classified PR map

Based on the PR map the study area was classified using a fuzzy- k -means classification (with $k = 2$) shown in Fig. 6.7(b). Visually, the PR zones correspond well to the validation image of the polygonal pattern (Fig. 6.4 (b)). A map comparison was performed to evaluate the accuracy of the PR classification. The confusion matrix is shown in Table 6.4.

TABLE 6.4 – Confusion matrix obtained after cross-tabulation of the pixels of the validation image (columns) and the PR classification (rows).

	Host material	Wedge	Total
PR zone 1	1081	142	1223
PR zone 2	181	336	517
Total	1262	478	1740

The PR classification predicted the polygonal pattern of wedges with an overall accuracy or observed proportion correct (P_o) of 81%, so 19% of the pixels were

misclassified. κ was 0.546, indicating a reasonable agreement between the two maps. Taking into account the categories of neighbouring pixels within a radius of 1 m (4 pixels) increased the accuracy: the fuzzy similarity S_a was 0.907 indicating a good spatial agreement between the two maps.

Some of the error could be attributed to:

- The creation of the validation image

The validation raster is not an absolute reference since it is based on a sketch of the visible polygonal network. Some polygons might not have been visible while other polygons may have been destroyed at the surface during excavation.

- The unknown depth occurrence of the wedges

The validation image is based on the superficial visibility of the wedges, but the depth occurrence of these wedges is not known. Some wedges quickly become smaller with depth, resulting in an EC_a based misclassification since the EM38DD sensor has an effective exploration depth of 1.6 and 0.75 m in the vertical and horizontal orientation respectively.

In an attempt to improve the prediction, an uncertainty zone surrounding the wedges was introduced in the classification. Therefore a map of the coefficient of variation (CV) of PR was used. The CV of a pixel is defined as the ratio of the standard deviation over the mean (expressed as percentage), calculated for the central pixel of a 3 x 3 moving window. Each pixel of the map is visited in turn. The result is given in Fig. 6.8(a). The highest CVs were found at the polygon boundaries since a change in PR characterizes the boundaries of the wedges. Then, the CV map was optimally classified into three zones using the fuzzy- k -means algorithm. The three classes had mean CVs of 0.9%, 2% and 3.4% and the last was defined as an uncertainty zone. This uncertainty zone was overlaid on the map of the PR classification (Fig. 6.8(b)). The proportion of pixels falling into the uncertainty zone was 18%, reducing the area of the other two zones.

TABLE 6.5 – Confusion matrix obtained after cross-tabulation of the pixels of the validation image (columns) and the PR classification with an uncertainty zone added (rows).

	Host material	Wedge	Total
PR zone 1	930	71	1001
PR zone 2	157	264	421
Uncertainty zone	175	143	318
Total	1262	478	1740

To evaluate the added value of this uncertainty zone a cross tabulation with the validation image was performed. The confusion matrix of this cross tabulation is

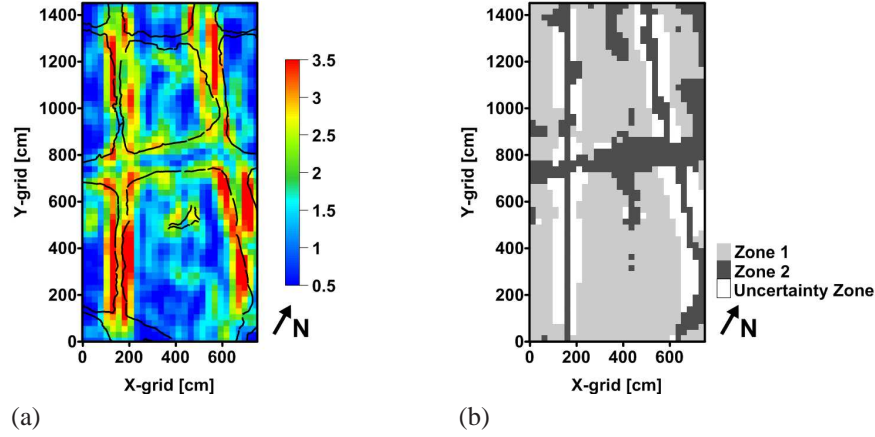


FIGURE 6.8 – (a) Map of the PR coefficient of variation (CV), expressed as percentage [%], overlaid with a vector of the wedges wider than 0.25 m and (b) fuzzy- k -means classification of the PR map with an uncertainty zone added based on the CV.

given in Table 6.5. The overall accuracy (based on the reduced area) increased to 84%, and κ increased to 0.6. So incorporating an uncertainty zone slightly improved the classification.

6.4 Conclusion

Electromagnetic induction was able to detect exposed near-surface frost-wedge pseudomorphs based on the differential soil conditions between wedge fillings and host materials. The dual dipole configuration of the EM38DD sensor was essential since it was found that the profile ratio of EC_a measurements revealed the polygonal pattern. The wedges were represented by local PR minima. Based on a fuzzy- k -classification of the PR map in combination with an uncertainty zone an overall accuracy of 84% was achieved.

It should be noted however that (i) only wedges wider than 0.25 m were identified with an high resolution EC_a survey and (ii) the methodology was tested over a small study area where the frost-wedge pseudomorphs were visible. A major advantage of the EMI technique is its suitability to survey large areas in a mobile way. However, future research has to focus on areas where the polygonal pattern is not visible at the surface. Nevertheless, this study clearly proved that the EM38DD is a promising soil sensor for locating frost-wedge pseudomorphs.

7

Locating clay in a sandy forest soil using the EM38DD soil sensor

The content of this chapter was published as:

Cockx, L., Van Meirvenne, M., and B. De Vos. (2007). Using the EM38DD soil sensor to delineate clay lenses in a sandy forest soil. *Soil Science Society of America Journal*, 71: 1314-1322.

7.1 Introduction

The spatial variability of forest soils is an important factor for forest site planning, quality, and productivity (Schoenholtz et al., 2000). Forest soils serve multiple production and environmental functions and maintaining these functions is crucial for sustainable forest management.

Accurate characterization of the soil is a preliminary step in site establishment, since both lateral and vertical soil heterogeneity may have an impact on tree growth patterns. Soil texture is a fundamental qualitative soil physical property, influencing many other properties and processes including soil formation, water movement, erosion potential and cation exchange capacity (CEC). It strongly affects soil moisture and controls the pool of nutrients available for plant uptake (McBride et al., 1990). Soil discontinuities may be defined by significant changes in soil texture and their identification and spatial delineation is important in land use decision making (Ogg et al., 2000). The presence of a clay layer in a sandy soil constitutes a discontinuity that restricts infiltration and influence the lateral movement of soil water and agrochemicals (Doolittle et al., 1994). Concerning forest productivity, Woolery et al. (2002) selected percentage clay in the B horizon as an important parameter for estimating species productivity. Also Bravo and Montero (2001) found texture to be an important factor for forest productivity. Moreover, soil texture is known to determine the impact of soil disturbances on tree growth (Gomez et al., 2002). However, it should be mentioned that forest productivity relies on the interplay of soil physical, chemical and biological properties and processes which can be complex and varying among forest ecosystems (Schoenholtz et al., 2000). Consequently, forest management practices should be site specific and account for the spatial variability of soil properties. Therefore, knowledge of the spatial variability of soil texture may be important for interpreting tree productivity and planning forest management strategies.

New techniques have evolved allowing soil spatial variability to be identified using non-invasive, geophysical soil sensors. Sensors based on EMI have already proved their utility for soil characterization in agricultural soils (Corwin and Lesch, 2005a). In this sensing technique the soil EC_a is measured, which is used as an indirect indicator of some soil properties (McNeill, 1980b). In non-saline soils the variation in EC_a is primarily determined by soil texture, moisture content and CEC, all of which are important to plant biomass production. The inverse of EC_a , the electrical resistivity, can be also considered as a surrogate for the variability of soil physical properties (Samouelian et al., 2005) but its depth of investigation depends on the distance between the electrodes, which is limited by the inter-tree distance in densely populated forests. On the other hand, GPR has proved to be an effective tool for exploring subsurface horizons, also in forests (Butnor et al., 2003). Kung and Donohue (1991) showed that GPR was able to locate soil lay-

ers with textural discontinuities whereas Boll et al. (1996) predicted the depth to textural interfaces using GPR. Both techniques detect changes in electromagnetic soil properties. However, the use of GPR is limited since only electrically resistive soils are amenable to study (Butnor et al., 2003). As a response to this disadvantage, EMI has been used as a precursory tool to guide the more costly, complex and time-consuming GPR measurements (Gish et al., 2002; Inman et al., 2002). Generally, it is recognized that EMI allows the detection of gradual lateral changes in textural properties (Doolittle and Collins, 1998), whereas GPR is better suited to vertical exploration of textural discontinuities. Nevertheless, several authors (Brus et al., 1992; Bork et al., 1998; Mueller et al., 2003) have reported the use of EMI to find textural discontinuities within the soil profile.

This study performed both lateral and vertical soil explorations using an EMI sensor. The sandy soil of the study area is characterized by the presence of a clay substratum at variable depth. We aimed to locate this textural discontinuity with the EM38DD sensor. This sensor has the advantage of measuring the EC_a simultaneously in two orientations. The ratio of these two orientations provides an indirect measure of the degree of soil profile heterogeneity. Additionally, we investigated different data analyzing techniques to delineate the spatial extent of the clay substratum.

7.2 Material and Methods

7.2.1 Study site

The study site is an even aged, 75-year-old Scots pine plantation of 2 ha in Brasschaat, Belgium (central coordinates: 51°18'33"N, 4°32'14"E). The plantation is part of a 150 ha mixed coniferous/deciduous state forest, called 'De Inslag', located in the Belgian Campine region (Fig. 7.1). Since 1988 the Flemish Research Institute for Nature and Forest has used this site as a research area and in 1992 it was integrated as a Level II Research Site into a European research program for monitoring forest ecosystems. Several studies concerning tree physiology, nutrient behaviour, CO₂ and water cycles, forest vitality and air pollution monitoring were conducted or are ongoing at this site. The topography of the area is flat with a maximum height variation of 33 cm. The soil is called a 'post-podzol' in the Belgian soil classification and symbolized as 'Zegb' on the Belgian soil map: it is a wet (drainage class 'e') sandy (textural class 'Z') soil, characterized by an anthropologically disturbed spodic horizon (profile development 'g'), with the final 'b' indicating a moderately sandy variant of the parent material. The soil consists of aeolian coarse sand, but irregularly located and at variable depths (usually between 1.5 to 2.5 m or more) clay with a thickness of at least 0.20 m can be found (> 40 % of clay) (Baeyens et al., 1993). The mineral soil is characterized by very low

pH values (pH-H₂O between 3.6 and 4.1) and a very low CEC; the forest floor is of the Mor type and varies in thickness across the site (between 1 and 13 cm). At some locations a thick moss layer (3–7 cm) covering the litter layer is present. In the northwestern corner traces of Second World War activities were found: bomb craters disturbed the soil and relict barbed wire was found at the site.

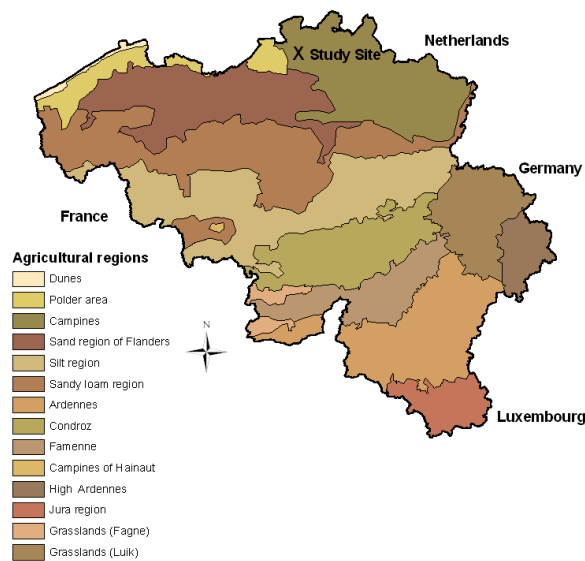


FIGURE 7.1 – Location of the study site in the Campine region of Flanders (Belgium).

7.2.2 EC_a survey

We conducted a soil sensing survey using the EM38DD soil sensor. With the EM38DD the soil EC_a is measured simultaneously in two orientations each having its own depth response profile. The sensitivity of the sensor can reach a depth of 2 m in low-conductive soils. The vertical orientation (EC_a-V) receives its major influence from deeper soil layers, while the horizontal orientation (EC_a-H) receives a dominant influence of the near surface soil. Combining the EC_a measured in the two orientations in a so-called profile ratio (PR) provides an indication of the heterogeneity of the soil profile (Corwin et al., 2003): $PR = EC_{a-H}/EC_{a-V}$. A PR close to one indicates a uniform profile, a $PR < 1$ indicates a more conductive subsoil compared to the topsoil and a $PR > 1$ indicates decreasing conductivity with depth. In this study site the PR is expected to be very informative.

In total, 156 EC_a measurements were taken by placing the EM38DD manually on the soil surface. Fifty-four locations were selected on the basis of a pre-defined grid with a 20 m by 20 m spacing; the other locations were taken in between

these grid nodes to complement the regular sampling design with shorter distance measurements.

7.2.3 Soil sampling

To provide an interpretation of the EC_a measurements, soil samples at 23 locations were taken over 50 cm intervals down to 2 m. Texture was analyzed using the conventional pipette method. Additionally, in the area where we expected to find clay, the presence of heavy clay within a depth of 2 m was checked by hand auguring at 60 locations. Where heavy clay was encountered, the depth to its top was registered. These presence/absence observations were used as validation data.

7.2.4 Interpolation techniques

Ordinary point kriging (OK) was used to interpolate the EC_a measurements (see section 3.1.3). Based on the PR values we aimed to generate the probability of the clay occurrence. Indicator kriging (IK) was used for two purposes: (i) to obtain a map of the probability of the clay lens to occur in terms of a critical PR value and (ii) to interpolate the binary presence/absence data to a validation map showing the presence of heavy clay within 2 m depth.

The principles of IK have been discussed in detail by Goovaerts (1997); Deutsch and Journel (1998), and others. IK obtains the probability of a certain critical threshold z_c being exceeded by building the conditional cumulative distribution function (ccdf) at each point based on the behaviour and correlation structure of indicator transformed data. In this case our interest was in the probability not to exceed a certain threshold: the clay layer is assumed to occur at locations with a PR value smaller than a critical PR threshold.

The ccdf $F[\mathbf{x}_0; z|(n)]$ signifies a probabilistic model for the uncertainty around the unknown value at \mathbf{x}_0 , where $|(n)$ represents the conditioning to local information:

$$F[\mathbf{x}_0; z|(n)] = \text{Prob}[Z(\mathbf{x}_0) \leq z|(n)] \quad (7.1)$$

Therefore, original data have to be transformed into indicators $i(\mathbf{x}_\alpha; z_k)$ in respect to a series of K threshold values z_k , selected across the range of data:

$$i(\mathbf{x}_\alpha; z_k) = \begin{cases} 1 & \text{if } z(\mathbf{x}_\alpha) \leq z_k \\ 0 & \text{if otherwise} \end{cases} \quad k = 1, 2, \dots, K \quad (7.2)$$

Generally, the quantiles of the global cdf are taken as a series of K threshold values, covering the range of variation of z (Van Meirvenne and Goovaerts, 2001).

For each threshold z_k an indicator variogram was calculated and modeled:

$$\hat{\gamma}_1(\mathbf{h}; z_k) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [i(\mathbf{x}_\alpha + \mathbf{h}; z_k) - z(\mathbf{x}_\alpha; z_k)]^2 \quad (7.3)$$

At every point \mathbf{x}_0 , the ordinary IK estimates of the indicators are used to approximate the ccdf. For threshold z_k this becomes:

$$F(\mathbf{x}_0; z|(n))^* = [i(\mathbf{x}_0; z_k)]^* = \sum_{\alpha=1}^{n(\mathbf{x}_0)} \lambda_\alpha(z_k) \cdot i(\mathbf{x}_\alpha; z_k) \quad (7.4)$$

The weights $\lambda_\alpha(z_k)$ are obtained by solving the ordinary IK system for each threshold z_k :

$$\begin{cases} \sum_{\beta=1}^{n(\mathbf{x}_0)} \lambda_\beta(z_k) \gamma_1(\mathbf{x}_\alpha - \mathbf{x}_\beta; z_k) + \varphi(z_k) = \gamma_1(\mathbf{x}_\alpha - \mathbf{x}_0; z_k) & \alpha = 1, \dots, n(\mathbf{x}_0) \\ \sum_{\beta=1}^{n(\mathbf{x}_0)} \lambda_\beta(z_k) = 1 \end{cases} \quad (7.5)$$

To construct the local ccdf, one has to deal with: (i) order-relation problems arising from the independent kriging of the indicator values and (ii) the interpolation and/or extrapolation between and outside the threshold values. The former is corrected by the upward/downward procedure as suggested by Goovaerts (1997). The latter interpolation was done using linear interpolation between the threshold values, for the lower tail extrapolation a power model was used whereas for the upper tail extrapolation an hyperbolic model was used. After the ccdf is built, the probability of not exceeding a critical threshold z_c can be assessed at every location \mathbf{x}_0 :

$$\text{Prob}[Z(\mathbf{x}_0) \leq z_c | (n)] = F[\mathbf{x}_0; z_c | (n)] \quad (7.6)$$

7.2.5 Classification and delineation methods

Two numerical methods were used for the delineation of the clay substratum from EC_a: (i) a fuzzy- k -means classification and (ii) a classification of the IK probability map. Using a fuzzy- k -means classification the PR values were classified into an optimum number of clusters (see section 3.2.1), from which one of the resulting clusters was retained as optimal, indicating the presence of the clay substratum. The second method is based on a map indicating the probability of the clay substratum to occur. This probability map was obtained through IK; the centroid of the optimal cluster from the fuzzy- k -means classification was taken as a critical threshold z_c . The resulting probability map was classified into a Boolean map showing predicted absence or presence of the clay substratum.

7.2.6 Accuracy measurements

The results of the delineation methods were compared to a validation image showing the experimentally observed presence of the clay substratum.

Categorical comparisons are generally based on a confusion matrix containing categorical similarities obtained from a pixel-by-pixel comparison. Table 7.1 shows a two by two confusion matrix in which the elements are the number of pixels that fall into each categorical combination.

TABLE 7.1 – A two by two confusion matrix.

		Observed		
Predicted	Presence	a = true positives	b = false positives	a+b
	Absences	c = false negatives	d = true negatives	c+d
	Total	a+c	b+d	n= a+b+c+d

Based on the confusion matrix a list of pixel-based accuracy measures was calculated (see section 3.2.2). A simple measure of accuracy is the proportion observed agreement (P_o) defined as the number of correctly classified pixels divided by the total number of pixels. The most popular accuracy measure however is the kappa statistic (κ) which eliminates classification agreement by chance (P_c) (Cohen, 1960). The main objection to these two statistics is their dependence on prevalence or the frequency of presences ($a+c/n$) (Fielding and Bell, 1997). A prevalence effect exists when there is an uneven distribution among observed presences and absences. κ^* as developed by Foody (1992) has also been described as an adjustment to κ for a prevalence effect since a skewed distribution of categories increases agreement by chance (P_c) (Byrt et al., 1993). Also κ_{histo} and κ_{loc} were calculated, but for delineation studies κ_{loc} was considered to be more relevant than κ_{histo} . More detail about these variants of kappa can be found in section 3.2.2. As discussed earlier in section 5.2.4 these measures however do not account for spatial relations. To account for similarities in spatial patterns, the average fuzzy similarity (S_a) as calculated by Eq. (5.6) was assessed by introducing fuzziness of location.

The second delineation method yielded in first instance a map showing the probability of the clay substratum to occur. To validate a probability map a Receiver Operating Characteristic (ROC) curve can be constructed (Pontius and Schneider, 2001). The ROC curve was first used in World War II to correctly discriminate between an object (passing aircrafts) or noise based on radar signals. Since then, ROC curves have been used in many other applications of signal detection, in particular for medical decision-making. Over the last few years, they have found increased interest in the machine learning and data mining communities for model evaluation and selection (Bradley, 1997; Perlich et al., 2004). A ROC curve re-

lates sensitivity to (1-specificity) over a continuous range of probability thresholds. Sensitivity is the proportion of observed presences that are predicted as such while specificity is the proportion of observed absences that are predicted as absences (Table 7.2). Sensitivity is also referred to as true positive rate and specificity as true negative rate. The complement of specificity, (1 - specificity), is then referred to as false positive rate. The occurrence of the clay substratum is predicted when its probability exceeds the probability threshold. The ROC curve evaluates the predicted probabilities by comparing them with the observed values over the whole range of predicted probabilities instead of only evaluating the percentage of correctly classified observations at one fixed threshold value. The clay substratum is predicted to occur if the predicted probability exceeds the probability threshold. For each possible probability threshold, sensitivity and specificity are calculated and as such on the ROC curve is constructed.

The area under the ROC curve (AUC) is often used as a threshold-independent measure for the ability of the method to correctly discriminate between absence and presence of an event of interest - in our case the clay substratum. Rather than specifying a threshold for converting predicted probabilities into either presences or absences, AUC provides a measure of how well the model discriminates the presence and absence across all possible thresholds. The AUC is a measure of the likelihood that a presence will have a higher predicted value from the method than an absence. An AUC value of 0.5 (a diagonal line on the ROC curve) would indicate that the method has no discriminating ability (i.e. it performs no better than chance), a value of 1 would indicate that a method always correctly assigns presences a higher probability than absences. Manel et al. (2001) showed that the AUC statistic is prevalence independent.

TABLE 7.2 – Indices related to the ROC curve and the F -measure. Their formula based on a two by two confusion matrix is given (in conformity with Table 7.1).

Index	Formula
Sensitivity = Recall (R) = True Positive Rate	$a/(a+c)$
Specificity = True Negative Rate	$d/(b+d)$
1- Specificity = False Positive Rate	$b/(b+d)$
Precision (P)	$a/(a+b)$

Besides for accuracy purposes, information from the contingency matrix was also used to determine the optimal probability threshold through the F -measure. The F -measure is defined as the harmonic mean of precision (P) and recall (R) (Table 7.2). Precision is the proportion of predicted presences that were observed to be present and is a measure of accuracy. Recall (also called sensitivity) on the other hand is the proportion of observed presences that were predicted to be present and measures the ability of the method to correctly predict all observed presences.

The trade-off between precision and recall is illustrated in Fig. 7.2. In case the clay substratum is predicted to be present in the whole study site, recall is perfect whereas precision is low. In case all the predicted clay is also observed, but not all observed clay is predicted, precision is perfect whereas recall is low. Using the harmonic mean of precision and recall implies that both need to be high to obtain a high F -measure.

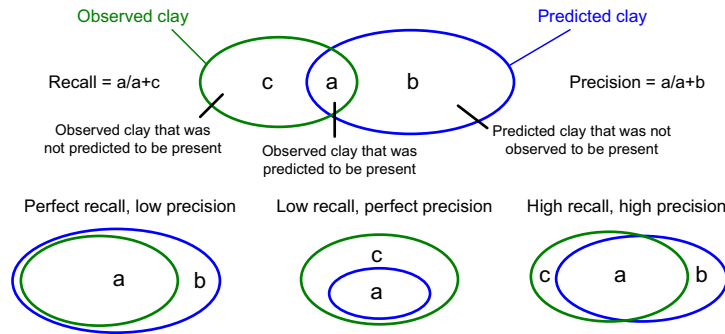


FIGURE 7.2 – Illustration of precision and recall with their trade-off.

A weighted version of the F -measure as developed by van Rijsbergen (1979) was used:

$$F_{\beta} = \frac{(\beta^2 + 1)PR}{\beta^2 P + R} \quad (7.7)$$

with $\beta \in [0, +\infty]$ as weighting factor controlling the relative importance of precision versus recall. With $\beta = 1$, precision and recall have equal weights: a smaller β emphasizes precision, a larger β emphasizes recall. Since the aim was to precisely predict the presence of the clay substratum, most weight was attached to precision and β was set to 0.5. To select the optimal probability threshold $F_{0.5}$ was maximized.

7.3 Results and Discussion

7.3.1 EC_a measurements

The locations of the 156 EC_a measurements are shown in Fig. 7.3a. The descriptive statistics of the EM38DD measurements (Table 7.3) show that the EC_a values of this sandy soil were very low (3–9 mS m⁻¹) with relatively small coefficients of variation (CVs of 22% and 29% for EC_a -H and EC_a -V, respectively). The EC_a -H and EC_a -V measurements had similar mean values within the same data range. Their distribution was symmetrical but platykurtic, whereas the PR measurements

showed a more skewed, leptokurtic distribution. In most studies the correlation between EC_a-H and EC_a-V is reported to be very strong (e.g. Triantafilis and Lesch (2005); Vitharana et al. (2008)), but in this study the correlation coefficient was only 0.74, indicating some deviation from a linear relationship.

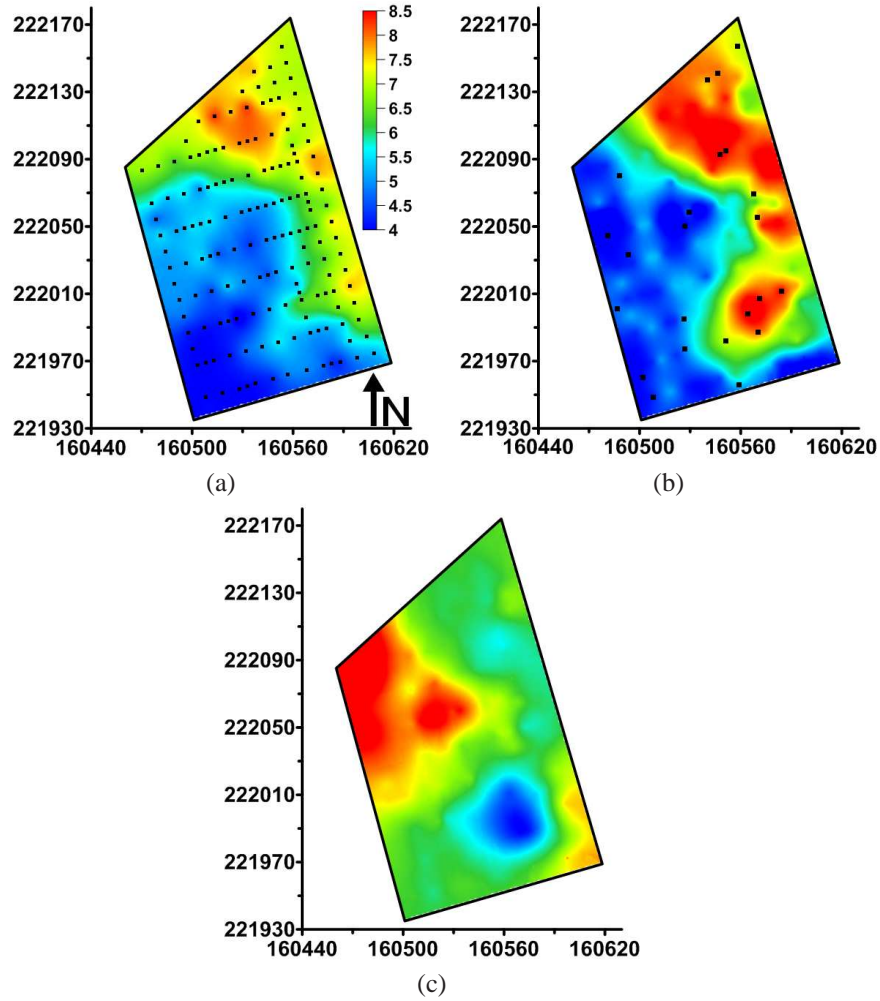


FIGURE 7.3 – Point ordinary kriging maps of (a) EC_a-H with indication of the EM38DD measurement locations, (b) EC_a-V with indication of the texture samples and (c) PR. The metric Lambert coordinates [m] are used on the X and Y axis.

Spherical variogram models with a nugget effect were found to represent the experimental variograms best. The variograms of EC_a-H and EC_a-V had a similar spatial structure with a strong spatial relation as indicated by their relative nugget

TABLE 7.3 – Descriptive statistics of EC_a-H, EC_a-V and PR .

	Summary statistics (<i>n</i> = 156)					
	Min	Max	Mean	CV [%]	Skewness	Kurtosis
EC _a -H	3	9	5.90	22	0.075	-0.776
EC _a -V	3	9	6.06	29	0.206	-1.287
PR	0.60	1.75	1.01	21	0.933	1.378

effect (RNE) of 10% and 6% respectively. The PR measurements on the other hand had a larger RNE (30%). The RNE measures the proportion of random or short-distance variability. The range of the variograms was in the order of 80 m for all three variables. Although the data range was rather small, maps of EC_a-H and EC_a-V, obtained by point OK with a pixel resolution of 1 by 1 m, showed clear and rather similar patterns (Fig. 7.3(a) and (b)). The highest EC_a values occurred in the east of the field, the lowest in the west. In contrast, the PR map indicates that most of the study site is characterized by a rather homogeneous soil (PR values around 1). However, in the northwest corner increased PR values were found (1.2–1.3) whereas in the southeast a circular phenomenon with decreased PR values (0.7–0.8) can be seen (Fig. 7.3(c)).

7.3.2 Clay content and EC_a variability

Based on the clay content of the 23 sampled locations (Fig. 7.3(b)), two types of textural profiles could be distinguished: one homogeneous with a low clay content down to 2 m (18 locations), the other one with a marked increase in clay content from a depth varying between 1–2 m onwards (5 locations) (Fig. 7.4). The standard deviation of the clay content of the homogeneous soil profile was quite uniform for all layers, on average 2.6%. The top 1 m of the heterogeneous soil profile had a similar standard deviation, but it increased strongly below 1 m to 6.5% on average. This increase in standard deviation resulted from the variable depth where the clay substratum occurred. The heterogeneous soil profiles were all characterized by an increased EC_a-V resulting in a decreased PR. The average PR of the heterogeneous profiles was 0.81, compared with an average PR value of 1.03 for the homogeneous soil profiles.

Fig. 7.3(c) shows that in the northwest corner the EC_a-H appeared to be higher than the EC_a-V although no substantial textural differences were found in the samples of that area. We assumed that the EC_a at those places was influenced by anthropogenic disturbances: in the northwest corner traces of the Second World War were found. Barbed wire acting as a metal artefact in the soil is known to disturb the electromagnetic signal strongly. Also the presence of bomb craters with associated differences in topography, moisture conditions and litter density is believed

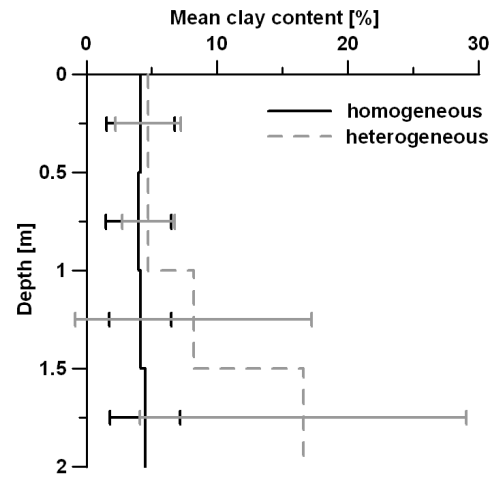


FIGURE 7.4 – Mean clay content of the heterogeneous and homogeneous soil profiles, shown in intervals of 0.50 m (errors bars represent the standard deviation).

TABLE 7.4 – Pearson correlation coefficients for mean clay content ($n = 23$) of different soil layers and EC_a -H and EC_a -V.

Depth of the mean clay content [m]	Pearson correlation coefficient	
	EC_a -H	EC_a -V
0-0.5	0.32	0.30
0.5-1	0.37	0.28
1-1.5	0.34	0.38
1.5-2	0.41	0.67**
0-1.5	0.44*	0.44*
0-2	0.40	0.56*

* Significant at the 0.05 level

** Significant at the 0.01 level

to influence the signal of the sensor. In the southeast the decreased PR was clearly caused by a textural discontinuity. All samples located in this circular pattern of small PR values had a substantial increase in clay content (between 13% and 27%) at a depth between 1.5–2 m. Table 7.4 shows the Pearson correlation coefficients (r) between the mean clay content in the different soil layers (0.5 m intervals) and EC_a -H, EC_a -V. Up to 1.5 m depth the correlations for both EC_a -H and EC_a -V were rather weak ($r < 0.50$) because of the textural homogeneity. In the 1.5–2 m layer the correlation increased. In particular, EC_a -V was well related with the more variable clay content in that layer ($r = 0.67$), while the correlation with EC_a -H was considerably less ($r = 0.41$). Looking at the correlation over the whole soil profile, averaged over the top 1.5 m and 2 m, the clay content in the 1.5–2 m layer seemed

to increase the correlation, especially for EC_a -V (r increased from 0.44 to 0.56).

7.3.3 Delineation of the clay lens

The presence of a clay substratum in the sandy soil substitutes a clear textural heterogeneity resulting in an increased EC_a -V compared to EC_a -H. Therefore PR minima were hypothesized to indicate the presence of this clay substratum. A fuzzy- k -means classification of the interpolated PR values was performed with a fuzzy exponent (ϕ) of 1.3. This value for ϕ is in the middle of the range 1.12–1.5 as suggested for soil data by Odeh et al. (1990). Minimizing the NCE (eq. (3.27)) and the FPI (eq.(3.25)) resulted in the least fuzzy and least disorganized number of classes (Fig. 7.5). The study site was optimally classified into five classes: two classes with a PR centroid around one, two classes with a PR centroid larger than one, and one class with a PR centroid smaller than one (Fig. 7.6(a)).

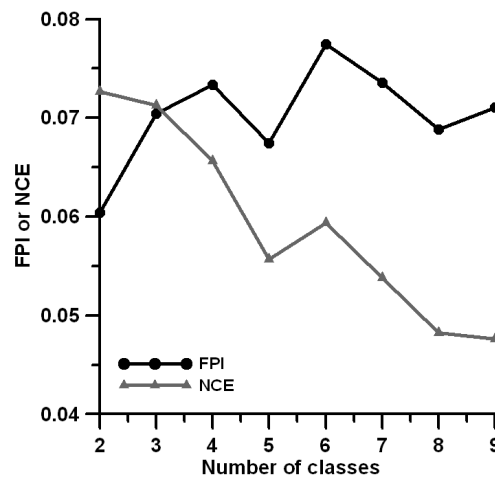


FIGURE 7.5 – Fuzziness performance index (FPI) and normalized classification entropy (NCE) as a function of the number of classes.

We were interested in the class with the lowest PR centroid (0.81), indicating an increasing conductivity with depth, suggesting the presence of a clay substratum. In a first, simple approach, a Boolean image was used to predict the presence of the clay substratum (Fig. 7.6(b)): if the centroid PR equals 0.81, the clay substratum was predicted to be present. This Boolean image resulted in two circular areas, of which the small did not seem to represent a deep clay substratum since both EC_a -H and EC_a -V had increased mean values in this area. It indicates a larger clay content in the entire profile, but with a slight increase with depth.

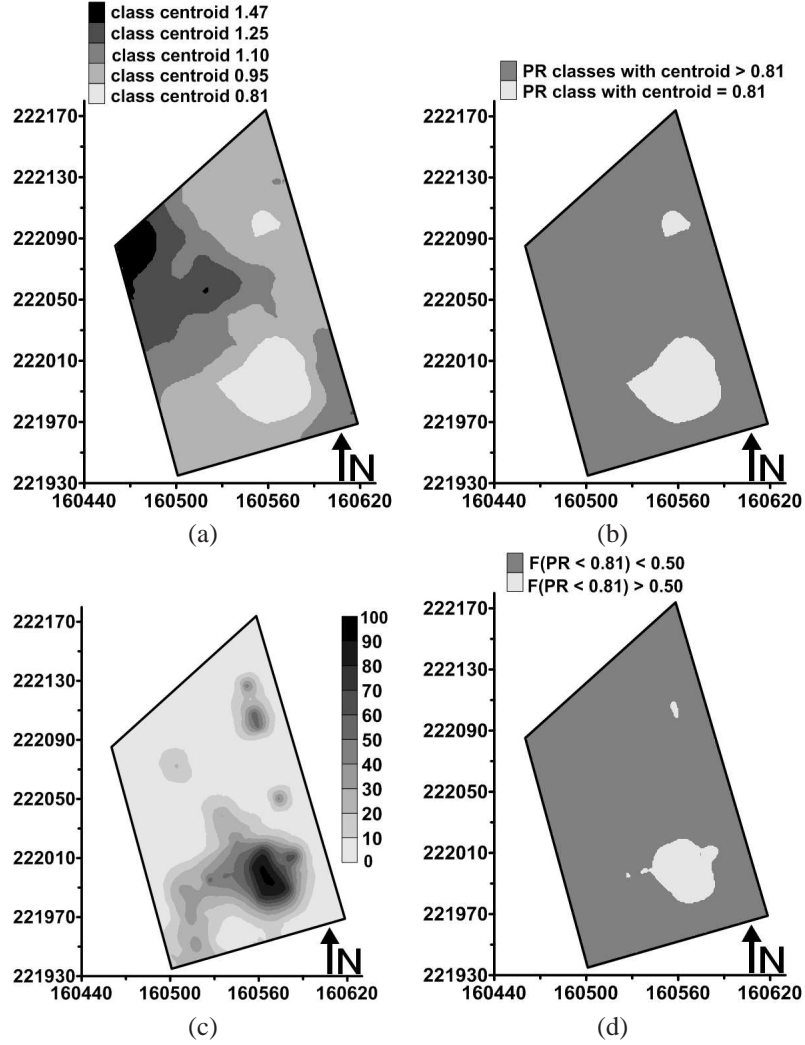


FIGURE 7.6 – Map of (a) the fuzzy classification of the profile ratio into five classes, (b) the predicted presence of the clay substratum according to the fuzzy-k-means reclassification, (c) the IK probability showing the probability (expressed as percentage) that the profile ratio ≤ 0.81 and (d) the predicted presence of the clay substratum according to the classification of the IK probability map with a probability threshold of 0.5.

The second method used this centroid value of 0.81 as the threshold value z_c in IK. This method introduced uncertainty in terms of the probability that the clay substratum is present. Based on the global cdf of the PR observations, seven thresholds (z_k) were selected for the indicator coding of PR. These thresholds correspond to the deciles of the PR distribution, since the 0.4, 0.5, 0.6 and 0.7 deciles

had the same quantile with a value of 1. Because a small PR indicates a textural heterogeneity, the clay substratum was assumed to occur at locations with a PR smaller than 0.81. IK provided the probability that $PR \leq 0.81$, which was taken as a measure of prediction that the clay substratum occurs (Fig. 7.6(c)).

7.3.4 Validation

All auger observations (those performed for the textural analyses and the binary observations used for the validation) were binary coded: 1 if clay occurred within a depth of 2 m, 0 if not. These values were interpolated using IK with a spherical variogram model having a RNE of 11% and a range of 90 m, resulting in an indicator map of the presence of the clay substratum (Fig. 7.7(a)). This indicator map was categorized by a threshold of 0.5: whenever the estimated indicator was at least 0.5, the clay substratum was expected to occur. The resulting map acted as the validation image (Fig. 7.7(b)). Fig. 7.7(b) also shows the sampled locations as black dots when clay was observed within 2 m and as empty dots if no clay was encountered.

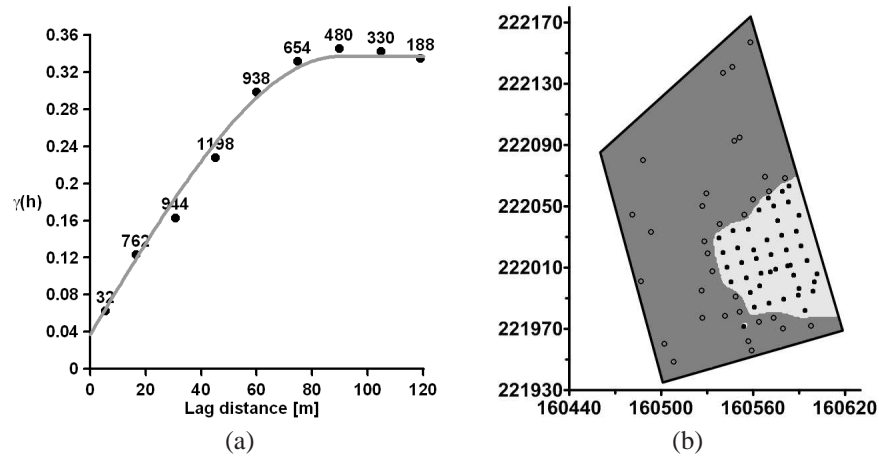


FIGURE 7.7 – (a) Indicator experimental (dots) and modeled variogram of the validation data and (b) the validation image with the validation samples (black dots are presences, while empty dots are absences).

Using this validation image the accuracy of the delineation methods was evaluated. First, the probability map of the second method was classified into a Boolean image based on an optimal probability value determined by the $F_{0.5}$ measure. The $F_{0.5}$ measure was maximized at a probability value of 0.50 (Fig. 7.8(a)). All pixels with a probability $\geq 50\%$ were classified into a class where the clay substratum was predicted to occur (Fig. 7.6(d)).

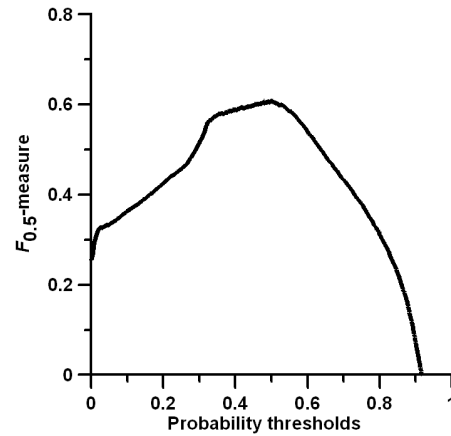


FIGURE 7.8 – The F -measure of the harmonic mean of precision and recall weighted ($F_{0.5}$) in function of the probability thresholds.

The similarity between the presence/absence pattern of the two Boolean classifications and the validation image can be seen clearly (Fig. 7.6 versus Fig. 7.7). Nevertheless there was a substantial difference in terms of the size of the clay lens: the area of the clay substratum was 0.25 ha and 0.14 ha for the two Boolean classifications respectively, whereas the validation image showed a clay substratum with an area of 0.47 ha. So the location of the clay substratum was correct, but the spatial extent of the clay substratum was underestimated by both methods. The main error occurs at the eastern boundary of the clay substratum. This side is near to the metal fence which delimits the study site. This might have influenced the electromagnetic field induced by the EM38DD and the resulting EC_a measurements.

To quantify the accuracy of the two methods, the Boolean maps were compared with the validation image on a pixel by pixel basis. Table 7.5 shows the resulting accuracy measures.

TABLE 7.5 – Accuracy measures of the two classification methods

Accuracy measure	Fuzzy- k -means	Classification of the IK probability map
P_o	0.817	0.838
κ	0.341	0.349
κ^*	0.634	0.676
κ_{loc}	0.529	0.864
S_a	0.819	0.841

The overall accuracy (P_o) reached high values for both methods (0.819–0.838), but it should be remarked that there was a prevalence effect of 21%; the proportion of pixels without a clay present in the subsoil (< 2 m) is larger than with presence

of the clay substratum. Predicting the absence of the clay substratum in the whole study site would still give an P_o of 0.79. κ values were 0.341 and 0.349 for the two methods, respectively, indicating a fair agreement (Landis and Koch, 1977). After correction for the prevalence effect a substantial strength of agreement is reached, indicated by κ^* values of 0.634 and 0.676. In terms of κ_{loc} the IK method scored best with a κ_{loc} of 0.86. The spatial context was taken into account with S_a . Fuzziness of category was not considered and fuzziness of location was set with an exponential decay function (with a halving distance of two cells and a neighbourhood defined by a radius of two cells). This measure was for both methods higher (0.819 and 0.841) than the indices calculated on a pixel basis. So we can conclude that the delineation was good in terms of spatial agreement. The IK based method was found to be the most accurate, but it is also a more elaborate method based on a PR threshold value obtained from the fuzzy- k -means method. Nevertheless, the IK based method was preferred to delineate the clay lens.

The accuracy of the IK probability method was also confirmed by its ROC curve (Fig. 7.9). The AUC value of this curve was 0.770, indicating a 77% likelihood that a presence of the clay substratum - randomly chosen from the map - has a higher predicted probability than a randomly chosen absence. Following Swets (1988), an AUC value between 0.7 and 0.9 indicates reasonable discriminating ability. Thus processing the EM38DD measurements allows for the delineation of a clay substratum in a sandy soil with an acceptable level of accuracy.

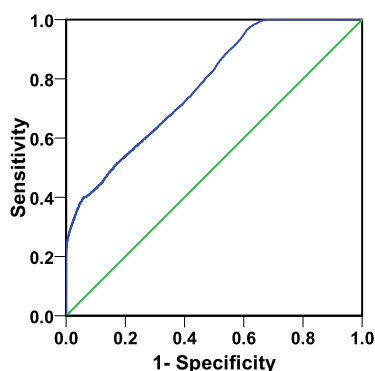


FIGURE 7.9 – The ROC curve of the IK probability method (AUC = 0.770) is represented by the blue line. The green line indicates a model with a classification performance no better than chance (AUC = 0.5).

7.3.5 Depth of the clay lens

The depth to the upper boundary of the clay layer (D_c) was reported to be an important factor for biomass development in pine stands (Usoltsev and Vanclay,

1995). In our study site a significant relationship between D_c (measured at 42

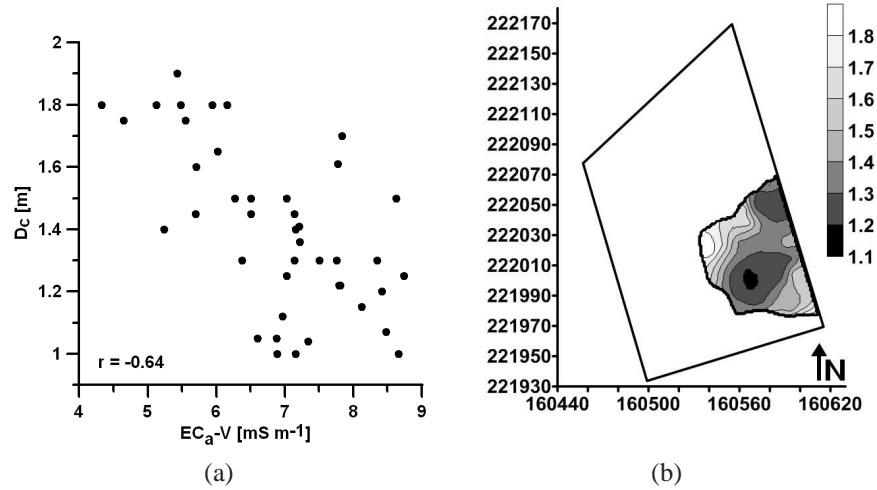


FIGURE 7.10 – (a) Scatterplot of EC_a-V vs the depth of the clay substratum (D_c) and (b) the depth of the textural discontinuity (clay substratum in meters below the surface).

locations) and EC_a-V was found, with a Pearson correlation coefficient of -0.64 (Fig. 7.10(a)). Following Doolittle et al. (1994), we fitted an exponential regression model which appeared to be the best model to predict D_c ($R^2 = 0.43$):

$$D_c = 0.986 + e^{1.438 - 0.347EC_a-V}. \quad (7.8)$$

The closer the textural discontinuity was to the surface, the more it contributed to the response of the sensor. Fig. 7.10(b) shows the predicted D_c . The central part of the clay lens occurred closest to the surface, with a minimum depth of 1.1 m.

7.4 Conclusions

Analyzing the EM38DD data resulted in an accurate identification of the location of the clay lens, but tended to underestimate the spatial extent of the clay substratum. The most appropriate processing technique appeared to be indicator kriging through the use of a probability map showing the probability of the clay substratum to occur. The fuzzy- k -means algorithm was used as an initial step to determine the optimal (and objective) PR threshold. In delineation studies the spatial context of similarities should be taken into account, both κ_{loc} and S_a were found to give good results.

The EM38DD sensor appeared to be a suitable instrument for detecting lateral and vertical soil textural variability important for forest management applications.

To determine a pedological discontinuity both orientations of the EM38DD sensor were essential because only the PR value was useful for analyzing the vertical heterogeneity of the soil profile. We concluded that the dual dipole version of the sensor allows extra opportunities for EMI applications. One field survey with the EM38DD is sufficient to characterize the spatial textural variability, to locate the presence of a clay substratum and to map its depth approximately.

8

Analyzing high resolution soil sensor data to optimize texture predictions

The content of this chapter was published as:

Cockx, L., Van Meirvenne, M., Vitharana, U.W.A., Verbeke, L.P.C., Simpson, D., Saey T. and F.M.B. Van Coillie. (2009). Extracting topsoil information from EM38DD sensor data using a neural network approach. *Soil Science Society of America Journal*, 73: 1-8.

8.1 Introduction

Detailed information on the variation of topsoil texture is important for optimizing soil and crop management due to its influence on soil fertility, water-holding capacity, and crop germination (Chen et al., 2004). A common approach to acquire fine scale soil textural information is based on multivariate (geo)statistical techniques given a significant relation between the target soil variable and an easy-to-measure ancillary variable (McBratney et al., 2006). This has led to an increasing interest in proximal soil sensors by which high resolution soil information can be collected in a short time (Adamchuk et al., 2004; Vitharana et al., 2006).

Proximal soil sensors based on EMI have been used widely to map within-field soil heterogeneity by measuring the EC_a of a soil (Abdu et al., 2007; Johnson et al., 2001; McCutcheon et al., 2006). Under non-saline conditions the soil properties which influence soil EC_a include the porosity, the amount of colloids, the soil moisture content and the soil temperature (Rhoades et al., 1976). Actually, an EMI sensor translates the depth-weighted average response of all these factors in one measurement of the soil EC_a . However, the response of an EMI sensor is non-linear with depth and depends on the coil orientation and intercoil spacing of the sensor (Wait, 1955). Besides, the degree of heterogeneity of the top- and subsoil properties is also a factor at play (Frogbrook and Oliver, 2007): large variations in one of the EC_a -influencing soil properties can dominate the response, regardless its depth. Therefore, the straightforwardness of extracting information about the top- or subsoil from EMI measurements also depends on the vertical soil variation.

Most studies that predicted the clay content by EMI-sensed EC_a were based on a single sensor orientation (Jung et al., 2006; Weller et al., 2007). Few studies combined EMI sensed data: at a within-field scale taking the ratio of two EM38 orientations increased the information present in each orientation separately (Cockx et al., 2007; Mankin and Karthikeyan, 2002) while at a landscape level the clay content of the upper 7 m was best predicted using a hierarchical spatial regression of a composite of EM34 and EM38 data (Triantafyllis and Lesch, 2005).

The approach in this study was to investigate the potential of artificial neural networks (ANNs) to extract topsoil information from two EC_a signals. Like a human brain system, ANNs are very efficient at processing data from a variety of sources. Intelligent data fusion through ANNs improves the process of information extraction (Dai and Khorram, 1999). Data fusion is defined as the combination of data from complementary sources and as such simultaneously obtained EC_a measurements each with a different depth-response curve can be treated as multi-source data. An advantage of working with ANNs is that they are able to model complex non-linear relationships without any a priori assumptions about the nature of the underlying processes (Haykin, 1998; Sy, 2006). The interactions of top- and subsoil properties influencing the EC_a signal represent such a com-

plex relationship. ANNs already proved their success in soil science, for example, relating soil EC_a to crop yield (Miao et al., 2006), classifying soil texture (Zhai et al., 2006), developing neural-network-based pedotransfer functions for the prediction of soil hydraulic properties (Pachepsky et al., 1996; Schaap et al., 1998). Still, this analysis has been much less explored for soil textural predictions and to our knowledge, no research focuses on the extraction of topsoil information from simultaneous EC_a signals.

In this study EM38DD based EC_a measurements were used in an ANN analysis to extract information about the topsoil clay content for an agricultural field in the Polder area of East-Flanders, Belgium. The aim was to increase the information available from the two sensor orientations separately by fusing the signals in an ANN. Therefore, we examined the sensitivity of the ANN predicted topsoil clay content to the number of sensor orientations (only horizontal or horizontal plus vertical) and the usefulness of incorporating contextual EC_a information in the ANN input.

8.2 Material and Methods

8.2.1 Site Description

The study was performed on a 10.5 ha arable field in Watervliet, Belgium ($51^{\circ}16'17''$ N, $3^{\circ}40'35''$ E). The field is located in the Polder area of East-Flanders (Fig. 8.1). The topography was nearly flat with a maximum height variation of 60 cm and the soil was classified as Aquic Udifluent (Soil Survey Staff, 1999). The dominant Belgian soil series of the field is sEdp, indicating a moderately wet (drainage class 'd') clayey topsoil texture (textural class 'E') with a shallow sandy substrate ('s') and little profile development (profile development 'p').

The top 110 cm are calcareous Holocene deposits, whereas deeper layers are Pleistocene sediments. Van Meirvenne et al. (1990) found in this area a lithological discontinuity at a depth between 40 and 50 cm where the textural composition changes from alluvial loamy material in the topsoil to aeolian sandy loam in the subsoil. A series of Holocene marine transgressions deposited loamy to clayey material on top of the Pleistocene, predominantly sandy, material. The transgressions occurred through a network of creeks which were subsequently filled with clayey material, resulting in a large subsoil textural variation, up to a 2 m depth. This creek pattern was clearly visible on an aerial photograph of the field, taken in the summer of 1989 (Fig. 8.2). At that time, the sugar beet crop exhibited a differential growth due to the large subsoil textural variability. The creek pattern is well represented by the crop marks: a darker green colour occurs in the clayey creeks. The narrow linear features on the photograph represent the former drainage ditch network constructed at the time of land reclamation.

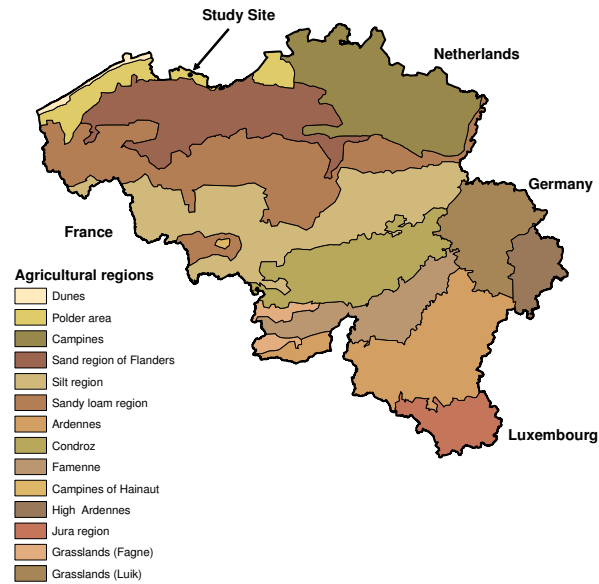


FIGURE 8.1 – Location of the study site in the Polder region of East-Flanders (Belgium).

The polder area is of high agricultural importance due to its fertile soil and from the 11th century on, dikes were constructed to protect the land against new marine incursions. A typical crop rotation in the area is potato, sugar beet, and winter wheat.



FIGURE 8.2 – Oblique aerial image of the arable field, taken in July 1989 (photo: J. Semey, source: Department of Archeology and Ancient History of Europe, Ghent University).

8.2.2 Electromagnetic Induction: Survey and Processing

During the summer of 2006, EC_a data were collected with the EM38DD sensor in a mobile configuration. The sensor was connected to a GPS and a field computer and pulled by an ATV at an average speed of 8 km h^{-1} . Every second a sensor measurement was recorded along lines about 2 m apart, resulting in a measurement resolution of 2 by 2 m. The EM38DD sensor measures simultaneously the soil EC_a in the horizontal ($EC_a\text{-H}$) and vertical ($EC_a\text{-V}$) modes of operation. With a fixed inter-coil spacing of 1 m, the response of the EM38DD is influenced by the coil orientation, each orientation having its own (theoretical) depth-response curve (see Fig. 2.10). Theoretically, the vertical orientation receives 50% of its response from the top 0.85 m of soil, while for the horizontal orientation a similar response is achieved from the top 0.40 m of soil (McNeill, 1980b). The response peaks at the surface for $EC_a\text{-H}$ and between the 0.3 and 0.5 m depth for $EC_a\text{-V}$. In general it is said that $EC_a\text{-H}$ is mainly influenced by topsoil properties, while $EC_a\text{-V}$ gets the highest contribution from the subsoil.

Before interpolating the high resolution EC_a -data, two preprocessing steps were taken. First, due to a lag of 2.5 m between the GPS and the sensor, a shift correction was applied to the recorded measurements. Second, the noise present in the $EC_a\text{-H}$ data was reduced using a Z-score filter: values deviating more than two standard deviations from their local mean (within a radius of 10 m) were replaced by the local mean $EC_a\text{-H}$ value.

Then, continuous EC_a maps with a pixel resolution of 1 by 1 m were obtained using ordinary point kriging, taking into account the spatial autocorrelation among the EC_a data (Goovaerts, 1997). The sensitivity of the measured EC_a to the operating speed induces a measurement error of about 1 mS m^{-1} (Sudduth et al., 2001). As in kriging with uncertain data, this uncertainty is taken into account by replacing the variances in the diagonal of the A matrix of the ordinary kriging system (Eq. 3.11) with the measurement error (Triantafyllis et al., 2001). This had a smoothening effect on the interpolation since no exact interpolations could be made. The EC_a maps were made using the VESPER-software (Minasny et al., 2002).

8.2.3 Soil Data Set

At 78 locations soil samples were taken according to an EC_a -directed sampling scheme to ensure that the existing soil variability was captured. Topsoil (0–40 cm) and subsoil (50–80 cm) samples were analyzed for texture by the conventional pipette-sieve method. At those 78 locations the $EC_a\text{-H}$ and $EC_a\text{-V}$ values were extracted from the interpolated maps to complete the soil data set that was used to predict the topsoil clay content through a series of ANNs.

8.2.4 Artificial Neural Networks: A Brief Introduction

Already in the 1990s, the potential of ANNs to predict difficult-to-measure soil properties was announced (Levine et al., 1996). The ANN is a computational approach, which mimics the capacity to learn through a complex mathematical structure of interconnecting layers, inspired by the biological neural system. A network structure exists of nodes which are organized into input, hidden, and output layers (Haykin, 1998). An example of an ANN structure with one hidden layer is given in Fig. 8.3.

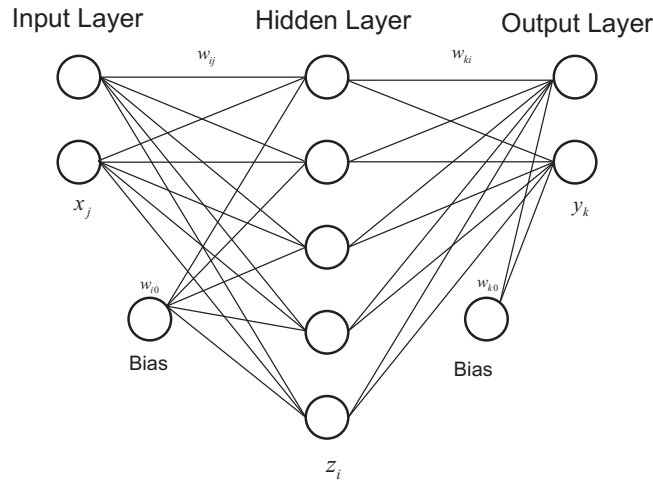


FIGURE 8.3 – Basic network architecture; the input layer (x_j) is connected with the hidden layer (z_i) through the weights w_{ij} , while w_{ki} are the weighted links that connect the hidden layer with the output layer (y_k). w_{i0} , w_{k0} are the bias terms added to the hidden, respectively output layer.

The most used network type is the feed-forward back-propagation network and its algorithm is based on two phases.

In the feed-forward phase, each neuron of the hidden layer ($i = 1, \dots, N_{hid}$) computes an output (z_i) by summarizing the input data ($x_j, j = 1, \dots, N_{in}$), multiplied with their respective weight (w_{ij}) and using this sum as an input in a nonlinear activation function (f_i):

$$z_i = f_i\left(\sum_j w_{ij}x_j + w_{i0}\right) \quad (8.1)$$

A bias term (w_{i0}) was added, serving as a constant added to the weighted sum. Initially, the weights are chosen randomly. Then, the same procedure is repeated in the output layer transforming the output from the hidden layer to the final output ($y_k, k = 1, \dots, N_{out}$), with weights w_{ki} and bias w_{k0} :

$$y_k = f_k\left(\sum_i w_{ki}x_i + w_{k0}\right) \quad (8.2)$$

The feed-forward phase stops after the output y_k has been predicted. A commonly used activation function to introduce non-linearity is the S-shaped hyperbolic tangent function:

$$f(\beta x) = \tanh \beta x = \frac{e^{\beta x} - e^{-\beta x}}{e^{\beta x} + e^{-\beta x}} \quad (8.3)$$

where x is the input of the activation function and β controls the slope of the linear part of the curve.

In the back-propagation phase, the error, defined as the difference between the predictions (y_k) and the actual values is minimized by optimizing the weights of the connections. In an iterative process, called training, the error is back-propagated from the output nodes to the hidden and from the hidden nodes to the input nodes and the weights are altered according to the generalized delta rule (Rumelhart et al., 1986):

$$\Delta w(t) = -\eta \frac{\delta E}{\delta w(t)} + \alpha \Delta w(t-1) \quad (8.4)$$

where $\Delta w(t-1)$ is the previous weight change, $\Delta w(t)$ is the actual weight change, η is the learning rate parameter, E is the error between predictions and actual values and α is the momentum parameter.

The adjustment of the weights should be proportional to the negative derivative of the error with respect to each weight and is set by the learning rate η . The smaller the learning rate, the smaller the changes to the weights will be from one iteration to the following and the smoother the error curve will be, but the slower the rate of training. A large learning rate on the other hand to speed up the the rate of learning might result in an unstable network (oscillating error curve). The momentum parameter α makes the change in weights dependent of past weight changes and improves as such the convergence rate of the network. The momentum parameter may also prevent the learning process to converge to a local minimum (Haykin, 1998).

We used the early stopping-cross-validation method for training, which seeks the minimum error of an independent test set to determine the number of training cycles (Fig. 8.4). The data are split into a training set which is used to train the network and a test set. The training is stopped periodically (after a fixed number of cycles) and the network is tested on the test set after each period of training. Typically the error of the training set is smaller than the error of the test set. The training learning curve decreases monotonically for an increasing number of cycles while the test learning curve decreases monotonically to a minimum and then it starts to increase as the training continues. At that point the training should stop to avoid overfitting: no better generalization will be obtained after this point.

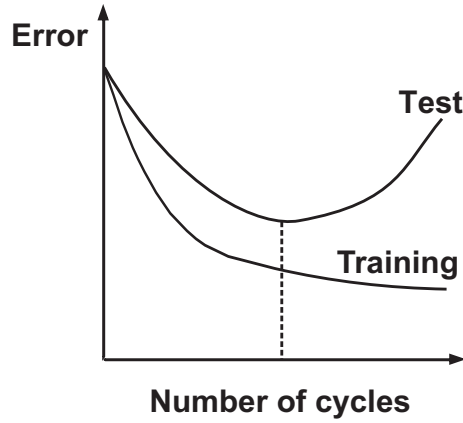


FIGURE 8.4 – Error curve with illustration of the early-stopping cross-validation method. The dotted line indicates when the training should stop.

8.2.5 Datasets, Network Architecture and Data Preparation

Due to its small size, the soil dataset was only split into two sets: a training set of 50 samples (n_{tr}) and a test set with the 28 others (n_{te}). In case of no data limitations an independent validation set is used for error estimation in a validation phase. To compensate for this lack, 10 splits ($n_{sp} = 10$) into training ($D_{tr,i}$, $i = 1, \dots, n_{sp}$) and testing ($D_{te,i}$, $i = 1, \dots, n_{sp}$) were randomly generated. This resampling method could be called a repeated split sample cross-validation. Besides, each of these training sets was run 10 times ($n_{in} = 10$) to account for different random initializations and eliminate the effect of the initial weights. In total 100 ($n_{sp} \cdot n_{in}$) neural network models were trained and the final prediction was obtained by averaging the outcome of all models. These two steps indirectly tested the robustness of the method.

The number of neurons in the hidden layer indicates the complexity of the relationship to be modeled. Too few hidden neurons will generally give high errors due to under-fitting, while too many hidden neurons will result in over-fitting (Bishop, 1997; Van Coillie et al., 2007). To determine the number of hidden neurons Hecht-Nielsen (1987) was followed who stated that $(2N_{in} + 1)$ hidden neurons should be sufficient to model any continuous function. The number of inputs (N_{in}) was altered in two ways:

- Contextual neural networks were used to incorporate spatial EC_a information around each soil sample. Around each soil sample a window was drawn to define the number of EC_a neighbours that were taken into account. The window size (WS) determines N_{in} and we tested a series of windows of x by x pixels, with $x = 1, 3, 5, 7, 9, 11$.

- The effect of using one (EC_a-H) or two (EC_a-H and EC_a-V) sensor orientations was investigated for a window size of x by x pixels, with $x = 1, 3, 5$. In the former case, the EC_a-V data were treated as unavailable and instead their mean was used.

According to the input, the models were abbreviated as H_WS _{x} or HV_WS _{x} and in total nine different input layers were tested. Due to the compositional nature of texture data, two textural fractions were used as output variables ($N_{out} = 2$): the topsoil clay and topsoil sand content. The results however only focus on the topsoil clay content (referred to as tclay). The values for the slope of the activation function, the learning rate and the momentum term were set to 0.9, 0.01 and 0.1, respectively.

Two preprocessing steps were performed: (i) the input data were normalized to a mean of zero and a unit standard deviation and (ii) the output data were rescaled to the interval [-0.9,0.9] to avoid the ends of the transfer function. This allows the network to produce an output when an observed EC_a-value lies outside the range used for the ANN development and avoids slow learning rates.

The ANN analysis was performed using LNNS, an artificial neural network simulator developed at the Laboratory of Forest Management and Spatial Information Techniques of the Ugent, Belgium, which is downloadable from <http://dfwm.ugent.be/forsit/>.

8.2.6 Validation Procedure

As stated above (section 8.2.5), there was no independent validation data set, instead the performance of the different neural network models was evaluated based on the average validation indices of the ten test sets. This introduced some bias in the validation indices, resulting in slightly better results than would obtained with an independent validation data set (Ozesmi et al., 2006). Four validation indices were calculated: (i) the mean estimation error (MEE), (ii) the mean squared estimation error (MSEE), (iii) the coefficient of determination (R^2), and (iv) the uncertainty or variance (VAR) among the $n_{sp} \cdot n_{in}$ realizations of each observation of the test datasets. The following formulae were used:

$$\text{MEE} = \frac{1}{n_{sp} \cdot n_{in}} \sum_{i=1}^{n_{sp}} \sum_{j=1}^{n_{in}} \frac{1}{n_{te}} \sum_{k=1}^{n_{te}} (y_{ik} - y_{ijk}^*) \quad (8.5)$$

$$\text{MSEE} = \frac{1}{n_{sp} \cdot n_{in}} \sum_{i=1}^{n_{sp}} \sum_{j=1}^{n_{in}} \frac{1}{n_{te}} \sum_{k=1}^{n_{te}} (y_{ik} - y_{ijk}^*)^2 \quad (8.6)$$

$$R^2 = \frac{1}{n_{sp} \cdot n_{in}} \sum_{i=1}^{n_{sp}} \sum_{j=1}^{n_{in}} \frac{\sum_{k=1}^{n_{te}} (y_{ijk}^* - \bar{y}_i)^2}{\sum_{k=1}^{n_{te}} (y_{ik} - \bar{y}_i)^2} \quad (8.7)$$

$$\text{VAR} = \frac{1}{n_{sp} \cdot n_{te}} \sum_{i=1}^{n_{sp}} \sum_{k=1}^{n_{te}} \frac{1}{n_{in} - 1} \sum_{j=1}^{n_{in}} (y_{ijk}^* - \bar{y}_{ik}^*)^2 \quad (8.8)$$

where y_{ik} is the observed tclay content of sample k of test set i , y_{ijk}^* is the predicted tclay content of sample k of test set i for initialization j , \bar{y}_i is the mean observed clay content of test set i , \bar{y}_{ik}^* is the mean predicted tclay content of sample k of test set i over all initializations, n_{te} is the number of samples in each test set, n_{sp} is the number of splits into test and training data sets and n_{in} is the number of initializations in each training process.

The MEE measures the biasness of the predictions; the MSEE quantifies the prediction accuracy, while the R^2 measures the variability of the tclay observations that is explained by the model. In addition, the VAR index indicates the reliability of the tclay predictions. Since all four indices had a different focus, we determined the best model performance using a rank score that pays equal attention to each of them. Suppose there are p models, for each validation index, the models were ranked from best to worst. A score of 1 is assigned to the best model and p to the worst model. For each model, the scores are then summed to obtain the overall rank score R_o . With q validation indices R_o values are in the range $[q, pq]$. For the ease of interpretation, R_o was normalized to the normalized rank score R_n (Parasuraman et al., 2006):

$$R_n = \frac{pq - R_o}{pq - p} \quad (8.9)$$

A model with a good overall performance has an R_n value close to 1.

8.2.7 Multiple Linear Regression

As an alternative for the prediction of the tclay content based on EM38DD data, the use of stepwise multiple linear regression (SMLR) was evaluated. In SMLR modeling, predictor variables are added one at a time to the regression to see if the model improves upon the addition of that variable. The model is fit based on least squares: the sum-of-squares of the differences between observed and predicted values is minimized. Adding the first predictor is based on the correlation with the output (dependent variable) and by this, the second predictor is fixed in case

of two predictor variables. A multiple linear regression model is defined by the following:

$$Y = a_0 + \sum_{k=1}^K a_k X_k \quad (8.10)$$

where Y is the dependent variable, X_k is the k^{th} independent variable, a_0, a_1, \dots, a_K are the regression coefficients and K is the number of independent variables.

The two EM38DD orientations were used as independent variables ($K = 2$) and due the limited soil dataset, all 78 soil samples were used for the SMLR analysis. The validation however was performed in an equivalent manner to the ANN analysis since the indices were calculated and averaged out for the 10 test sets. It should be noted that no VAR index could be calculated since there are no initializations in SMLR.

8.3 Results and Discussion

8.3.1 EC_a and Texture

Table 8.1 shows that the values of EC_a -V were slightly larger than those of EC_a -H and skewed to the right. Nevertheless the coefficients of variation (CVs) were in the same order and the linear correlation between them was 0.78 which is rather low compared with other studies (Korsaeth, 2005; Triantafilis and Lesch, 2005).

TABLE 8.1 – Descriptive statistics of the EC_a -H, EC_a -V ($n = 19694$) and the textural fractions of top- and subsoil.

		Min	Max	Mean	Skewness	CV [%]
		EC_a [mS m ⁻¹]				
EC_a -H		10.9	30.1	18.0	0.38	15
EC_a -V		14.7	33.7	21.2	1.07	13
		Textural fractions				
Topsoil	clay	14.30	23.71	19.12	-0.04	13
	silt	20.80	54.20	35.06	0.58	21
	sand	23.10	64.80	45.82	-0.23	20
Subsoil	clay	3.20	22.11	9.07	0.85	45
	silt	3.70	46.30	16.80	0.95	55
	sand	31.70	92.90	74.14	-0.91	18

The interpolated EC_a maps (Fig. 8.5) show that both orientations were clearly influenced by the subsoil creek pattern: a similar pattern can be observed, with the highest EC_a values occurring in the channels. On both maps two dominant linear features (channels) of increased EC_a are present: one in the eastern part of the field, parallel to the boundary with clear side branches and one diagonally

crossing the western part of the field. So regardless the theoretical depth-response curve, the subsoil variability also influenced the EC_a -H data strongly.

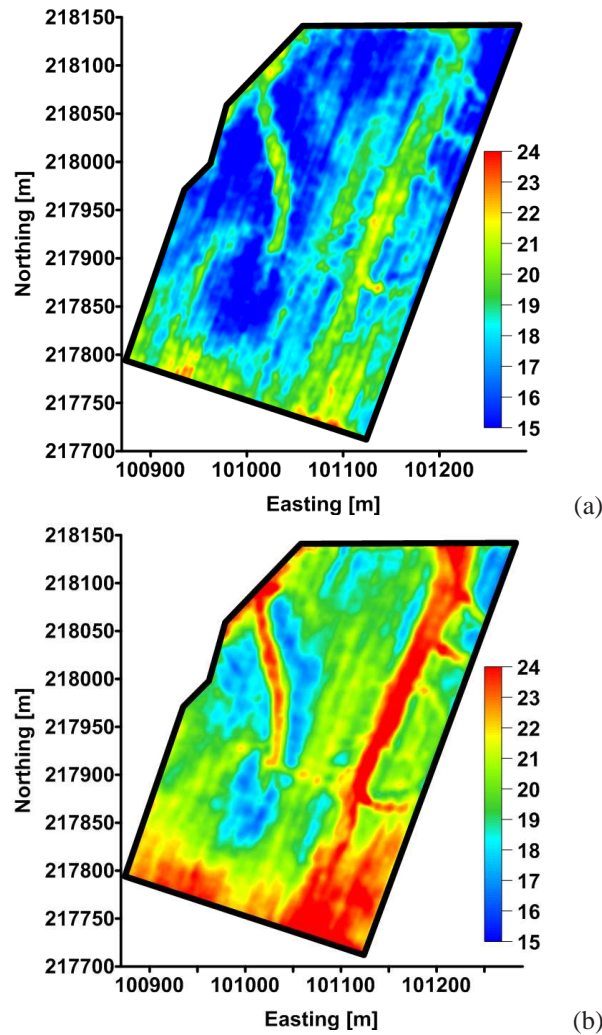


FIGURE 8.5 – Maps of (a) EC_a -H and (b) EC_a -V. The map units are $ms\ m^{-1}$.

Both EC_a signals exhibited the same spatial behaviour represented by a double nested spherical variogram model, but more noise was registered by the horizontal orientation of the sensor (Fig. 8.6). The relative nugget effect (RNE) of the EC_a -V measurements was 4%, while for the original EC_a -H data the RNE was 30%. After filtering the EC_a -H data, its RNE reduced to 23%. A first scale of variability could be related to the creek pattern of the subsoil which is present within the range of 35

m, while the second range of 200 m represents the wider within-field variability.

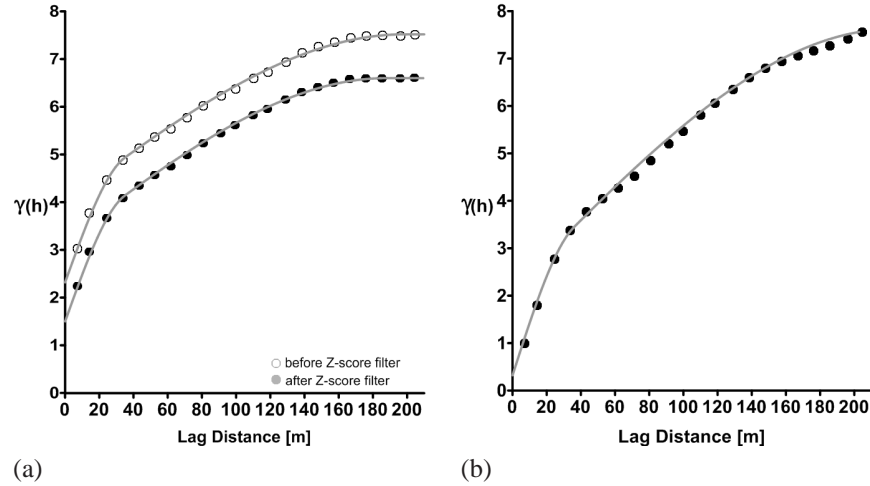


FIGURE 8.6 – Variograms of (a) EC_a -H before and after the application of a Z-score filter, (b) EC_a -V.

The large textural variability of the subsoil is expressed by high CVs for clay and silt in the subsoil (Table 8.1). For the clay fraction, the CV of the subsoil was 45%, compared to a CV of 13% for the topsoil. The topsoil texture was loam and clay loam, while in the subsoil it ranged over five USDA textural classes (sand, loamy sand, sandy loam, loam and silt loam) (Vitharana et al., 2006).

The 78 top- and subsoil samples are presented in Fig. 8.7. Based on these samples only, no channel pattern could be recognized in the topsoil, the highest tclay values were situated around a northeast line in the center and in the bottom left corner of the field. The subsoil clay content on the other hand followed the observed EC_a variability.

The Pearson (r) and Spearman (ρ) correlation coefficients between the textural fractions and extracted EC_a at the sampled locations are given in Table 8.2. The differences between the Pearson and Spearman correlation coefficients were limited, so the degree of linearity is expressed well by the Pearson correlation coefficient. Subsoil texture correlated strongly with EC_a -H and EC_a -V, with an r -value around 0.7. This also points to the influence of the subsoil on the EC_a -H measurements. EC_a -H correlated even better with the subsoil clay content ($r = 0.70$) than with tclay ($r = 0.61$). The relation of tclay observations and the two sensor orientations was shown in Fig. 8.8. The topsoil clay content had the most linear relation with EC_a -H although the correlation coefficient was moderate ($r = 0.61$) and quite some spread was found. With EC_a -V, the cloud of data points became more scattered and the r value was only 0.44. This rather limited linearity

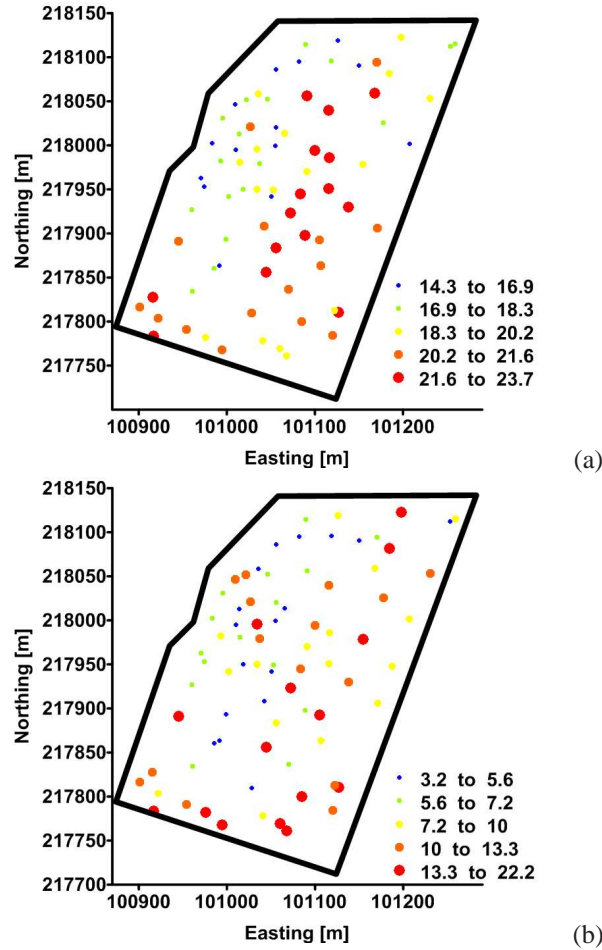


FIGURE 8.7 – Map indicating the location and the clay content [%] of (a) the topsoil samples and (b) the subsoil samples.

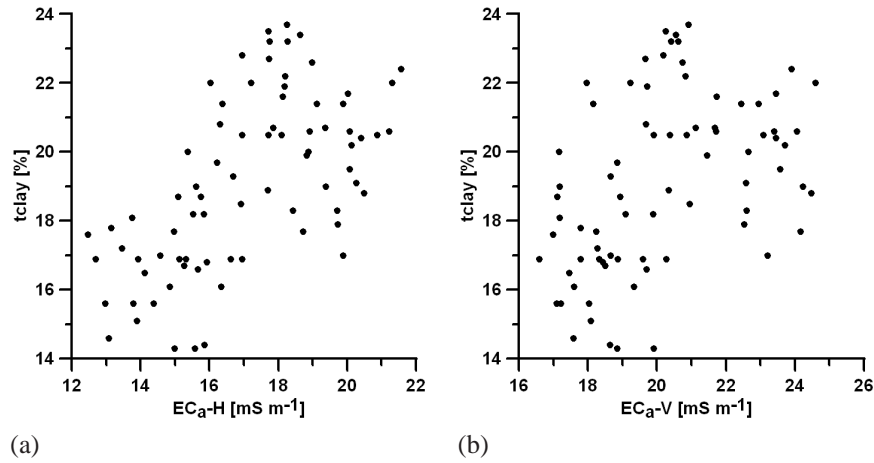
motivates the use of ANNs.

8.3.2 Artificial Neural Network Analysis

The effect of changing the input layer of the ANN by increasing the window size or eliminating the vertical sensor orientation is given in Table 8.3. The validation indices of the tclay predictions are shown for the nine different model architectures. The MEE of all models was smaller than the precision of clay determined using the pipette procedure, which is 1% (Gee and Bauder, 1986), indicating the low mean error of all network models. As a consequence, the MEE is not used

TABLE 8.2 – Pearson and Spearman correlation coefficients between EC_a and the textural fractions ($n = 78$).

	Topsoil			Subsoil		
	clay	silt	sand	clay	silt	sand
Pearson (r)						
EC_a -H	0.61	0.74	-0.77	0.70	0.72	-0.72
EC_a -V	0.44	0.66	-0.65	0.69	0.71	-0.72
Spearman (ρ)						
EC_a -H	0.63	0.77	-0.79	0.75	0.76	-0.76
EC_a -V	0.49	0.69	-0.69	0.74	0.75	-0.75

**FIGURE 8.8** – Scatterplots of (a) the topsoil clay content and EC_a -H, (b) the topsoil clay content and EC_a -V.

further to compare the models. Concerning the different changes in the input layer, the MSEE, the R^2 , and the VAR followed the same trend. Increasing the amount of contextual information or ignoring the information obtained with the vertical EM38DD orientation deteriorated the accuracy and the reliability of the tclay predictions. According to the R^2 and the VAR indices, the best model was HV_WS1 explaining 65.5% of the tclay variability with an uncertainty of 0.052%². Its MSEE of 2.829%² was slightly larger than the MSEE of the HV_WS3 model, so based on the MSEE only no differentiations could be made between HV_WS1 and HV_WS3.

Focusing on the effect of an increasing input window size, the least significance difference (LSD) test ($P < 0.05$) was used to compare the different levels of contextual information. A least significant difference between the means of the validation indices is calculated as the product of the standard deviation of the mean

TABLE 8.3 – Validation indices of the different network architectures, ordered according to a change in the input window size or a change in the number of sensor orientations.

		MEE	MSEE	R ²	VAR	R _n
Effect of window size						
HV_WS ₁		0.459	2.829 ^{a(*)}	0.655 ^a	0.052 ^a	0.96
HV_WS ₃		0.130	2.769 ^a	0.550 ^b	0.222 ^b	0.83
HV_WS ₅		0.210	3.531 ^a	0.549 ^b	0.801 ^b	0.54
HV_WS ₇		-0.329	6.057 ^b	0.356 ^c	3.046 ^c	0.25
HV_WS ₉		-0.356	7.751 ^c	0.303 ^c	4.408 ^d	0.13
HV_WS ₁₁		0.325	11.591 ^d	0.208 ^d	7.675 ^e	0.00
Effect of EM38DD orientation						
WS ₁	H_WS ₁	-0.179	3.045 ^{a(**)}	0.472 ^a	0.059 ^a	0.75
	HV_WS ₁	0.459	2.829 ^a	0.655 ^b	0.052 ^a	0.96
WS ₃	H_WS ₃	0.159	3.241 ^a	0.466 ^a	0.091 ^a	0.63
	HV_WS ₃	0.130	2.769 ^a	0.550 ^a	0.222 ^b	0.83
WS ₅	H_WS ₅	0.206	4.323 ^a	0.441 ^a	0.234 ^a	0.42
	HV_WS ₅	0.210	3.531 ^a	0.549 ^a	0.801 ^b	0.54

* Values with the same letter are not significantly ($P < 0.05$) different by an LSD-test.

** Values with the same letter are not significantly ($P < 0.05$) different according to a Student's t -test.

difference and its associated t -value (Webster, 2007). Any difference that exceeds this LSD is considered to be significant. Up to a window size of 5 by 5 pixels the MSEE values were comparable, higher window sizes resulted in a significant increase of the MSEE. The R^2 and the VAR significantly differed between models with ($WS > 1$) and without ($WS = 1$) contextual input information. So, contextual information did not contribute to an improved model. These unexpected results were probably due the small dataset which resulted in too complex network architectures when taking into account contextual information. It is known that ideally the size of the dataset should exceed the number of weights in the network (Burnham and Anderson, 2002). So if the number of soil samples could be extended, the influence of contextual information would probably increase. Besides, the reliability of the predictions clearly depended on the number of input nodes as the VAR index increased when a more complex input layer was used. This conclusion was valid regardless the nature of the input data: contextual or extra sensor information.

The relative importance of incorporating both sensor orientations in the network was indicated by a deterioration of the MSEE, R^2 and VAR indices when treating the EC_a-V data as unavailable. This trend was valid for all tested window sizes. Nevertheless, almost no significant differences were found with the Student's t test ($P < 0.05$).

Comparing the overall performance of all the models was done by integrating

the different indices into the R_n -score. Based on the R_n , a well-defined order in the performance of the network models was obtained (Table 8.3, Fig. 8.9). The best model was HV_WS1 with an R_n of 0.96, followed by HV_WS3 ($R_n = 0.83$). The negative effect of including contextual information increased with an increasing window size, due to the increasing complexity of the network. On the other hand the effect of including the vertically sensed EC_a had a stable and positive effect on the tclay predictions. On average the R_n -score increased with 30% (with a standard deviation of 2%) when both EC_a -V and EC_a -H data were used instead of only EC_a -H. Information obtained with the vertical sensor orientation seemed to be essential to optimize the tclay predictions. This also confirmed our hypothesis that under circumstances of very heterogeneous subsoil properties, the EC_a -H is strongly influenced by the subsoil, masking information about the topsoil variability. A neural network analysis seemed to be an appropriate technique to extract the topsoil information from the two signals obtained with the EM38DD sensor.

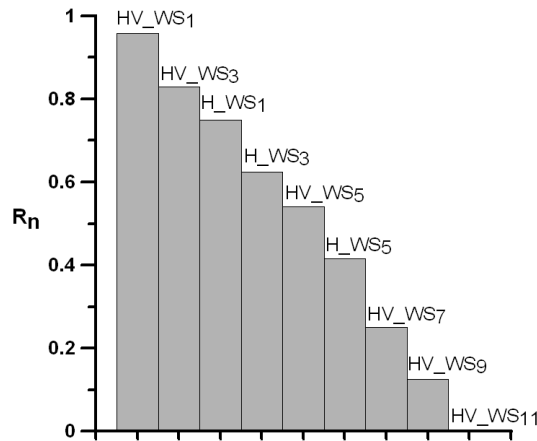


FIGURE 8.9 – The R_n -score of all models.

8.3.3 Comparison with SMLR

Concerning the advantage of using a more complex technique like ANNs the comparison was made with stepwise multiple linear regression. The SMLR was performed with EC_a -H and EC_a -V as predictor variables. Both orientations contributed significantly to the model and following regression equation was found:

$$\text{tclay} = 13.25 + 1.53EC_a\text{-H} - 0.99EC_a\text{-V} \quad (8.11)$$

The goodness-of-fit of the model (R^2) was 0.488 and the adjusted R^2 was 0.47. The goodness-of-fit of a univariate regression based on EC_a -H data only was 0.37

(adjusted $R^2 = 0.36$), indicating that also in this analysis adding the vertical orientation improved the tclay prediction. Nevertheless, the individual importance of each orientation was not assessed due to a multicollinearity concern since the linear correlation between the two predictor variables was considerably high (0.78). The predictor variables in any regression problem might be intercorrelated and the term multicollinearity is used to describe the case when the intercorrelation of predictor variables is high. Multicollinearity does not preclude the use of regression, but can make it impossible or difficult to assess the relative importance of individual predictor variables from the estimated coefficients of the regression equation. With variance inflation factors of 6.96 no serious problems of multi-collinearity arise however (Neter et al., 1996).

The model was validated based on the average indices of the ten test data sets: the MVL model had a MEE of -0.045, an MSEE of 3.72 and 41% of the tclay variability was explained. Compared to the HV_WS1 model, the MSEE and R^2 indices were considerable worse. So using ANNs clearly improved the clay predictions. This indicated the outperformance of ANNs to stepwise multiple linear regression, probably due to the nonlinear depth response of an EMI sensor, the limited linear relation with the tclay content and the complex interactions of all the EC_a -influencing soil properties.

8.3.4 Mapping the topsoil clay content

The predictions of tclay obtained with the HV_WS1 model are shown in Fig. 8.11(a). The tclay data ranged between 16 and 23.5%, with the highest clay content in the center of the field. Despite the integration of the EC_a -V and EC_a -H data, a linear channel pattern of increased tclay content was still visible in the west of the field. A map indicating the locations and the magnitude of the estimation errors was projected on a classified EC_a -V map with the mean EC_a -V value (21.2 mS m^{-1}) as cutoff value (Fig. 8.11(b)). This boolean classification shows the most clayey zones of the field and as such the subsoil creek pattern is apparent. The largest errors occurred in the channel of the west: there the tclay content of the samples was clearly overestimated by our model. So, the effect of the subsoil was not filtered out completely by the neural network technique.

However, if only the EC_a -H data (H_WS1) were used, the maximum predicted tclay reduced to 21% and the channel pattern was clearly visible in the tclay predictions (Fig. 8.11(c)). The errors of this model indicated that not only the topsoil clay content in the western channel was overestimated, but also in the central part of the field the errors increased (mainly underestimations) (Fig. 8.11(d)). This prompted the need for subsoil textural information when predicting the clay content of the topsoil. Therefore, fusing the two EM38DD orientations in an ANN masked out the subsoil influence on the EC_a -H signal considerably, increasing

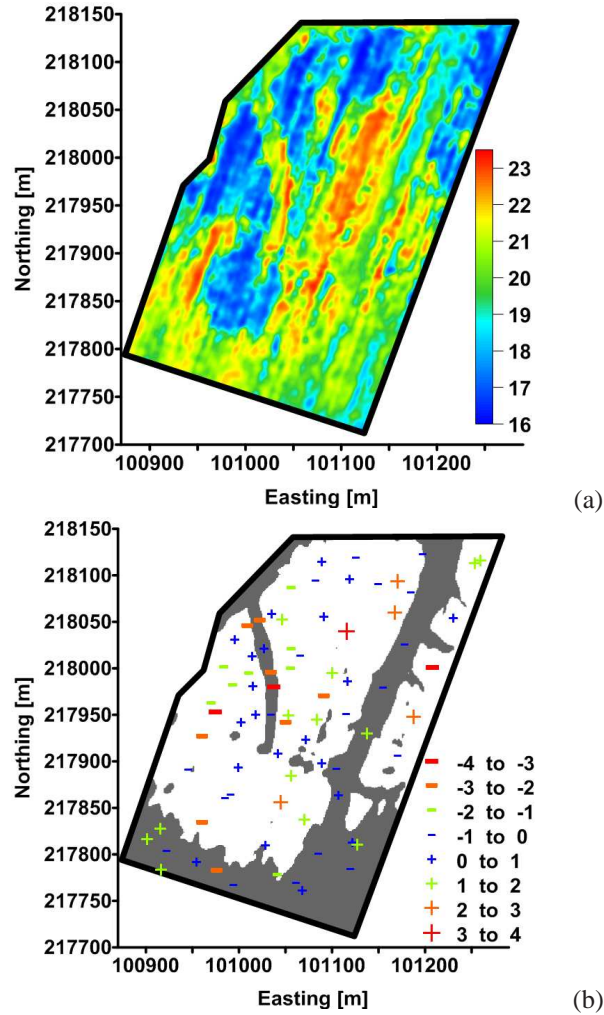


FIGURE 8.10 – (a) Map of the tclay content predicted by the HV_WS₁ model and (b) Location map of the estimation error of the HV_WS₁ model, projected on a classified EC_a-V map.

consequently the accuracy of the predicted tclay content.

8.4 Conclusions

In this study, the two signals (EC_a-V and EC_a-H) obtained with an EM38DD sensor were combined in an ANN model to estimate the clay content in the topsoil. Based on both orientations, the influence of the heterogeneous subsoil on the tclay

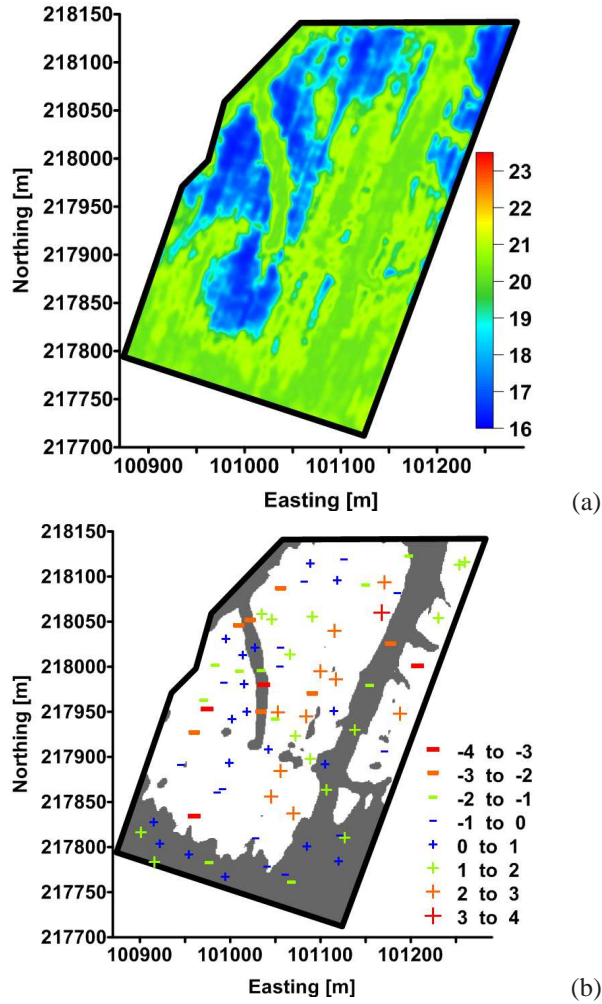


FIGURE 8.11 – (a) Map of the tclay content predicted by the H_WS₁ model and (b) Location map of the estimation error of the H_WS₁ model, projected on a classified EC_a-V map.

prediction reduced considerably. Compared with an ANN analysis based on EC_a-H data only, the tclay predictions were more accurate in terms of MSEE, R^2 , and also the uncertainty of the predictions was lowest. The normalized rank score increased with 30% when two sensor orientations were used instead of only the horizontal one. Integrating contextual information in the ANN model did not improve the tclay predictions: the validation indices deteriorated with an increasing input window size. This was probably caused by the small size of the soil dataset, but

further research to explore the possibilities of large datasets for use in contextual neural networks is required.

It can be concluded that fusing the two EMI signals by an ANN analysis increased the topsoil information present in only one signal. This approach opens perspectives for the optimal use of proximally sensed soil data. Artificial neural networks appeared to be a useful tool for optimizing the prediction of textural information from depth-weighted EMI responses.

9

General conclusions and future research

9.1 Introduction

Knowledge of the spatial variability of a soil is the key to efficiently managing natural soil resources. Soils are inherently highly variable and this characteristic complicates the spatial characterization of a soil. Soil sensors were developed as an alternative for intensive soil sampling since it is an easy-to-obtain source of high-resolution soil information. This thesis concentrates on a new type of EMI-based soil sensor, the EM38DD, developed in 2001. Different approaches to analyzing EM38DD signals were addressed.

The final chapter reflects on the most important conclusions of this thesis and discusses the research questions as postulated in Chapter 1. Also some ideas for future research are put forward.

9.2 General conclusions

9.2.1 Answers to the research questions

The general aim of the thesis was to investigate the use of a proximal EMI soil sensor to inventory and characterize soil properties at a high resolution. Nine research questions were posed to reach this aim. Through the different research chapters answers were given to these questions and summarized here.

Question 1. *Which soil management applications can benefit from an EMI sensor-based approach?*

In this thesis, five soil management applications were addressed:

- Site specific soil management by adjusting agricultural management to the observed within-field soil variability
- Soil quality evaluation through the spatial variability of earthworms occurrence
- Detecting geogenic traces in soil to provide information about Pleistocene periglacial periods
- Forest site planning through the characterization of textural heterogeneities in forest soils
- Improved mapping of soil textural properties by fusing dual EMI signals

The identification of within-field soil classes for use in site specific management can be performed successfully with EMI-based measurements. In Chapter 4 this method was applied to the domain of soil chemistry, while Chapter 5 focused on soil biology. Using an EMI sensor to characterize soil chemical properties only

works if the following condition is met: the soil chemical property of interest has to interact with at least one of the EC_a influencing properties, as is the case for soil nitrate. EMI clearly improved the characterization of nitrate dynamics since nitrate movement depends on the soil texture: the risk to nitrate leaching decreases in soils with a higher silt and clay content. Also the characterisation of soil biological variability by EMI performs well through the interaction with soil texture. Deep burrowing species were found to live in spatially separated clusters and this occurrence pattern was in accordance with the subsoil textural variability.

Chapter 6 illustrated that former geogenic phenomena like fossil periglacial features can be detected by an EMI sensor. Their detection is based on a contrast between properties of these features and the surrounding soil. Frost-wedge pseudomorphs introduce a textural heterogeneity in the host soil, since the cracks were filled with wind-blown sand and silt material. Besides, geomorphological features (like the former creeks in Chapter 8) were also visible on high resolution EC_a maps.

Site characterization by EMI was applied successfully in a forestry context (Chapter 7). More specifically, a textural discontinuity was delineated by a probability approach based on the ratio of the horizontal to the vertical signals. Although EC_a measurements were taken manually at a resolution of 20 m by 20 m, the presence of a clay substratum in the sandy forest soil was identified well by analyzing both the EM38DD signals.

In Chapter 8, the dual EM38DD signals were fused in an ANN and this helped to filter out the dominant influence of a very heterogeneous subsoil on the horizontal signal. The accuracy of topsoil clay predictions increased by using two instead of one EM38DD signals as an input in an ANN.

Yet, the here mentioned applications do not form a comprehensive list. However, they all have one important characteristic in common: a lateral or vertical change in one of the EC_a influencing factors.

Question 2. *Does the within-field soil variation of Flanders (Belgium) require high resolution soil data?*

Yes.

A common criticism to site specific soil management in Flanders is the relatively small size of the agricultural fields. Nevertheless the study sites used in this thesis have areas ranging from 1 ha to 10.5 ha and all of them exhibited reasonable EC_a variability. In Chapter 4 the difference between the information present on a conventional soil map and on an EC_a map is illustrated. Even this small field (1 ha) exhibits a considerable soil textural variability. A differential N management in the two distinct EC_a zones could be an efficient way to reduce the amount of nitrate leached in autumn. To capture that variability using a classical soil sampling approach, the labour and economic costs are high due to the large number of

soil samples. Here high resolution soil sensor fit perfectly: they provide a quick view on the variability present in the soil which can be used to efficiently reduce the number of soil samples and determine their locations.

Question 3. *Lateral and vertical soil heterogeneity: a single or dual EM38 problem?*

Both lateral and vertical soil variability were addressed in this thesis: lateral soil variability was studied in Chapter 4, 5 and 6, while in Chapter 7 and 8 both lateral and vertical soil variability were addressed.

In Chapter 4 and 5, one single orientation was sufficient to characterize the lateral soil variation. The EC_a patterns were similar for the dual signals. Notwithstanding, the presence of nested textural variability in a field can complicate a single signal interpretation. Nested soil variability occurs when spatially continuous patterns of soil variability are present within the 'normal' lateral soil variability (as in Chapter 6 and Chapter 8). In these chapters, the variograms of the EC_a measurements were all modeled using a nested structure. For the study site with the pleistocene wedge casts the textural variability of the host material appeared to be larger than the textural contrast of wedge versus host material. A single EM38 signal did not reveal the polygonal pattern. To eliminate the dominant textural variability of the host material, the dual signals were compared in the profile ratio (PR). Looking relatively at the two EM38DD signals revealed the polygonal pattern.

However, the presence of a strong textural discontinuity in the soil profile results in differential spatial patterns for both orientations. The relation between the dual signals expresses the degree of vertical soil variability: the higher EC_a values differ, the stronger the textural discontinuity. In Chapter 7 the ratio of the horizontal to the vertical signal (PR) was used to reveal the vertical profile variability since the delineation of a clay was not obvious from a single orientation. In Chapter 8 combining the dual signals in ANNs improved the prediction of topsoil clay content.

It was concluded:

- for homogeneous soil profiles without nested variability a single EM38DD orientation suffices;
- for homogeneous soil profiles with nested variability the ratio of the dual EM38DD signals can eliminate the dominant lateral variability and reveal the remaining variability;
- for heterogeneous soil profiles the dual EM38DD signals are essential to capture the vertical soil variability

In this context homogeneous profiles refer to soil profiles without a strong textural discontinuity.

Question 4. *Is there a difference in characterizing top- or subsoil properties?*

Yes.

From the theoretical response curves (Fig. 2.10) it was concluded that the horizontal EM38 signal is dominantly influenced by the topsoil, while the vertical signal has a dominant influence of the soil at 0.30–0.50 m depth. However, these EM38 signals also have contributions from the entire soil profile. Material below a depth of 2 m even contributes 12 % and 25 % to the horizontal, respectively vertical response.

In Chapter 8 the horizontal signal was dominantly influenced by the very heterogeneous subsoil. So although the peak response to the horizontal signal comes from the near surface soil, it is too easily ignored that the subsoil still contributes significantly. The degree of influence of the subsoil on the horizontal reading depends on the conductivity values of the subsoil layer. In Watervliet with a highly variable subsoil texture due to the presence of a former water channels, the horizontal reading reflected mainly the subsoil variability. So the theoretical response curves should be interpreted with caution to the conductivity of the different soil layers (top- and subsoil). For homogeneous soil layers, it is evident that the horizontal signal provides information about the topsoil properties and vice versa, the vertical signal has to be used to characterize subsoil properties. For heterogeneous soil profiles, one has to take into account the conductivity differences between sub- and topsoil to decide whether one signal is sufficient or whether dual signals are required to obtain the desired sub- or topsoil information.

Question 5. *Does the dual dipole character supply complementary soil information?*

Yes.

In Chapter 6, 7 and 8 dual signals were needed to reach the aim of the study. These chapters were concerned with the inventory of both heterogeneous soil profiles and homogeneous soil profiles characterized by a nested variability. A single EMI signal only provides information concerning the lateral soil variability. No conclusions related to the soil profile characteristics can be made. Dual EMI signals allow the characterization of a soil profile: homogeneous soil profiles have similar EC_a patterns while differing EC_a patterns indicate a soil textural discontinuity in the soil profile.

Question 6. *Are soil samples and lab analyses now redundant?*

No.

High resolution EC_a data allow for a fast and cost-efficient characterization of a site. Nevertheless the EC_a values are not interpretable as such since they are controlled by a variety of factors. Always soil samples are necessary to interpret the absolute EC_a values. However, compared to a traditional soil survey the number

of soil samples reduces considerably in an EMI-based soil survey. In a traditional way, soil samples are the only means of soil information, and as such the number of soil samples needed is proportional to the required detail of soil information. So mostly the economic consequences of high resolution soil surveying in a traditional way are large. To that aspect, EMI soil sensors offer a solution. An EC_a map is a major source of information concerning the spatial variability of soil properties. This EC_a map serves as a guide to locate a limited number of soil samples. In fact, the EC_a data range should be captured with special attention to extremes and spatial outliers. So, an EMI soil sensor does not replace the soil auger; soil samples still need to be collected and analyzed.

As an illustration, the costs and quality associated with soil textural mapping are discussed for both a traditional and an EMI-based approach in a Belgian context. Traditionally, soil textural classes were determined by palpation during soil auger observations, at an average density of one observation per ha. In this way one person can map the textural classes of approximately 25 ha per day with a resolution of 100 by 100 m. To map the same area by an EMI survey in one day a line interval of 5 m can be obtained at a speed of 10 km/h. Within the lines the EC_a data are separated 3 m when measurements are taken every second. The quality of the soil classes can be expressed by the resolution of the data: with an EMI-based approach 20 times more detail is obtained. Nevertheless, the cost of an EMI survey at this resolution is €100/ha including the variable costs of transport and fuel. It is upon the user of the map to decide what amount of detail is needed. For each purpose one should balance the quality and the economic costs to map soil texture.

If a high resolution map of the percentage clay is required for the same area of 25 ha, the samples should be analysed in the laboratory and afterwards the resulting clay fraction has to be interpolated to a grid. With no prior (ancillary) information available, 100 locations separated by an appropriate interval are recommended for reliable variogram modeling and high resolution mapping of soil properties (Webster and Oliver, 2001). Therefore, an additional collection of 75 soil samples or one extra day of field work is required in a traditional approach. With the EMI approach the resulting EC_a map can be used for targeted sampling in areas of significantly different EC_a . Kerry and Oliver (2007) demonstrated that based on ancillary data a sampling intensity of 50 samples should suffice for kriging soil properties for precision agriculture purposes. However other geostatistical methods using ancillary information (like regression kriging) can further reduce this number. With the cost for one textural analysis in the laboratory being €82 per sample, the two approaches can be compared economically (for a 25 ha area):

- Traditional soil survey
 - 100 samples to be analysed in the laboratory: €8200
 - 2 days of fieldwork for the collection of the soil samples

- EMI-based soil survey
 - laboratory cost for analysing a maximum of 50 samples: €4100 max
 - 2 days of fieldwork: 1 day for the EMI soil survey plus 1 day for the collection of the samples
 - EMI survey: €2500

It can be concluded that an EMI-based soil mapping approach complemented with a useful number of soil samples reduces the costs associated with a traditional approach. An EMI based approach to target soil samples is believed to be an appropriate way to balance the accuracy and the costs of high resolution soil mapping.

Question 7. *Which processing techniques are suitable to analyse EM38DD data?*

First, a preprocessing phase investigates the distribution and spatial behaviour of the collected EC_a measurements. Here outliers, erroneous measurements or noise in the data can be removed or filtered. Then, the EC_a measurements are interpolated. Depending on the resolution of the EMI soil survey the interpolation phase becomes more or less crucial. Finally, the EC_a maps should be interpreted. A fuzzy- k -means classification of one or more EC_a signals identifies groups of pixels with similar EC_a values and this facilitates the interpretation of the spatial EC_a variability. It should be noted however that a classification implies a simplification of the reality (i.e. the EC_a variability) which is not always wanted. The resulting EC_a map is in fact a source of exhaustive, ancillary information which can be used to predict other soil properties not intensively sampled like soil nitrate (Chapter 4), soil texture (Chapter 8) or the presence/absence of a textural discontinuity (Chapter 7). Here geostatistical methods play an important role: depending on the goal of the study stratified simple kriging, ANN or indicator kriging was used. Additionally, map comparison statistics allow the determination of similarities between an EC_a map and a map of the soil property of interest. In a spatial context, map comparison statistics that go beyond the pixel by pixel comparison assessed the map similarities better. Also in the presence/absence studies the prevalence effect can play a major role and κ^* gains interest.

Question 8. *What is the optimal procedure to increase information extraction from dual signals?*

It was concluded that rather than interpreting dual signals next to each other, integrating them improved the soil inventory. A different level of integration was applied throughout the thesis. In Chapter 5 with a homogeneous soil profile, the arithmetic average of the dual signals was used. Herewith, no extra information concerning the soil variability was obtained, it only combined the information of

both EM38DD signals. In Chapter 6 with a homogeneous soil profile but a continuous pattern present in the soil, the dual EM38DD signals were rationalised and clearly the information increased. The soil variability of the host material was filtered out on the PR map and by this the polygonal pattern of frost-wedge pseudomorphs became visible. In Chapter 7 with a heterogeneous soil profile, also the PR appeared to be most informative. On neither of the original EC_a maps the clay substratum could be delineated. In Chapter 8 the methodology however oversteps the simple combination of the dual signals in a ratio. ANNs were proven to be the optimal procedure to extract information about the topsoil texture from the dual EM38DD signals by filtering out the influence of the subsoil. Compared to predictions based on single signal, the topsoil clay predictions increased when fusing both the horizontal and the vertical EC_a measurements in an ANN. ANNs offer perspectives to optimally combine multi-sensor data to increase the information present in one sensor.

Question 9. *What are the limitations of the EM38DD soil sensor?*

The main limitation of the EM38DD sensor is the fact that its EC_a reading is an integrated value over all soil layers. Two signals are measured by the sensor orientations but each signal has contributions from the whole soil profile according to their depth response curves. As such, the electrical conductivity value of each soil layer also plays an important role in the final measurement. Consequently the EM38DD sensor is unable to link its readings directly to a soil property at a specified soil depth. The soil profile cannot be reconstructed by EM38DD data solely. Top- or subsoil influences can be enhanced or filtered (like in Chapter 8), but for accurate depth profiling additional information is needed, either by soil sampling, either by another complimentary proximal soil sensor. Nevertheless, the EM38DD soil sensor provides indications about the degree of soil heterogeneity and thus enables the identification of zones of interest. Also absolute EC_a values are not interpretable in se. But, the power of an EMI soil survey lies in the relative EC_a changes in field, reflecting the soil spatial variability.

9.2.2 Impact of conclusions

This thesis clearly illustrated that soil EC_a data are a valuable source of ancillary information to improve soil inventories. The objective of the inventory appeared to be a crucial factor determining the approach to EM38DD data processing. A wide range of soil management applications can benefit from an EMI based soil inventory. The condition to be fulfilled is that one of the EC_a influencing factors plays a key role in the application. Moreover, the dual characteristic of the EM38DD soil sensor was found to be primordial for adequate lateral and vertical soil inventories. High resolution soil EC_a data were proved to increase the quality of soil inventories for multiple purposes.

9.3 Future perspectives

*When you use information from one source, it's plagiarism;
When you use information from many, it's information fusion -*
Belur V. Dasarathy

In general, it can be concluded that a multi-sensor approach should be supported on the condition that the sensor signals are complementary. Soil sensors not only produce large amounts of soil data in a relatively fast and economic way, combining complementary signals seeks to obtain more inferences than can be derived from a single sensor. The challenge now lies in stipulating the optimal sensor combination with the maximum predictive capacity suiting the postulated objective. New data analyzing techniques should be explored in this context. Soil survey methods are changing from strong interpolation methods needed in the situation of poor soil data availability to methods focusing on information extraction, data fusion and filtering in the situation of high resolution multi-dimensional sensor data (in information space). Spatial wavelets analyses as a multi-resolution based method, factorial kriging as a scale-decomposing method, artificial neural networks as a data fusion method and principal component analysis as an information extraction method all appear powerful techniques to be applied to proximally sensed soil data. Besides, techniques applied successfully to remote sensing data can be explored for adaptation to proximally sensed soil management applications.

Another idea for future research refers to the scale of the study. In this thesis the within-field scale was focused. However proximal soil sensors also have potential on a landscape and even a regional scale. A major challenge in using EC_a to model soil spatial variability across a larger scale is how to integrate and utilize data collected in different fields and/or at different times. For proximal soil sensors, a larger scale entails a loss of data resolution if the advantage of relatively quick surveying should be maintained. At a larger scale, fusion of remote and proximal sensors can be targeted for processes acting at a regional level. Similarly, if the focus is on a smaller scale, very detailed EC_a data are needed. Proximal sensing for archeological purposes is an example of the latter case and has gained a lot of interest the last years. The implementation of the Treaty of Malta (1992) state that the archaeological potential of a site to be cleared or reconstructed should be assessed prior to the works. Through this convention, geophysical soil sensors are now successfully used in archeological prospections. It is believed that the conclusions of this thesis can be translated to solutions for soil mapping problems on a landscape and regional scale. It is hoped that this work can contribute to policy measures and incentives for proximal soil sensor based inventories.

Bibliography

- Abdu, H., Robinson, D. A., and Jones, S. (2007). Comparing bulk soil electrical conductivity determination using the DUALEM-1S and EM38-DD electromagnetic induction instruments. *Soil Science Society of America Journal*, 71(1):189–196.
- Adamchuk, V. I., Hummel, J. W., Morgan, M. T., and Upadhyaya, S. K. (2004). On-the-go soil sensors for precision agriculture. *Computers and Electronics in Agriculture*, 44(1):71–91.
- Adamchuk, V. I., Lund, E. D., Reed, T. M., and Ferguson, R. B. (2007). Evaluation of an on-the-go technology for soil pH mapping. *Precision Agriculture*, 8(3):139–149.
- 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. *Computers and Electronics in Agriculture*, 48(3):272–294.
- Adamchuk, V. I., Morgan, M. T., and Ess, D. R. (1999). An automated sampling system for measuring soil pH. *Transactions of the Asae*, 42(4):885–891.
- Adamchuk, V. I., Morgan, M. T., and Sumali, H. (2001). Application of a strain gauge array to estimate soil mechanical impedance on-the-go. *Transactions of the Asae*, 44(6):1377–1383.
- Addiscott, T. M. (1996). Fertilizers and nitrate leaching. *Issues in Environmental Sciences*, 5(1):1–26.
- Addiscott, T. M., Withmore, A. P., and Powlson, D. . (1991). *Farming, fertilizers and the nitrate problem*. CAB international, Wallingford, UK.
- Allred, B. J., Ehsani, M. R., and Saraswat, D. (2006). Comparison of electromagnetic induction, capacitively-coupled resistivity, and galvanic contact resistivity methods for soil electrical conductivity measurement. *Applied Engineering in Agriculture*, 22(2):215–230.

- Allred, B. J., Eshani, M., and Daniels, J. J. (2008). General considerations for geophysical methods applied to agriculture. In Allred, B. J., Daniels, J. J., and Eshani, M. R., editors, *Handbook of Agricultural Geophysics*, pages 3–16. CRC Press, Boca Raton, Florida.
- Aubert, M., Hedde, M., Decaens, T., Margerie, P., Alard, D., and Bureau, F. (2005). Factors controlling soil macrofauna spatial pattern in a pure beech and a mixed beech-hornbeam forest. *Comptes Rendus Biologies*, 328(1):57–74.
- Baeyens, L., Van Slycken, J., and Stevens, D. (1993). *Description of the soil profile in Brasschaat*. Internal research paper, Institute of Forestry and Game Management, Geraardsbergen, Belgium.
- Barnes, E. M. and Baker, M. G. (2000). Multispectral data for mapping soil texture: Possibilities and limitations. *Applied Engineering in Agriculture*, 16(6):731–741.
- Barot, S., Rossi, J. P., and Lavelle, P. (2007). Self-organization in a simple consumer-resource system, the example of earthworms. *Soil Biology and Biochemistry*, 39(9):2230–2240.
- Barrow, N. J. (1967). Some aspects of the effects of grazing on the nutrition of pastures. *Journal of the Australian Institute of Agricultural Science*, 33:254–262.
- Benayas, J. M. R., Sanchez-Colomer, M. G., and Escudero, A. (2004). Landscape- and field-scale control of spatial variation of soil properties in Mediterranean montane meadows. *Biogeochemistry*, 69(2):207–225.
- Bezdek, J. C. (1981). Pattern recognition with fuzzy objective function algorithms. Plenum Press, New York.
- Birrell, S. J. and Hummel, J. W. (2001). Real-time multi ISFET/FIA soil analysis system with automatic sample extraction. *Computers and Electronics in Agriculture*, 32(1):45–67.
- Bishop, C. M. (1997). *Neural networks for pattern recognition (fourth edition)*. Oxford University Press, Oxford.
- Blackshaw, R. P., Donovan, S. E., Hazarika, S., Bol, R., and Dixon, E. R. (2007). Earthworm responses to long term agricultural management practices: Spatial relationships with soil properties. *European Journal of Soil Biology*, 43:S171–S175.
- Bogaert, N., Salomez, J., Vermoesen, A., Hofman, G., Van Cleemput, O., and Van Meirvenne, M. (2000). Within-field variability of mineral nitrogen in grassland. *Biology and Fertility of Soils*, 32(3):186–193.

- Boike, J. and Yoshikawa, K. (2003). Mapping of periglacial geomorphology using kite/balloon aerial photography. *Permafrost and Periglacial Processes*, 14(1):81–85.
- Bork, E. W., West, N. E., Doolittle, J. A., and Boettinger, J. L. (1998). Soil depth assessment of sagebrush grazing treatments using electromagnetic induction. *Journal of Range Management*, 51(4):469–474.
- Bouché, B. M. (1977). Stratégies lombriciennes. In Lohm, V. and Persson, P., editors, *Soil Organisms as Components of Ecosystems*, volume 25, pages 122–132. Ecological Bulletin, Stockholm, Sweden.
- Bouma, J. and Finke, P. A. (1993). Origin and nature of soil resource variability. In Robert, P., Rust, R., and Larson, W., editors, *Proceedings of soil specific crop management*, pages 3–13. ASA - CSSA - SSSA, Madison, Wisconsin.
- Bradley, A. P. (1997). The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7):1145–1159.
- Bravo, F. and Montero, G. (2001). Site index estimation in Scots pine (*Pinus sylvestris* L.) stands in the High Ebro Basin (northern Spain) using soil attributes. *Forestry*, 74(4):395–406.
- Brennan, R. L. and Prediger, D. J. (1981). Coefficient kappa - Some uses, misuses, and alternatives. *Educational and Psychological Measurement*, 41(3):687–699.
- Brevik, E. C., Fenton, T. E., and Horton, R. (2004). Effect of daily soil temperature fluctuations on soil electrical conductivity as measured with the Geonics EM-38. *Precision Agriculture*, 5:145–152.
- Brus, D. J., Knotters, M., Vandooremolen, W. A., Vankernebeek, P., and Vanseeters, R. J. M. (1992). The use of electromagnetic measurements of apparent soil electrical-conductivity to predict the boulder clay depth. *Geoderma*, 55(1-2):79–93.
- Buchleiter, G. W. and Farahani, H. (2002). *Comparison of electrical conductivity measurements from two different sensing technologies*. Paper No. 02-1056. ASAE, St. Joseph, Michigan.
- Burnham, K. P. and Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. Springer-Verlag, New York.
- Burrough, P. A. (1983). Multiscale sources of spatial variation in soil. 1. The application of fractal concepts to nested levels of soil variation. *Journal of Soil Science*, 34(3):577–597.

- Burrough, P. A. (1993). Soil variability: a late 20th century view. *Soil and Fertilizers*, 56:529–562.
- Burrough, P. A., Bouma, J., and Yates, S. R. (1994). The state-of-the-art in pedometrics. *Geoderma*, 62(1-3):311–326.
- Burrough, P. A., van Gaans, P. F. M., and Hootsmans, R. (1997). Continuous classification in soil survey: Spatial correlation, confusion and boundaries. *Geoderma*, 77(2-4):115–135.
- Bushnell, T. M. (1932). A new technique in soil mapping. *American soil survey association bulletin*, 13:74–81.
- Butnor, J. R., Doolittle, J. A., Johnsen, K. H., Samuelson, L., Stokes, T., and Kress, L. (2003). Utility of ground-penetrating radar as a root biomass survey tool in forest systems. *Soil Science Society of America Journal*, 67(5):1607–1615.
- Byrt, T., Bishop, J., and Carlin, J. B. (1993). Bias, prevalence and kappa. *Journal of Clinical Epidemiology*, 46(5):423–429.
- Cambardella, C. A., Moorman, T. B., Novak, J. M., Parkin, T. B., Karlen, D. L., Turco, R. F., and Konopka, A. E. (1994). Field-scale variability of soil properties in central Iowa soils. *Soil Science Society of America Journal*, 58:1501–1511.
- Cannavacciuolo, M., Bellido, A., Cluzeau, D., Gascuel, C., and Trehen, P. (1998). A geostatistical approach to the study of earthworm distribution in grassland. *Applied Soil Ecology*, 9(1-3):345–349.
- Chen, F., Kissel, D. E., West, L. T., and Adkins, W. (2000). Field-scale mapping of surface soil organic carbon using remotely sensed imagery. *Soil Science Society of America Journal*, 64(2):746–753.
- Chen, F., Kissel, D. E., West, L. T., Adkins, W., Clark, R., Rickman, D., and Luval, J. C. (2004). Field scale mapping of surface clay concentration. *Precision Agriculture*, 5:7–26.
- Clement, B. R. and Stombaugh, T. S. (2000). *Continuously measuring soil compaction sensor development*. Paper. No. 00-1041. ASAE, St. Joseph, Michigan.
- Cline, M. G. (1949). Basic principles of soil classifications. *Soil Science*, 2:81–91.
- Cockx, L., Van Meirvenne, M., and De Vos, B. (2007). Using the EM38DD soil sensor to delineate clay lenses in a sandy forest soil. *Soil Science Society of America Journal*, 71(4):1314–1322.

- Cockx, L., Van Meirvenne, M., and Hofman, G. (2005). Characterization of nitrogen dynamics in a pasture soil by electromagnetic induction. *Biology and Fertility of Soils*, 42:24–30.
- Cockx, L., Van Meirvenne, M., and Vitharana, U. W. A. (2008). Extracting topsoil information from high resolution EM38DD sensor data using contextual neural networks. In *Proceedings of the first global workshop on High Resolution Digital Soil Sensing and Mapping*, volume I, page 8.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46.
- Congalton, R. and Green, K. (1999). *Assessing the accuracy of remotely sensed data: Principles and practices*. Lewis Publisher, New York.
- Cook, S. E., Corner, R. J., Grealish, G., Gessler, P. E., and Chartres, C. J. (1996). A rule-based system to map soil properties. *Soil Science Society of America Journal*, 60(6):1893–1900.
- Corwin, D. and Lesch, S. M. (2003). Application of soil electrical conductivity to precision agriculture: Theory, principles, and guidelines. *Agronomy Journal*, 95(3):455–471.
- Corwin, D. L. (2005). Geospatial measurements of apparent soil electrical conductivity for characterizing soil spatial variability. In J., A. and Munoz-Carpena, R., editors, *Soil-Water-Solute Process Characterization*, pages 639–672. CRC Press, Boca Raton, FL.
- Corwin, D. L., Kaffka, S. R., Hopmans, J. W., Mori, Y., van Groenigen, J. W., van Kessel, C., Lesch, S. M., and Oster, J. D. (2003). Assessment and field-scale mapping of soil quality properties of a saline-sodic soil. *Geoderma*, 114(3-4):231–259.
- Corwin, D. L. and Lesch, S. M. (2005a). Apparent soil electrical conductivity measurements in agriculture. *Computers and Electronics in Agriculture*, 46(1-3):11–43.
- Corwin, D. L. and Lesch, S. M. (2005b). Characterizing soil spatial variability with apparent soil electrical conductivity part I. survey protocols. *Computers and Electronics in Agriculture*, 46(1-3):103–133.
- Cressie, N. A. C. (1991). *Statistics for spatial data*. page 900. New York.
- Dabas, M. and Tabbagh, A. (2003). Comparison of EMI and DC methods for soil mapping in precision agriculture. In Stafford, J. and Werner, A., editors, *Precision Agriculture*, pages 121–127. Wageningen Academic Publishers, Wageningen, The Netherlands.

- Dai, X. and Khorram, S. (1999). Data fusion using artificial neural networks: a case study on multitemporal change analysis. *Computers, Environment and Urban Systems*, 23(1):19–31.
- Dalsted, K., Paris, J. F., Clay, D. E., Reese, C. L., and Chang, J. (2003). Selecting the appropriate satellite remote sensing product for precision agriculture. In Potash and Phosphate Institute, editor, *Site-specific management guidelines No. 40*, page 6.
- Dansart, A. M., Bahr, J. M., and Attig, J. W. (1999). Using ground-penetrating radar to map fossil ice wedges that are preferential paths for leaching to groundwater. *Geological Society of America Abstracts with Programs*, 31:A–76.
- Davis, J. L. and Annan, A. P. (2002). Ground penetrating radar to measure soil water content. In Dane, J. H. and Topp, G. C., editors, *Methods of soil analysis. Part 4 - Physical Methods. Soil Sci. Soc. Am. Book Ser. 5.*, pages 446–463. SSSA, Madison, Wisconsin.
- Decaens, T., Bureau, F., and Margerie, P. (2003). Earthworm communities in a wet agricultural landscape of the Seine Valley (Upper Normandy, France). *Pedobiologia*, 47(5-6):479–489.
- Demas, G. P. and Rabenhorst, M. C. (2001). Factors of subaqueous soil formation: a system of quantitative pedology for submersed environments. *Geoderma*, 102(3-4):189–204.
- Deutsch, C. V. and Journel, A. G. (1998). *GSLIB: Geostatistical software library and user's guide. 2nd edition*. Applied Geostatistics Series. Oxford University Press, New York.
- Doolittle, J. A. and Collins, M. E. (1998). A comparison of EMinduction and GPR methods in areas of karst. *Geoderma*, 85(1):83–102.
- Doolittle, J. A., Indorante, S. J., Potter, D. K., Hefner, S. G., and McCauley, W. M. (2002). Comparing three geophysical tools for locating sand blows in alluvial soils of southeast Missouri. *Journal of Soil and Water Conservation*, 57(3):175–182.
- Doolittle, J. A., Sudduth, K. A., Kitchen, N. R., and Indorante, S. J. (1994). Estimating depths to claypans using electromagnetic induction methods. *Journal of Soil and Water Conservation*, 49(6):572–575.
- D'Urso, G. and Minacapilli, M. (2006). A semi-empirical approach for surface soil water content estimation from radar data without a-priori information on surface roughness. *Journal of Hydrology*, 321(1-4):297–310.

- Earl, R., Taylor, J. C., Wood, G. A., Bradley, I., James, I. T., Waine, T., Welsh, J. P., Godwin, R. J., and Knight, S. M. (2003). Soil factors and their influence on within-field crop variability, part 1: Field observation of soil variation. *Biosystems Engineering*, 84(4):425–440.
- Eigenberg, R. A., Doran, J. W., Nienaber, J. A., Ferguson, R. B., and Woodbury, B. L. (2002). Electrical conductivity monitoring of soil condition and available N with animal manure and a cover crop. *Agriculture, Ecosystems and Environment*, 88:183–193.
- Eigenberg, R. A., Korthals, R. L., and Nienaber, J. A. (1998). Geophysical electromagnetic survey methods applied to agricultural waste sites. *Journal of Environmental Quality*, 27(1):215–219.
- Ettema, C. H. and Wardle, D. A. (2002). Spatial soil ecology. *Trends in Ecology and Evolution*, 17(4):177–183.
- Evans, R. (1990). Crop patterns recorded on aerial photographs of England and Wales - Their type, extent and agricultural implications. *Journal of Agricultural Science*, 115:369–382.
- Fielding, A. H. and Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24(1):38–49. MAR.
- Finke, P. A., Brus, D. J., Bierkens, M. F. P., Hoogland, T., Knotters, M., and de Vries, F. (2004). Mapping groundwater dynamics using multiple sources of exhaustive high resolution data. *Geoderma*, 123(1-2):23–39.
- Foody, G. M. (1992). On the compensation for chance agreement in image classification accuracy assessment. *Photogrammetric Engineering and Remote Sensing*, 58(10):1459–1460.
- Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80(1):185–201.
- French, H. M. (1996). *The Periglacial Environment - 2nd edition*. Addison Wesley Longman, Harlow, UK.
- Frogbrook, Z. L. (1999). The effect of sampling intensity on the reliability of predictions and maps of soil properties. In Stafford, J. V., editor, *2nd European conference on precision agriculture*, Denmark. Silsoe Research Institute.
- Frogbrook, Z. L. and Oliver, M. A. (2007). Identifying management zones in agricultural fields using spatially constrained classification of soil and ancillary data. *Soil Use and Management*, 23(1):40–51.

- Gaines, T. P. and Gaines, S. T. (1994). Soil texture effect on nitrate leaching in soil percolates. *Communications in Soil Science and Plant Analysis*, 25(13-14):2561–2570.
- Garrett, M. K., Watson, C. J., Jordan, C., J., S. R. W., and Smith, R. V. (1992). The nitrogen economy of grazed grassland. In *The Fertilizer Society Proceedings*, volume 326, pages 1–32.
- Gee, G. W. and Bauder, J. W. (1986). Particle size analysis. In Klute, A., editor, *Methods of soil analysis. Part 1 - Physical and Mineralogical Methods. Agronomy Monograph no. 9 (second edition)*, pages 383–411. Madison, WI.
- Geonics Limited (2001). EM38-DD Ground conductivity meter. Dual Dipole Version. Operating Manual. Technical report.
- Ghysels, G. and Heyse, I. (2006). Composite-wedge pseudomorphs in Flanders, Belgium. *Permafrost and Periglacial Processes*, 17(2):145–161.
- Gish, T. J., Dulaney, W. P., Kung, K. J. S., Daughtry, C. S. T., Doolittle, J. A., and Miller, P. T. (2002). Evaluating use of ground-penetrating radar for identifying subsurface flow pathways. *Soil Science Society of America Journal*, 66(5):1620–1629.
- Godwin, R. J. and Miller, P. C. H. (2003). A review of the technologies for mapping within-field variability. *Biosystems Engineering*, 84(4):393–407.
- Gomez, A., Powers, R. F., Singer, M. J., and Horwath, W. R. (2002). Soil compaction effects on growth of young ponderosa pine following litter removal in california’s sierra nevada. *Soil Science Society of America Journal*, 66(4):1334–1343.
- Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York.
- Govers, G., Vandale, K., Desmet, P., Poesen, J., and Bunte, K. (1994). The role of tillage in soil redistribution on hillslopes. *European Journal of Soil Science*, 45(4):469–478.
- Greenhouse, J. P. and Morgan, A. V. (1977). Resistivity mapping of fossil permafrost patterns in southwestern ontario. *Canadian Journal of Earth Sciences*, 14(3):496–500.
- Hagen, A. (2003). Fuzzy set approach to assessing similarity of categorical maps. *International Journal of Geographical Information Science*, 17(3):235–249.

- Hall, G. F. and Olson, C. G. (1991). Predicting variability of soils from landscape models. In Mausbach, M. and Wilding, L., editors, *Spatial variabilities of soils and landforms. SSSA special publication No. 28*. SSSA, Madison, Winsconsin.
- Harry, D. G. and Gozdzik, J. S. (1988). Ice-wedges: growth, thaw transformation, and paleoenvironmental significance. *Journal of Quaternary Science*, 3:39–55.
- Haykin, S. (1998). *Neural networks: a comprehensive foundation (second edition)*. Prentice Hall, Upper Saddle River, NJ.
- Hecht-Nielsen, R. (1987). Kolmogorov's mapping neural existence theorem. In Caudill, M. and Butler, C., editors, *IEEE International conference on Neural Networks*, volume 3, pages 11–14, San Diego, CA.
- Heiniger, R. W., McBride, R. G., and Clay, D. E. (2003). Using soil electrical conductivity to improve nutrient management. *Agronomy Journal*, 95(3):508–519.
- Hemmat, A. and Adamchuk, V. I. (2008). Sensor systems for measuring soil compaction: Review and analysis. *Computers and Electronics in Agriculture*, 63:89–103.
- Hendrickx, J. M. H. and Kachanoski, R. G. (2002). Solute content and concentration - indirect measurement of solute concentration - nonintrusive electromagnetic induction. In Dane, J. H. and Topp, G. C., editors, *Methods of Soil Analysis, Part 4 - Physical Methods. Soil Sci. Soc. Am. Book Ser. 5.*, pages 1297–1306. Soil Science Society of America, Madison, Winsconsin.
- Hengl, T., Heuvelink, G. B. M., and Stein, A. (2004). A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*, 120(1-2):75–93.
- Hernández, P., Fernandez, R., Novo, M., Trigo, D., and Diaz Cosin, D. J. (2007). Geostatistical and multivariate analysis of the horizontal distribution of an earthworm community in el molar (madrid, spain). *Pedobiologia*, 51(1):13–21.
- Heyse, I. and Ghysels, G. (2003). Fossil polygonal periglacial structures in Flanders (Belgium). In Phillips, M., Springman, S. M., and Arenson, L. U., editors, *8th International Conference on Permafrost*, pages 395–400, Zurich, Switzerland. A a Balkema Publishers.
- Inman, D. J., Freeland, R. S., Ammons, J. T., and Yoder, R. E. (2002). Soil investigations using electromagnetic induction and ground-penetrating radar in southwest Tennessee. *Soil Science Society of America Journal*, 66(1):206–211.

- Isaaks, E. H. and Srivastava, R. M. (1989). *An introduction to applied geostatistics*. Oxford University Press, Oxford.
- Jenerette, G. D. and Wu, J. (2000). On the definitions of scale. *Bulletin of the Ecological Society of America*, 81:104–105.
- Jenny, H. (1941). *Factors of soil formation*. McGraw-Hill, New York.
- Jiménez, J.-J., Decaens, T., and Rossi, J.-P. (2006). Stability of the spatio-temporal distribution and niche overlap in neotropical earthworm assemblages. *Acta Oecologica*, 30(3):299–311.
- Johnson, C. K., Doran, J. W., Duke, H. R., Wienhold, B. J., Eskridge, K. M., and Shanahan, J. F. (2001). Field-scale electrical conductivity mapping for delineating soil condition. *Soil Science Society of America Journal*, 65(6):1829–1837.
- Johnson, C. K., Wienhold, B. J., Doran, J. W., Drijber, R. A., and Wright, S. F. (2004). Linking microbial-scale findings to farm-scale outcomes in a dryland cropping system. *Precision Agriculture*, 5:311–328.
- Jung, W. K., Kitchen, N. R., Sudduth, K. A., and Anderson, S. H. (2006). Spatial characteristics of claypan soil properties in an agricultural field. *Soil Science Society of America Journal*, 70(4):1387–1397.
- Kalra, Y. P. and Maynard, D. G. (1991). Methods manual for forest soil and plant analysis. In Region, F. C. N., editor, *Information report NOR-X-319*, pages 66–68. Northern Forestry Centre, Edmonton, Alberta.
- Kerry, R. and Oliver, M. A. (2003). Variograms of ancillary data to aid sampling for soil surveys. *Precision Agriculture*, 4(3):261–278.
- Kerry, R. and Oliver, M. A. (2007). Comparing sampling needs for variograms of soil properties computed by the method of moments and residual maximum likelihood. *Geoderma*, 140(4):383–396.
- Korsaeth, A. (2005). Soil apparent electrical conductivity (ECa) as a means of monitoring changes in soil inorganic N on heterogeneous morainic soils in SE Norway during two growing seasons. *Nutrient Cycling in Agroecosystems*, 72(3):213–227.
- Kung, K. J. S. and Donohue, S. V. (1991). Improved solute-sampling protocol in a sandy vadose zone using ground-penetrating radar. *Soil Science Society of America Journal*, 55(6):1543–1545.
- Lam, L. S. and Quattrochi, D. (1992). On the issues of scale, resolution, and fractal analysis in the mapping sciences. *Professional Geographer*, 44:88–89.

- Landis, J. . and Koch, G. G. (1977). Measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.
- Lark, R. M. (2005). Exploring scale-dependent correlation of soil properties by nested sampling. *European Journal of Soil Science*, 56(3):307–317.
- Lavelle, P., Decaens, T., Aubert, M., Barot, S., Blouin, M., Bureau, F., Margerie, P., Mora, P., and Rossi, J.-P. (2006). Soil invertebrates and ecosystem services. *European Journal of Soil Biology*, 42(Supplement 1):S3–S15.
- Lesch, S. M. and Corwin, D. L. (2003). Using the dual-pathway parallel conductance model to determine how different soil properties influence conductivity survey data. *Agronomy Journal*, 95(2):365–379.
- Lesch, S. M., Corwin, D. L., and Robinson, D. A. (2005). Apparent soil electrical conductivity mapping as an agricultural management tool in arid zone soils. *Computers and Electronics in Agriculture*, 46(1-3):351–378.
- Levin, S. A. (1992). The problem of pattern and scale in ecology. *Ecology*, 73(6):1943–1967.
- Levine, E. R., Kimes, D. S., and Sigillito, V. G. (1996). Classifying soil structure using neural networks. *Ecological Modelling*, 92(1):101–108.
- Lillesand, T. M. and Kiefer, R. W. (1995). *Remote sensing and image interpretation 3rd edition*. John Wiley & Sons, New York.
- Lin, H. S., Wheeler, D., Bell, J., and Wilding, L. (2005). Assessment of soil spatial variability at multiple scales. *Ecological Modelling*, 182(3-4):271–290.
- Logsdon, S. D., Perfect, E., and Tarquis, A. M. (2008). Multiscale soil investigations: Physical concepts and mathematical techniques. *Vadose Zone Journal*, 7(2):453–455.
- Lopez-Granados, F., Jurado-Exposito, M., Pena-Barragan, J. M., and Garcia-Torres, L. (2005). Using geostatistical and remote sensing approaches for mapping soil properties. *European Journal of Agronomy*, 23(3):279–289.
- Manel, S., Williams, H. C., and Ormerod, S. J. (2001). Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*, 38(5):921–931.
- Mankin, K. R. and Karthikeyan, R. (2002). Field assessment of saline seep remediation using electromagnetic induction. *Transactions of the Asae*, 45(1):99–107.

- Maselli, F., Conese, C., and Petkov, L. (1994). Use of probability entropy for the estimation and graphical representations of the accuracy of maximum-likelihood classifications. *ISPRS Journal of Photogrammetry and Remote Sensing*, 49(2):13–20.
- Matheron, G. (1973). The intrinsic random functions and their applications. *Advances in Applied Probability*, 5:439–468.
- McBratney, A. B. (1992). On variation, uncertainty and informatics in environmental soil-management. *Australian Journal of Soil Research*, 30(6):913–935.
- McBratney, A. B. and de Gruijter, J. J. (1992). A continuum approach to soil classification by modified fuzzy k-means with extragrades. *Journal of Soil Science*, 43(1):159–175.
- McBratney, A. B., Minasny, B., and Rossel, R. V. (2006). Spectral soil analysis and inference systems: A powerful combination for solving the soil data crisis. *Geoderma*, 136(1-2):272–278.
- McBratney, A. B., Odeh, I. O. A., Bishop, T. F. A., Dunbar, M. S., and Shatar, T. M. (2000). An overview of pedometric techniques for use in soil survey. *Geoderma*, 97(3-4):293–327.
- McBratney, A. B., Santos, M. L. M., and Minasny, B. (2003). On digital soil mapping. *Geoderma*, 117(1-2):3–52.
- McBride, R. A., Gordon, A. M., and Shrive, S. C. (1990). Estimating forest soil quality from terrain measurements of apparent electrical conductivity. *Soil Science Society of America Journal*, 54(1):290–293.
- McCutcheon, M. C., Farahani, H. J., Stednick, J. D., Buchleiter, G. W., and Green, T. R. (2006). Effect of soil water on apparent soil electrical conductivity and texture relationships in a dryland field. *Biosystems Engineering*, 94(1):19–32.
- McNeill, J. D. (1980a). Electrical conductivity of soils and rocks. In Geonics Limited, editor, *Technical Note TN-5*. Mississauga.
- McNeill, J. D. (1980b). Electromagnetic terrain conductivity measurement at low induction numbers. In Geonics Limited, editor, *Technical Note TN-6*. Mississauga.
- Meinardi, C. R., Beusen, A. H. W., Bollen, M. J. S., Klepper, O., and Willems, W. J. (1994). Vulnerability to diffuse pollution and average nitrate contamination of european soils and groundwater. In *Meeting on Modern Approaches to Water Management - Living with Water, at the International Conference on Integrated Water Resources Management*, pages 159–165, Amsterdam, Netherlands.

- Miao, Y., Mulla, D. J., and Robert, P. C. (2006). Spatial variability of soil properties, corn quality and yield in two Illinois, USA fields: implications for precision corn management. *Precision Agriculture*, 7(1):5–20.
- Minasny, B. and McBratney, A. B. (2002). FuzME version 3.0. Australian Center for Precision Agriculture, McMillan Building A05, The University of Sydney, Australia. <http://www.usyd.edu.au/su/agric/acpa>.
- Minasny, B. and McBratney, A. B. (2007). Spatial prediction of soil properties using EBLUP with the Matern covariance function. *Geoderma*, 140(4):324–336.
- Minasny, B., McBratney, A. B., Mendonca-Santos, M. L., Odeh, I. O. A., and Guyon, B. (2006). Prediction and digital mapping of soil carbon storage in the Lower Namoi Valley. *Australian Journal of Soil Research*, 44(3):233–244.
- Minasny, B., McBratney, A. B., and Whelan, B. M. (2002). VESPER version 1.62. Australian Center for Precision Agriculture, McMillan Building A05, The University of Sydney, Australia. <http://www.usyd.edu.au/su/agric/acpa>.
- 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 & Tillage Research*, 93(1):13–27.
- 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 & Tillage Research*, 86(2):218–229.
- Mueller, T. G., Hartsock, N. J., Stombaugh, T. S., Shearer, S. A., Cornelius, P. L., and Barnhisel, R. I. (2003). Soil electrical conductivity map variability in limestone soils overlain by loess. *Agronomy Journal*, 95(3):496–507.
- Murton, J. B. and French, H. M. (1993). Thaw modification of frost-fissure wedges, Richards Island, Pleistocene Mackenzie Delta, Western Arctic Canada. *Journal of Quaternary Science*, 8(3):185–196.
- Neter, J., Wasserman, W., and Kutner, M. (1996). *Applied Linear Statistical Models- 4th edition*. WCB McGraw-Hill, Boston, Massachusetts.
- Nuutinen, V., Pitkänen, J., Kuusela, E., Widbom, T., and Lohilahti, H. (1998). Spatial variation of an earthworm community related to soil properties and yield in a grass-clover field. *Applied Soil Ecology*, 8:85–94.
- Odeh, I. O. A., Mcbratney, A. B., and Chittleborough, D. J. (1992). Soil pattern-recognition with fuzzy-c-means - Application to classification and soil-landform interrelationships. *Soil Science Society of America Journal*, 56(2):505–516.

- Ogg, C. M., Edmonds, W. J., and Baker, J. C. (2000). Statistical verification of soil discontinuities in Virginia. *Soil Science*, 165(2):170–183.
- Oliver, M. A. (1987). Geostatistics and its application to soil science. *Soil Use and Management*, 3:8–20.
- Oliver, M. A. (1999). Exploring soil spatial variation geostatistically. In Stafford, J. V., editor, *2nd European conference on precision agriculture*, pages 3–18, Denmark. Silsoe Research Institute.
- Orion Research (1991). Model 93-07 nitrate electrode instruction manual. Orion Research, Boston. Technical report.
- Oyama, M. and Takehara, H. (1970). *Revised Standard Soil Color Charts*. Research Council for Agriculture, Forestry and Fisheries, Japan.
- Ozesmi, S. L., Tan, C. O., and Ozesmi, U. (2006). Methodological issues in building, training, and testing artificial neural networks in ecological applications. *Ecological Modelling*, 195(1-2):83–93.
- Pachepsky, Y. A., Timlin, D., and Varallyay, G. (1996). Artificial neural networks to estimate soil water retention from easily measurable data. *Soil Science Society of America Journal*, 60(3):727–733.
- Panissod, C., Dabas, M., Jolivet, A., and Tabbagh, A. (1997). A novel mobile multipole system (MUCEP) for shallow (0-3 m) geoelectrical investigation: the 'Vol-de-canards' array. *Geophysical Prospecting*, 45(6):983–1002.
- Pannatier, Y. (1996). Variowin. Springer-Verlag, New York.
- Parasuraman, K., Elshorbagy, A., and Si, B. C. (2006). Estimating saturated hydraulic conductivity in spatially variable fields using neural network ensembles. *Soil Science Society of America Journal*, 70(6):1851–1859.
- Park, S. J. and Vlek, P. L. G. (2002). Environmental correlation of three-dimensional soil spatial variability: a comparison of three adaptive techniques. *Geoderma*, 109(1-2):117–140.
- Peng, W. L., Wheeler, D. B., Bell, J. C., and Krusemark, M. G. (2003). Delineating patterns of soil drainage class on bare soils using remote sensing analyses. *Geoderma*, 115(3-4):261–279.
- Perlich, C., Provost, F., and Simonoff, J. S. (2004). Tree induction vs. logistic regression: A learning-curve analysis. *Journal of Machine Learning Research*, 4(2):211–255.

- Petersen, R. G. and Calvin, L. D. (1986). Sampling. In Klute, A., editor, *Methods of soil analysis, Part 1 - Physical and Mineralogical Methods. Agronomy Monograph no. 9 (Second Edition)*, pages 33–52. ASA - SSSA, Madison, Wisconsin.
- Pierce, F. J. and Nowak, P. (1999). Aspects of precision agriculture. In *Advances in Agronomy, Vol 67*, volume 67 of *Advances in Agronomy*, pages 1–85.
- Poier, K. R. and Richter, J. (1992). Spatial distribution of earthworms and soil properties in an arable loess soil. *Soil Biology and Biochemistry*, 24:1601–1608.
- Pontius, R. G. (2000). Quantification error versus location error in comparison of categorical maps. *Photogrammetric Engineering and Remote Sensing*, 66(8):1011–1016.
- Pontius, R. G. and Schneider, L. C. (2001). Land-cover change model validation by an roc method for the Ipswich watershed, Massachusetts, USA. *Agriculture Ecosystems and Environment*, 85(1-3):239–248.
- Power, C., Simms, A., and White, R. (2001). Hierarchical fuzzy pattern matching for the regional comparison of land use maps. *International Journal of Geographical Information Science*, 15:77–100.
- Reynolds, J. M. (1997). *An introduction to applied and environmental geophysics*. John Wiley and sons, New York.
- Rhoades, J. D., Manteghi, N. A., Shouse, P. J., and Alves, W. J. (1989). Soil electrical-conductivity and soil-salinity - new formulations and calibrations. *Soil Science Society of America Journal*, 53(2):433–439.
- Rhoades, J. D., Raats, P. A. C., and Prather, R. J. (1976). Effects of liquid-phase electrical-conductivity, water-content, and surface conductivity on bulk soil electrical-conductivity. *Soil Science Society of America Journal*, 40(5):651–655.
- Rhoades, J. J., Chanduvi, F., and Lesch, S. M. (1999). *Soil salinity assessment: methods and interpretation of electrical conductivity measurements*. FAO Irrigation and Drainage Paper 57. Food and Agricultural Organization of the United Nations, Rome.
- Robinson, D. A., Lebron, I., Lesch, S. M., and Shouse, P. (2004). Minimizing drift in electrical conductivity measurements in high temperature environments using the EM-38. *Soil Science Society of America Journal*, 68(2):339–345.
- Rossi, J.-P., Lavelle, P., and Albrecht, A. (1997). Relationships between spatial pattern of the endogeic earthworm *polypheretima elongata* and soil heterogeneity. *Soil Biology & Biochemistry*, 29:485–488.

- Rossiter, D. G. (2005). Digital soil mapping: Towards a multiple-use soil information system. *Análisis Geográficos*, 32(1):7–15.
- Roubens, M. (1982). Fuzzy clustering algorithms and their cluster validity. *European Journal of Operational Research*, 10(3):294–301.
- Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning internal representations by error propagation. In Rumelhart, D. and McClelland, J., editors, *Parallel distributed processing*, volume 1, pages 318–362. MIT Press, Cambridge, M.A.
- Sabatier, J. M., Hess, H., Arnott, W. P., Attenborough, K., Romkens, M. J. M., and Grissinger, E. H. (1990). Insitu measurements of soil physical-properties by acoustical techniques. *Soil Science Society of America Journal*, 54(3):658–672.
- Samouelian, A., Cousin, I., Tabbagh, A., Bruand, A., and Richard, G. (2005). Electrical resistivity survey in soil science: a review. *Soil & Tillage Research*, 83(2):173–193.
- Schaap, M. G., Leij, F. J., and van Genuchten, M. T. (1998). Neural network analysis for hierarchical prediction of soil hydraulic properties. *Soil Science Society of America Journal*, 62(4):847–855.
- Schoenholtz, S. H., Van Miegroet, H., and Burger, J. A. (2000). A review of chemical and physical properties as indicators of forest soil quality: challenges and opportunities. *Forest Ecology and Management*, 138(1-3):335–356.
- Scull, P., Franklin, J., Chadwick, O. A., and McArthur, D. (2003). Predictive soil mapping: a review. *Progress in Physical Geography*, 27(2):171–197.
- Sethuramasamyraja, B., Adamchuk, V. I., Dobermann, A., Marx, D. B., Jones, D. D., and Meyer, G. E. (2008). Agitated soil measurement method for integrated on-the-go mapping of soil pH, potassium and nitrate contents. *Computers and Electronics in Agriculture*, 60(2):212–225.
- Sheets, K. R. and Hendrickx, J. M. H. (1995). Noninvasive soil-water content measurement using electromagnetic induction. *Water Resources Research*, 31(10):2401–2409.
- Simpson, D., Lehouck, A., Van Meirvenne, M., Bourgeois, J., Thoen, E., and Vervloet, J. (2008). Geoarchaeological prospection of a medieval manor in the dutch polders using an electromagnetic induction sensor in combination with soil augerings. *Geoarchaeology: An International Journal*, 23(2):305–319.
- Sims, R. W. and Gerard, B. M., editors (1999). *Earthworms*. Synopses of the British Fauna (New Series). Published for The Linnean Society of London

- and The Estuarine and Coastal Sciences Association by Field Studies Council, Shrewsbury.
- Slavich, P. G. (1990). Determining ECa-depth profiles from electromagnetic induction. *Australian Journal of Soil Research*, 28(3):443–452.
- Smith, J. L. and Doran, J. W. (1996). Measurement and use of pH and electrical conductivity for soil quality analysis. In Doran, J. W. and Jones, A. J., editors, *Methods for assessing soil quality*, pages 169–185. Soil, Sci. Soc. Am. Spec. Publ. 49, Madison, WI.
- Soil Survey Staff (1999). Soil taxonomy, A basic system of soil classification for making and interpreting soil surveys, 2nd edition. In USDA-NRCS, editor, *Agricultural Handbook*, volume 436. US Government Printing Office, Washington, DC, USA.
- Sommer, M. (2006). Influence of soil pattern on matter transport in and from terrestrial biogeosystems - a new concept for landscape pedology. *Geoderma*, 133(1-2):107–123.
- Stevenson, F. (1982). Origin and distribution of nitrogen in soil. In Stevenson, F., editor, *Nitrogen in agricultural soils*, pages 1 – 42. American Society of Agronomy, Madison, Wisconsin.
- Sudduth, K. A., Drummond, S., and Kitchen, N. R. (2001). Accuracy issues in electromagnetic induction sensing of soil electrical conductivity for precision agriculture. *Computers and Electronics in Agriculture*, 31(3):239–264.
- Sudduth, K. A. and Hummel, J. W. (1991). Evaluation of reflectance methods for soil organic-matter sensing. *Transactions of the Asae*, 34(4):1900–1909.
- Sudduth, K. A., Kitchen, N. R., Bollero, G. A., Bullock, D. G., and Wiebold, W. J. (2003a). Comparison of electromagnetic induction and direct sensing of soil electrical conductivity. *Agronomy Journal*, 95(3):472–482.
- Sudduth, K. A., Kitchen, N. R., Bollero, G. A., Bullock, D. G., and Wiebold, W. J. (2003b). Comparison of electromagnetic induction and direct sensing of soil electrical conductivity. *Agronomy Journal*, 95(3):472–482.
- Sullivan, D. G., Shaw, J. N., and Rickman, D. (2005). Ikonos imagery to estimate surface soil property variability in two Alabama physiographies. *Soil Science Society of America Journal*, 69(6):1789–1798.
- Svensson, H. (1972). Use of stress situations in vegetation for detecting ground conditions on aerial photographs. *Photogrammetria*, 28(3):75–87.

- Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240(4857):1285–1293.
- Sy, N. L. (2006). Modelling the infiltration process with a multi-layer perceptron artificial neural network. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, 51(1):3–20.
- Triantafilis, J. and Lesch, S. M. (2005). Mapping clay content variation using electromagnetic induction techniques. *Computers and Electronics in Agriculture*, 46(1-3):203–237.
- Triantafilis, J., Odeh, I. O. A., and McBratney, A. B. (2001). Five geostatistical models to predict soil salinity from electromagnetic induction data across irrigated cotton. *Soil Science Society of America Journal*, 65(3):869–878.
- Upchurch, D. R. and Edmonds, W. J. (1991). Statistical procedures for specific objectives. In Mausbach, M. J. and Wilding, L. P., editors, *Spatial variabilities of soils and landforms. SSSA special publication No. 28*. SSSA, Madison, Winsconsin.
- Usoltsev, V. A. and Vanclay, J. K. (1995). Stand biomass dynamics of pine plantations and natural forests on dry steppe in Kazakhstan. *Scandinavian Journal of Forest Research*, 10(4):305–312.
- Van Coillie, F. M. B., Verbeke, L. P. C., and De Wulf, R. R. (2007). Feature selection by genetic algorithms in object-based classification of IKONOS imagery for forest mapping in Flanders, Belgium. *Remote Sensing of Environment*, 110(4):476–487.
- Van Meirvenne, M. (2003). Is soil variability within the small fields of flanders structured enough to allow precision agriculture? *Precision Agriculture*, 4:193–201.
- Van Meirvenne, M., Hofman, G., Van Hove, J., and Van Ruymbeke, M. (1990). A continuous spatial characterization of textural fractions and caco3 content of the topsoil of the polder region of northwest east-flanders, belgium. *Soil Science*, 150(4):710–716.
- van Rijsbergen, C. J. (1979). *Information retrieval*. Butterworths, London.
- Vandenberghe, J. and Pissart, A. (1993). Permafrost changes in europe during the last glacial. *Permafrost and Periglacial Processes*, 4:121–135.
- Veris Technologies Inc. (2003). *Veris Mobile Sensor Platform. Publication BAC 28542/1/04*. Salina, Kansas.

- Verschoore, R., Pieters, J. G., Seps, T., Spriet, Y., and Vangeyte, J. (2003). Development of a sensor for continuous soil resistance measurement. In Stafford, J. V., editor, *4th European Conference on Precision Agriculture*, pages 689–695, Berlin, Germany.
- Viscara Rossel, R. A., Jeon, Y. S., Odeh, I. O. A., and McBratney, A. B. (2008). Using a legacy soil sample to develop a mid-IR spectral library. *Australian Journal of Soil Research*, 46(1):1–16.
- Viscara 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(1-2):59–75.
- Viscarra Rossel, R. A., Fouad, Y., and Walter, C. (2008). Using a digital camera to measure soil organic carbon and iron contents. *Biosystems Engineering*, 100(2):149–159.
- Viscarra Rossel, R. A. and McBratney, A. B. (1998). Laboratory evaluation of a proximal sensing technique for simultaneous measurement of soil clay and water content. *Geoderma*, 85(1):19–39.
- Viscarra Rossel, R. A., Taylor, H. J., and McBratney, A. B. (2007). Multivariate calibration of hyperspectral γ -ray energy spectra for proximal soil sensing. *European Journal of Soil Science*, 58(1):343–353.
- Visser, H. and de Nijs, T. (2006). The Map Comparison Kit. *Environmental Modelling & Software*, 21(3):346–358.
- Vitharana, U. W. A., Saey, T., Cockx, L., Simpson, D., Vermeersch, H., and Van Meirvenne, M. (2008). Upgrading a 1/20,000 soil map with an apparent electrical conductivity survey. *Geoderma*, 148(1):107–112.
- Vitharana, U. W. A., Van Meirvenne, M., Cockx, L., and Bourgeois, J. (2006). Identifying potential management zones in a layered soil using several sources of ancillary information. *Soil Use and Management*, 22(4):405–413.
- Wait, J. R. (1955). Mutual electromagnetic coupling of loops over a homogeneous ground. *Geophysics*, 20:630–637.
- Walker, J. P. and Houser, P. R. (2002). Evaluation of the OhmMapper instrument for soil moisture measurement. *Soil Science Society of America Journal*, 66(3):728–734.
- Walters, J. C. (1994). Ice-wedge casts and relict polygonal patterned-ground in north-east iowa, usa. *Permafrost and Periglacial Processes*, 5(4):269–281.

- Webster, R. (2007). Analysis of variance, inference, multiple comparisons and sampling effects in soil research (vol 58, pg 74, 2007). *European Journal of Soil Science*, 58(6):1549–1549.
- Webster, R. and Oliver, M. (2001). Geostatistics for environmental scientists. page 271. John Wiley and Sons, Chichester.
- Weller, U., Zipprich, M., Sommer, M., Castell, W. Z., and Wehrhan, M. (2007). Mapping clay content across boundaries at the landscape scale with electromagnetic induction. *Soil Science Society of America Journal*, 71:1740–1747.
- Whalen, J. K. (2004). Spatial and temporal distribution of earthworm patches in corn field, hayfield and forest systems of southwestern quebec, canada. *Applied Soil Ecology*, 27(2):143–151.
- Wilding, L. P. and Drees, L. R. (1978). Spatial variability: a pedologist's viewpoint. In *Diversity of soils in the tropics. American society of agronomy special publication No. 34*, pages 1–12. SSSA, Madison, Winsconsin.
- Wollenhaupt, N. C., Mulla, D. J., and Gotway Crawford, C. A. (1997). Soil sampling and interpolation techniques for mapping spatial variability of soil properties. In Pierce, F. J. and Sadler, E. J., editors, *The state of site specific management for agriculture*, pages 19–54. ASA-CSSA-SSSA, Madison.
- Woolery, M. E., Olson, K. R., Dawson, J. O., and Bollero, G. (2002). Using soil properties to predict forest productivity in southern Illinois. *Journal of Soil and Water Conservation*, 57(1):37–45.
- Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8:338–353.
- Zhai, Y. S., Thomasson, J. A., Boggess, J. E., and Sui, R. X. (2006). Soil texture classification with artificial neural networks operating on remote sensing data. *Computers and Electronics in Agriculture*, 54(2):53–68.
- Zhang, R. and Wienhold, B. J. (2002). The effect of soil moisture on mineral nitrogen, soil electrical conductivity, and pH. *Nutrient Cycling in Agroecosystems*, 63(2-3):251–254.

Curriculum vitae

Contact

Liesbet Cockx
Lentestraat 22
9000 Gent
Belgium
Phone.: +32 499 416527
Liesbet.Cockx@Ugent.be

Education

Bio-engineer in Land and Forest Management, July 2002
Katholieke Universiteit Leuven
Leuven, Belgium
Thesis: *Evaluation of the carbon balance of a forestry plantation in South Africa*. Promotor: Prof. Bart Muys
Diploma in Education (AILO), July 2006
Ghent University
Ghent, Belgium

Professional Experience

06/2003–06/2010 Assisting Academic Personnel
Department of Soil Management
Ghent University
Ghent, Belgium

Scientific Publications

International publications with peer review¹

Meerschman, E., Cockx, L. and M. Van Meirvenne. A geostatistical two-phase sampling strategy to map soil heavy metal concentrations in a former war zone. *European Journal of Soil Science*. (submitted, 2009)

¹The here mentioned journals have an ISI impact factor.

Cockx, L., Van Meirvenne, M., Vitharana, U.W.A., Verbeke, L.P.C., Simpson, D., Saey, T. and F.M.B. Van Coillie. 2009. Extracting topsoil information from EM38DD sensor data using a neural network approach. *Soil Science Society of America Journal*, 73(6): 1-8.

Mouazen, A. M. , Maleki, M. R., Cockx, L., Van Meirvenne, M., Van Holm, L. H. J., Merckx, R., De Baerdemaeker, J. and H. Ramon. 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 Research*, 103: 144-152.

Saey, T., Simpson, D., Vermeersch, H., Cockx, L. and Van Meirvenne M. 2009. Comparing the the EM38DD and DUALEM-21S sensors for depth-to-clay mapping. *Soil Science Society of America Journal*, 73: 7-12.

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

Simpson, D., Van Meirvenne, M., Saey, T., Vermeersch, H. , Bourgeois, J., Lehouck, A., Cockx, L. and U.W.A. Vitharana. 2009. Evaluating the multiple coil configurations of the EM38DD and the DUALEM-21S sensors to detect archeological anomalies. *Archaeological Prospection*, 16: 1-13.

Valckx, J., Cockx, L., Wauters, J., Van Meirvenne, M., Govers, G., Hermy, M. and B. Muys. 2009. Within-field spatial variability of earthworm populations related to species interactions and soil apparent electrical conductivity. *Applied Soil Ecology*, 41(3): 315-328.

Peters J., Verhoest, N.E.C., Samson, R., Van Meirvenne, M., Cockx, L. and B. De Baets. 2009. Uncertainty propagation in vegetation distribution models based on ensemble classifiers. *Ecological modelling*, 220: 791-804.

Vitharana, U.W.A, Van Meirvenne, M., Simpson, D. and L. Cockx. 2008. Agronomic consequences of potential management zones delineated on the basis of EM38DD measurements. *Near Surface Geophysics*, 6(5): 289-296.

Vitharana, U.W.A., Saey T., Cockx L., Simpson D. and M. Van Meirvenne. 2008. Upgrading a 1/20000 soil map with an apparent electrical conductivity survey. *Geoderma*, 148(1): 107-112.

Vitharana, U.W.A., Van Meirvenne, M., Simpson, D., Cockx, L. and J. De Baerdemaeker. 2008. Key soil and topographic properties to delineate potential management classes for precision agriculture in the European loess area. *Geoderma*, 143: 206-215.

Cockx, L., Van Meirvenne, M. and B. De Vos. 2007. Using the EM38DD soil sensor to delineate clay lenses in a sandy forest soil. *Soil Science Society of America Journal*, 71: 1314-1322.

Garcia-Quijano, J.F., Peters, J., Cockx, L., Van Wyk, G., Rosanov, A., Deckmyn, G., Ceulemans, R., Ward, S.M., Holden, N.M., Van Orshoven, J. and B. Muys. 2007. Carbon sequestration and environmental effects of afforestation with *Pinus radiata* D. Don in the Western Cape, South Africa. *Climatic Change*, 83: 323-355.

Vitharana, U.W.A., Van Meirvenne, M., Cockx L. and J. Bourgeois. 2006. Identifying potential management zones in a layered soil using multiple sources of ancillary information. *Soil Use and Management*, 22: 405-413.

Cockx, L., Ghysels, G., Van Meirvenne, M. and I. Heyse. 2006. Prospecting frost-wedge pseudomorphs and their polygonal network using the electromagnetic induction sensor EM38DD. *Permafrost and Periglacial Processes*, 17(2): 163-168.

Cockx, L., Van Meirvenne, M. and G. Hofman. 2005. Characterization of nitrogen dynamics in a pasture soil by electromagnetic induction. *Biology and Fertility of Soils*, 42(1): 24-30.

Conference and workshop proceedings

Cockx L., Van Meirvenne, M. and U.W.A., Vitharana 2008. Extracting top-soil information from high resolution EM38DD sensor data using contextual neural networks. In: R. Viscarra Rossel, First Global Workshop on High Resolution Digital Soil Sensing and Mapping, 5-8 February, 2008, Sydney, Australia, Volume I, 8 pgs.

Mouazen, A. M., Maleki, M. R., Cockx, L., Van Meirvenne, M., Van Holm, L. H. J., Merckx, R., De Baerdemaeker, J. and H. Ramon. 2008. Optimisation of sensor position for high quality of on-line collected soil spectra. In: R. Viscarra Rossel, First Global Workshop on High Resolution Digital Soil Sensing and Mapping, 5-8 February, 2008, Sydney, Australia, Volume II, 11 pgs.

Vitharana, U.W.A., Van Meirvenne, M., Amakor, X.N.C., Saey, T. and D. Simpson. 2007. Potency of proximal soil sensing to upgrade the soil map of Belgium: test case UGent experimental farm at Melle. Thematic Day 2007: Soil resources in Belgium - Current and future issues Soil Science Society of Belgium, 6th December 2007. Brussels, p. 10.

Vitharana U.W.A., Van Meirvenne, N, Amakor, X.N.C., Simpson, D. and L. Cockx. 2007. Evaluating two scales of polygon soil maps and an electrical

conductivity Survey in Characterizing Within-field Soil Textural Variability. Pedometrics 2007, 27-30 August 2007. Tübingen, Germany, p.41.

Van Meirvenne, M., Vitharana, U.W.A. and L. Cockx. 2006. Developments in soil sampling and proximal sensing with applications in precision agriculture. In: Langouche D. and Van Ranst E. (Eds.) *New Waves in Physical Land Resources*, Gent 3-9 September 2006, pp. 226-227, ISBN: 9789076769950.

Vitharana, U.W.A., Van Meirvenne, M., Simpson, D., Cockx, L. and J. De Baerdemaeker. 2006. Influence of the topography on the within-field soil variability in the Loess region in Belgium. In: *2nd Global Workshop on Digital Soil Mapping: book of abstracts*, Ed.: Mendonça-Santos M. et al., Rio de Janeiro, Brazil, 4-7 July 2006, p. 56. ISBN: 85-85864-21 (on CD).

Simpson, D., Van Meirvenne, M., De Baerdemaeker, J., Vitharana, U.W.A. and L. Cockx. 2006. Soil-crop-landscape relationships within an agricultural field in the loess region of Belgium. Day of the Young Soil Scientists, Belgian Soil Science Society, 22 February 2006. Brussels, p. 10.

Valckx, J., Cockx, L., Wauters, J. Van Meirvenne, M., Hermy, M. and B. Muys. 2006. Spatial variability of earthworm communities in relation to soil bulk electrical conductivity. Krakow, Poland, 4-9 september 2006, ISEE8 Abstracts Book, p 190.

Cockx, L., Van Meirvenne, M. and B. De Vos. 2006. Quantifying forest soil variability and its influence on tree diameter. *Spatial Data Methods for Environmental and Ecological Processes*. Foggia, Italy, 14-15 September 2006, CD-rom.

Van Meirvenne, M., Cockx, L. and Vitharana, U.W.A. 2005. Pedometrics in transition: from too few to too many data? *Pedometrics 2005: Frontiers in Pedometrics*, Naples, USA. 12-14 September 2005, Abstract book p. 81-82.

Cockx, L., Van Meirvenne, M. and G. Hofman. 2005. Within-field nitrogen management based on a non invasive soil sensor. In: L.Cockx, M. Van Meirvenne, T. Toth, G. Hofman and T. Nemeth (Eds.), *workshop proceedings Monitoring space-time dynamics of soil chemical properties to improve soil management and environmental quality*, Ghent, 8-9 December 2005, p. 71-82. ISBN: 9059890973.

Vitharana, U.W.A., Simpson, D., Van Meirvenne, M. and L. Cockx. 2005. Evaluation of multiple ancillary information to support potential management zones delineation: case study in the loess region of Belgium. In: L. Cockx, M. Van Meirvenne, T. Toth, G. Hofman and T. Nemeth (Eds.), *workshop proceedings Monitoring space-time dynamics of soil chemical properties to*

improve soil management and environmental quality, Ghent, 8-9 December 2005, p. 57-69. ISBN: 9059890973.

Vitharana, W.A.U., Van Meirvenne, M. and L. Cockx. 2005. Using secondary information sources to improve the within-field soil textural mapping in a layered alluvial soil. Day of the Young Soil Scientists, Belgian Soil Science Society, 23 February 2005. Brussels, p. 5.

Vitharana, W.A.U., Van Meirvenne, M. and L. Cockx. 2005. Using secondary information sources to improve the within-field soil textural mapping in a layered alluvial soil. In: Stafford J. (Ed.), *Precision Agriculture '05*. Wageningen Academic Publishers, pp. 425-432. ISBN: 9076998698.

Cockx, L., Ghysels, G., Van Meirvenne, M. and I. Heyse. 2005. Detecting frost-wedge pseudomorphs using electromagnetic induction. *Terra Nostra*, European Conference on Permafrost. Potsdam, Germany, 12-16 June 2005, p.10.

Cockx, L., Van Meirvenne, M. and G. Hofman. 2004. The use of electromagnetic induction in delineating nitrogen management zones. In: *Proceedings of the 7th International Conference on Precision Agriculture and Other Resources Management*, 25-28 July 2004, Minneapolis, USA, p. 468-480.

Other Publications

Simpson D., Van Meirvenne, M., De Baerdemaeker, J., Vitharana, U.W.A. and L. Cockx. 2006. Soil-crop-landscape relationships within an agricultural field in the loess region of Belgium. *Pedologia-Themata*, 13, Soil Science Society of Belgium.

Oral presentations to workshops and conferences

Cockx, L., Van Meirvenne, M. and U.W.A. Vitharana. A neural network approach to topsoil clay prediction using an EMI-based soil sensor. 1st global workshop on high resolution digital soil sensing and mapping, Sydney, Australia, 5-8 February 2008.

Cockx, L., Van Meirvenne, M. and B. De Vos. Using proximal soil sensing to quantify forest soil variability and its influence on tree diameter. International workshop: Spatial Data Methods for Environmental and Ecological Processes. Foggia, Italy, 14-15 September 2006.

Cockx, L., Ghysels, G., Van Meirvenne and M., Heyse, I. Detecting frost-wedge pseudomorphs using electromagnetic induction. EUCOP 2005 - 2nd European Conference on Permafrost. Potsdam, Germany, 12-16 June 2005.

Cockx, L., Van Meirvenne, M. and Hofman, G. Within-field nitrogen man-

agement based on a non-invasive soil sensor. Workshop on monitoring space-time dynamics of soil chemical properties to improve soil management and environmental quality. Ghent, Belgium, 8-9 December 2005.

Cockx, L., Van Meirvenne, M. and Hofman, G.. The use of electromagnetic Induction in delineating nitrogen management zones. 7th International Conference on Precision Agriculture, Minneapolis, Minnesota, USA, 25-28 July 2004.

Cockx, L., Van Meirvenne, M. and Hofman, G. Within-field nitrogen management supported by electromagnetic induction. Day of the Young Soil Scientists, Belgian Soil Science Society, Brussels, Belgium, 18 February 2004.

Cockx, L., Peters, J, and Muys, B. International conference: Interactions of forest carbon management and maintenance of biodiversity, Tuusula, Finland, 4-6 September 2002.

February 10, 2010

