ADVANCES IN PROXIMAL SOIL SENSING THROUGH INTEGRATED SYSTEMS APPROACH

by

Nandkishor Motiram Dhawale

Department of Bioresource Engineering

Macdonald Campus of McGill University

Montreal, Canada

April 2015

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of

Doctor of Philosophy

This dissertation is dedicated to my loving wife and dad.

Their support, encouragement, and constant love sustained me throughout my life.

हा प्रबंध माझ्या प्रेम ळ पत्नी आणि वडीलास, समर्पित आहे.

त्यांचे समर्थन, उत्तेजन आणि सतत प्रेम माझ्या आयुष्यात मला लाभले आहे.

ABSTRACT

Precision agriculture uses information technologies to improve the economic, environmental and social aspects of modern farming. Among other innovations, the ability to account for spatial variability in the crop growing environment has been the most challenging. The main limitation to the widespread adoption of these technologies has been the lack of economically feasible spatial data to reveal the heterogeneity of soil properties related to local fertilization needs. To increase the implementation of this technology, advanced equipment and methods to obtain high density measurements of the physical and chemical properties of soil at a relatively low cost are essential. The quality of data needed by decision-making algorithms must be improved; this needs the development of, an automated sensing platform for *in situ* soil measurements, similar to the rovers deployed on Mars. The project focuses on this challenge. The thesis addressed three key aspects of the development of the technology: 1) design of an automated on the-spot soil analyser, 2) evaluation of the potential for deploying different hyperspectral sensor systems, and 3) development of an algorithm to cluster multilayer high density spatial data to maximize the value of prescribed point-based measurements.

A new automated, on-the-spot soil analyser (OSA) was developed to perform subsurface measurements of soil chemical properties. It used an array of ion-selective electrodes and could simultaneously determine pH and nitrate, on-the-spot, in an automated manner. The system can be easily extended to use other sensors, such as active, hyperspectral sensors.

A set of field deployment-ready instruments for active hyperspectral measurements of soil reflectance were tested in terms of both their reproducibility (precision) and predictability (accuracy) of the measurements and evaluated in terms of their ability to predict soil particle size distribution, soil carbon, soil organic matter, and total soil phosphorous content. This research focused on instruments operating in the visible, near-infrared and mid-infrared parts of the electromagnetic spectrum.

Finally, a new algorithm was developed to process traditional high-density proximal soil sensing data, such as apparent soil electrical conductivity (ECa) and field elevation. Preliminary tests of

the algorithm using six production fields illustrated the algorithm's robustness when delineating field areas with different field elevations and ECa measurements. This algorithm has the potential to optimize the number of point-based measurements with OSA, other *in situ* sensing systems, or conventional soil sampling.

Six individual research activities resulted in six different manuscripts that constitute the main body of this thesis.

RÉSUMÉ

L'agriculture de précision utilise les technologies de l'information pour améliorer les aspects économiques, environnementaux et sociaux de l'agriculture moderne. Parmi différentes innovations, la capacité de prendre en considération la variabilité spatiale des conditions de croissances des cultures a été particulièrement ambitieuse. La principale contrainte à l'adoption massive de ces technologies a été le manque de données économiquement viables révélant l'hétérogénéité des propriétés du sol et se rapportant aux besoins locaux de fertilisants.

Des équipements et méthodes avancés servant à l'obtention de mesures de haute densité des propriétés physiques et chimiques du sol à des coûts relativement bas sont essentiels pour augmenter l'utilisation de la technologie. Pour améliorer la qualité des données requises par les algorithmes de prise de décision, une plateforme de détection automatisée, semblable aux robots déployés sur Mars, doit être développée pour mesurer le sol in situ. Le projet se concentre sur ce défi. La thèse a abordé trois aspects clés du développement de la technologie : 1) la conception d'un analyseur de sol sur place automatisé, 2) l'évaluation du potentiel de déploiement de différents systèmes de capteurs hyper-spectraux et 3) le développement d'un algorithme pour regrouper des données spatiales multicouches de haute densités pour maximiser la valeur des mesures par points recommandées.

Un nouvel analyser sur place (OSA) a été développé pour effectuer des mesures souterraines des propriétés chimiques du sol en utilisant un ensemble d'électrodes sélectives. Ces électrodes peuvent simultanément déterminer de manière automatisée l'acidité et nitrate. Le système peut aisément être élargi à d'autres capteurs, comme des capteurs hyper-spectraux actifs.

Un ensemble d'instruments de terrain prêts-à-déployer pour des mesures de réflectance du sol hyperspectrales a été testé en termes de répétabilité (précision) et de prévisibilité (exactitude) des mesures. Ces capteurs ont été évalués en termes de leur capacité à prédire la distribution de la taille des particules de sol, le carbone du sol, la matière organique et le contenu en phosphore extrait. Cette recherche était centrée sur des instruments fonctionnant dans le visible, l'infrarouge proche et l'infrarouge moyen du spectre électromagnétique.

Finalement, un nouvel algorithme a été développé pour utiliser des données traditionnelles de détection du sol à proximité, comme la conductivité électrique apparente du sol (ECa) et l'élévation des champs. Des tests préliminaires ont été effectué pour six terrains de production. Les résultats ont confirmé la robustesse de l'algorithme lors de la délimitation des zones de terrain avec différentes élévations et mesures ECa. Cet algorithme a le potentiel d'optimiser la valeur des mesures par points en utilisant l'OSA, d'autres systèmes de capteurs in situ, ou l'échantillonnage du sol.

Six activités individuelles de recherche ont résulté en six différents manuscrits qui constituent le corps de cette thèse.

ACKNOWLEDGEMENTS

As the author of this thesis I would like to acknowledge those who have provided help and support throughout this research. It is to them that I owe my deep gratitude.

This project is supported in part by funds provided through National Science and Engineering Research Council (NSERC) of Canada Discovery as well as Agriculture and Agri-Food Canada (AAFC) Agricultural Greenhouse Gases Program (AGGP) and Deere and Company grants using infrastructure supported by Canada Foundation for Innovation (CFI) Leader Opportunity Fund (LOF).

I express my deepest gratitude to my research advisor, Prof. Viacheslav I. Adamchuk, for his great instruction, caring, patience, and providing me with an excellent atmosphere for doing research. I appreciate the opportunity to work closely with him during the Ph.D. program to learn from his knowledge, expertise, problem-solving capability, and especially, his enthusiasm for pursuing research work. I always learnt from him, both from his academic knowledge and from his attitudes toward life.

On this occasion, I would like to extend my sincere thanks and gratitude to Professor Shiv O. Prasher. As a co-supervisor committee member, he constantly provided his technical guidance throughout my Ph.D. program at McGill University. I am also grateful to Dr. Ken Sudduth, Prof. Asim Biswas and Prof. Caroline Begg for being my dissertation examination committee members and I offer my appreciation to Prof. William Hendershot, Prof. Ian Adomowski and Prof. Vijaya Raghavan for being the committee members for my Ph.D. comprehensive examination and who provided suggestions on the preliminary work of this research.

I also acknowledge the valuable contributions from the co-authors of the manuscripts of this research, Dr. Raphael A. Viscarra Rossel, Prof. Ashraf A. Ismail, Prof. Joan K. Whalen, Prof. Pierre R.L. Dutilleul, Prof. Richard F. Ferguson, Jasmeen Kaur, and Frédéric René-Laforest. They contributed their knowledge and expertise in this research. They

were always kind and prompt in responding to my questions regarding my research and in reviewing my writing for publication. Their suggestions and comments helped me understand my research from different perspectives.

I am very thankful to all my colleagues and friends; Helene Lalande, Marc Samoisette, Sedigheh Zarayan, Sam Sotocinal, Scott Manktelow, Dr. Abdul Mannan Ehsan, Dr. Luan Pan, Dr.Wenjun Ji, Ahmad Matsu, Antoine Pouliot, Bharat Sudarsan, Christopher Nzediegwu, Evan Francisco Ruiz de la Macorra, Eman ElSayed, Evan Henry, Florian Reumont, Frédéric René-Laforest, Hsin-Hui Huang, Ian Chan, Jaskaran Dhiman, Jasmeen Kaur, Marine Louargant, Maria Mastorakos, Michael Saminsky, Muhammad Azhar Baig, Sanaz Alizadeh, Trevor Stanhope, Rahul Suresh, Sara Tawil, Yue Su, Yasmeen Hitti, and all of the precision agriculture group at McGill for their assistance and support on the research. Dr. Ashutosh Singh, Dr. Gopu Nair, Dhananjai Dhokne, and many-many other friends have always been a source of motivation and inspiration for me and a special thanks to Darlene Canning and Helen Fyles who helped proof reading and editing this dissertation.

I would also like to thank Prof. Rolf Wuthrich, the advisor for my Master's program, who opened my eyes to see the beauty of the world of applied research.

Finally, my sincere gratitude goes to the creator of this universe, my late mother, dear father, sisters, aunts and inlaws and to my beloved wife and son. Their unconditional love and support have always been my source of inspiration.

FORMAT OF THE THESIS

This thesis is submitted in the format of papers suitable for journal publication. This thesis format has been approved by the Faculty of Graduate and Postdoctoral Studies, McGill University, and follows the conditions outlined in the Guidelines: Concerning Thesis Preparation, which are as follows:

“As an alternative to the traditional thesis format, the dissertation can consist of a collection of papers of which the student is an author or co-author. These papers must have a cohesive, unitary character making them a report of a single program of research. The structure for the manuscript-based thesis must conform to the following:

Candidates have the option of including, as part of the thesis, the text of one or more papers submitted, or to be submitted, for publication, or the clearly duplicated text (not the reprints) of one or more published papers. These texts must conform to the "Guidelines for Thesis Preparation" with respect to font size, line spacing and margin sizes and must be bound together as an integral part of the thesis. (Reprints of published papers can be included in the appendices at the end of the thesis).

The thesis must be more than a collection of manuscripts. All components must be integrated into a cohesive unit with a logical progression from one chapter to the next. In order to ensure that the thesis has continuity, connecting texts that provide logical bridges between the different papers are mandatory.

The thesis must conform to all other requirements of the "Guidelines for Thesis Preparation" in addition to the manuscripts.

The thesis must include the following

(a) A table of contents;

(b) An abstract in English and French;

(c) An introduction which clearly states the rational and objectives of the research;

(d) A comprehensive review of the literature (in addition to that covered in the introduction to each paper);

(e) A final conclusion and summary;

As manuscripts for publication are frequently very concise documents, where appropriate, additional material must be provided (e.g., in appendices) in sufficient detail to allow a clear and precise judgment to be made of the importance and originality of the research reported in the thesis.

In general, when co-authored papers are included in a thesis the candidate must have made a substantial contribution to all papers included in the thesis. In addition, the candidate is required to make an explicit statement in the thesis as to who contributed to such work and to what extent. This statement should appear in a single section entitled "Contributions of Authors" as a preface to the thesis. The supervisor must attest to the accuracy of this statement at the doctoral oral defense. Since the task of the examiners is made more difficult in these cases, it is in the candidate's interest to clearly specify the responsibilities of all the authors of the co-authored papers".

CONTRIBUTIONS OF AUTHORS

The chapters of this thesis have been presented at scientific conferences and submitted for publication to peer reviewed journals. The author of this thesis was responsible for; design, development and field testing of the On-the-spot Soil Analyser (OSA), design, development and testing of the muti-dimensional spatial data clustering application and modeling and analysis of hyperspectral data. The author also took charge of the study and carried out other research and analytical work in both the field and the laboratory. He was responsible for the preparation of all manuscripts and presentations listed under this topic. Prof. Viacheslav I. Adamchuk is the thesis supervisor. He was actively involved in all stages of the study, providing scientific advice and technical supervision. He was also directly associated with editing and reviewing the prepared manuscripts. Prof. Shiv O. Prasher is his co-supervisor; he constantly provided scientific suggestions for this study.

Dr. Raphael A.Viscarra Rossel is pedrometrician from Bruce E. Butler laboratory, CSIRO Land and Water, Australia. He provided technical advice on hyperspectral data analysis. Prof. Ashraf A. Ismail is a professor in the Department of Food Science and Agriculture Chemistry at McGill University. He supported by providing access to various hyperspectral instruments. Prof. Joann K. Whalen is a professor in the Department of Natural Resource and Sciences at McGill University. She supported by providing archived soil samples. Prof. Pierre R.L. Dutilleul, a professor in the Department of Plant Science at McGill University, contributed with his research expertise on geostatistics at different stages of this study. Prof. Richard F. Ferguson a professor in the Department of Agronomy and Horticulture at University of Nebraska, USA. He supported by providing readily available proximal soil sensing data sets. Marine Louargant, is former summer-intern student from Instrumentation and Control Systems Laboratory, Macdonald Campus, McGill University. She supported by collecting soil spectral measurements using the visible/near-infrared spectrometer. Jasmeen Kaur is a collegue and a graduate student at McGill University, she supported by collecting data from the mid-infrared spectrometer. Frédéric René-Laforest is also my collegue and graduate student at McGill University. Beside data collection activities, he supported with Computer Aided Drawings (CAD) and fabrication tasks of the the On-the-spot Soil Analyser (OSA).

List of Publications and Scientific Presentations Related to the Thesis

A. Thesis Chapters Submitted for Publication in Peer Reviewed Journals

1. Dhawale, N.M., F. Rene-Laforest, V.I. Adamchuk, & S.O. Prasher. 2015. Development and evaluation of an automated On-the-spot Soil Analyser (OSA) for precision agriculture. *Computers and Electronics in Agriculture Journal* (to be submitted).

2. Dhawale, N.M., V.I. Adamchuk, S.O. Prasher, P.R.L. Dutilleul, & R.B. Ferguson. 2015. Spatially contrained data clustering for multilayer high-density proximal soil sensing data. *Geoderma* (to be submitted).

3. Dhawale, N.M., V.I. Adamchuk, S.O. Prasher, & R.A.Viscarra Rossel. 2015. Precision and accuracy of visible/near-infrared soil reflectance measurements. *Computers and Electronics in Agriculture Journal* (under review).

4. Dhawale, N.M., V.I. Adamchuk, R.A. Viscarra Rossel, S.O. Prasher, A.A. Ismail, & J.K. Whalen. 2015. Predicting plant available phosphorus using vis-NIR hyperspectral measurements obtained using a soil profiling tool. *Canadian Biosystems Engineering Journal* (under review).

5. Dhawale, N.M., V.I. Adamchuk, V.I., S.O. Prasher, R.A. Viscarra Rossel, A.A. Ismail, J.K. Whalen, & M. Louargant. 2015. A comparative study between a field deployment ready visible/near-infrared and prototyped mid-infrared hyperspectral instruments. *Journal of Near Infrared Spectroscopy* (under review).

6. Dhawale, N.M., V.I Adamchuk, S.O. Prasher, R.A. Viscarra Rossel, A.A. Ismail, & K. Jasmeen. 2015. Proximal soil sensing of soil texture and organic matter with a prototype portable mid-infrared spectrometer. *European Journal of Soil Science* (in press).

B. Refereed Conference Proceedings

7. Dhawale, N.M., V.I. Adamchuk, S. O. Prasher, P.R.L. Dutilleul, & R.B. Ferguson. 2014. Spatially constrained geospatial data clustering for multilayer sensor based measurements. *Paper No. MTSTC2-127*. ISPRS/IGU Joint International Conference on Geospatial Theory, Processing, Modelling and Applications, 6-8 October 2014, Toronto, Canada.

C. Non-refereed Conference Proceedings

8. Adamchuk, V., Dhawale, N., & Rene-Laforest, F. 2014. Development of on-the-spot soil analyser for measuring soil chemical properties. In: *proceedings of 12th International Conference on Precision Agriculture*, July 20-23, 2014, Sacramento. California. USA.

9. Dhawale, N.M., Adamchuk, V.I., Prasher, S.O., Viscarra Rossel, R.A., Ismail, A.A., Whalen, J.K., & Louargant, M. 2014. Comparative analysis of vis/NIR/mid-IR hyperspectrometry for measuring soil physical properties. *Paper No. 141909453.* St. Joseph. Michigan: ASABE.

10. Dhawale, N.M., Adamchuk, V.I., Viscarra Rossel, R.A., Prasher, S.O., Ismail, A.A., & Whalen, J.K. 2013. Predicting extractable soil phosphorus using visible/near-infrared hyper spectral soil reflectance measurements. *Paper No. 13-047.* Orleans. Ontario: CSBE.

11. Dhawale, N.M., Adamchuk, V.I., Prasher, S.O., Viscarra Rossel, R.A., & Ismail, A.A. 2013. Analysis of the repeatability of soil spectral data obtained using different measurement techniques. In: *Proceedings of the 3rd Global Workshop on Proximal Soil Sensing*, Potsdam, Germany, 26-29 May 2013, eds. R. Gebbers, E. Luck, and J. Ruhlmann, 161-165. Potsdam, Germany: ATB Leibniz-Institut fur Agratechnik Potsdam-Bornim.

12. Dhawale, N.M., Adamchuk, V.I., Prasher, S.O., & Dutilleul, P.R.L. 2012. Spatial data clustering using neighbourhood analysis. *Paper No. 121337939.* St. Joseph, Michigan: ASABE.

D. Published Abstracts

13. Dhawale, N.M., Adamchuk, V.I., Prasher, S.O., Whalen, J.K., Pan, L., & Mat Su, A.S. 2013. Rapid measurement of nitrate ion activity using a direct soil sensing approach. In: *poster of Scientific Program of CSSS/MSSS/CSAFM Joint Meeting*, Winnipeg, Manitoba, 21-25 July 2013, 99. Winnipeg. MB. Canada.

14. Dhawale, N.M., Adamchuk, V.I., Prasher, S.O., & Ismail, A.A. 2013. Analysis of repeatability of potentiometric and optical soil sensors used for measuring agricultural soil properties. I*n: poster of Integrated Sensor Systems (ISS) Training program, Summer School Event*. Montreal, QC. Canada: University of Montreal.

15. Dhawale, N.M., Adamchuk, V.I., & Prasher, S.O. 2013. Preparing soil samples to measure ex situ soil physical, soil chemical and soil biological attributes. *In: photo Competition, of Integrated Sensor Systems (ISS) Training program Winter Networking Event on - Bringing Sensor technology to Market*. Montreal, QC. Canada: McGill University

16. Herzellah, S., Dhawale, N., He, H., Whalen, J., Adamchuk, V., Prasher, S., Rintoul, S., Pinchuk, D., Sedman, J., & Ismail. A. 2013. Comparative assessment of Visible/Near-Infrared/Mid-Infrared reflectance techniques for the rapid analysis of soil texture. In*: poster of Pittcon Conference and Expo. 2013.* Philadelphia. PA. USA.

17. Dhawale, N.M., Adamchuk, V.I., & Prasher, S.O. 2012. A review on electro-chemical sensors for real-time mapping of agriculture soil nutrients on-the-go. *In: poster of Integrated Sensor Systems (ISS) Training program, Summer School Event.* Montreal, QC. Canada: University of Montreal.

18. Dhawale, N.M., Adamchuk, V.I., & Prasher, S.O. 2011. Measuring near surface soil organic matter content using an active optical crop canopy sensor. *In: Poster of Second Global Workshop on Proximal Soil Sensing.* 2011. Montreal, QC. Canada: McGill University.

E. Patent

19. Adamchuk, V.I., Dhawale, N.M., Rene-Laforest, F., Prasher S.O., & Pouliot, A. 2014. On the Spot Soil Analyser. *US Patent Application No. 14/494,719*.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

ATR - Attenuated Total Reflectance

ATV - All Terrain Vehicle

C - Carbon

CEC - Cation Exchange Capacity

CHEMFET - Chemically Field Effect Transistor

CI - Cone Index

DeepECa - Apparent soil Electrical Conductivity from deep layer

DRIFTS - Diffuse Reflectance Infrared Fourier Transform Spectroscopy

EC - Electrical Conductivity (mS m^{-2})

ECa - Apparent soil Electrical Conductivity

EMI - Electromagnetic Induction

FTIR - Fourier Transform Infrared Spectroscopy

GIS - Geographical Information System

GNSS - Global Navigation Satellite System

GPS - Global Positioning System

H - Hydrogen

H_2PO_4 - Hydrogen phosphate

HPO_4^{2-} - Hydrogen di-phosphate

ISE - Ion Selective Electrode

ISFET - Ion Selective Field Effect Transistor

K - Potassium

K^+ - Potassium ion

LVF - Linear Variable Filter

MC - Mean Centering

mid-IR - Mid Infrared

MP - Measurement Precision

MSC - Multiplicative Scattering Correction

MSE - Mean Squared Error

MSP - Mobile Soil Sensing Platform

N - Nitrogen

Na^+ - Sodium

NH_4^+ - Ammonium ion

NIR - Near Infrared

NO_3^- - Nitrate ion

OSA - On-the-spot Soil Analyser

P - Phosphorus

PA - Precision Agriculture

pH -$\log_{10}$ Concentration of Hydrogen ions

PLSR - Partial Least Squares Regression

$P_{Mehlich-3}$ - Mehlich 3 extractable phosphorus

PSS - Proximal Soil Sensing

QC - Québec

RSE - Ratio of Spread over Error

RTK - Real Time Kinematic

SEP - Standard Error of Prediction

Shallow - ECa Apparent soil Electrical Conductivity from shallow layer

SLR - Simple Linear Regression

SOC - Soil Organic Carbon

Soil C - Carbon Content in soil

SOM - Soil Organic Matter

SRS - Soil Reflectance Spectroscopy

SSM - Site Specific Crop Management

SWC - Soil Water Content

TC - Total Carbon

TN - Total Nitrogen

URT - Uniform Rate Technology

VFA - Variable Filter Array

VIMP - Variable Importance of Projections

vis - Visible

VRT - Variable Rate Technology

ZnSe - Zinc Selenide

Chapter 1

INTRODUCTION

1.1 General Introduction

Today's food production demands, driven by a rapidly increasing population, costly agricultural inputs and environmental risks, result in major challenges for agricultural and biological engineers, who must devise enhanced agricultural practices to maximize yield and production efficiency (Gebbers and Adamchuk, 2010).

Nitrogen (N), phosphorus (P), and potassium (K) are the three major crop nutrients essential to healthy crop growth. The predominant forms of N, P, and K in the soil are ammonium (NH_4^+) and nitrate (NO_3^-) ions, phosphate ions (H_2PO_4 and HPO_4^{2-}) and K^+ ions, respectively. Soil texture, soil organic matter (SOM) and soil pH are the three main soil properties controlling plant nutrient availability (Sparks, 2003). Both soil electrical conductivity (EC) and field topography correlate with soil properties that affect crop productivity. These properties include soil moisture, soil texture, cation exchange capacity (CEC), drainage conditions, SOM, salinity, and subsoil characteristics.

Conventionally, agricultural production is increased by practicing intense mechanization and by augmenting the quantities of crop inputs. A number of studies have shown that one of the problems associated with conventional organic and/or inorganic fertilization schemes is the risk of excess nutrients (*i.e*. those not taken up by crops or other plants) reaching watercourses (Hooda *et al.,* 2000). Excess N in the form of NO_3^- is very mobile in agricultural soils and predominantly reaches watercourses by leaching through the soil and into subsurface drains. Phosphates are comparatively immobile and bind easily to soil particles. They tend to remain in the first few centimetres of topsoil, where they can pose a non-point source pollution risk when rain-driven soil erosion carries surface runoff to streams (Chambers *et al.,* 1997; Carefoot and Whalen 2003). In some situations, N, particularly in the form of soil-bound NH_4^+, but also

recently applied NO_3^-, can also be lost in runoff (Canadian Council of Ministers for the Environment, 2007). K is not usually considered a nutrient of environmental (water quality) concern, although when present in excessive amounts, its uptake into forages can exceed acceptable concentrations and have a significant impact on cattle health (Schmidt and Hughes-Games, 2010). Reduced water quality, cost of applying extra inputs, and loss of land due to soil degradation can cause a negative socio-economic impact for the farming community.

Site-specific management (SSM), also referred to as Precision Agriculture (PA), holds considerable potential for improving agricultural production and reducing the cost and risks posed by excess application of agriculture inputs. It accounts for spatial and temporal variability in a site specific manner and therefore, promotes the efficient use of fertilizers, pesticides, and other crop inputs (Pierce and Nowak, 1999). In simple terms, SSM is about applying soil or crop inputs in the Right Amount, at the Right Time, at the Right Place, from the Right Source and in the Right Manner (Robert *et al.,* 1994; Khosla, 2008).

Measuring soil properties and mapping their variability across landscapes is a powerful means to allow farmers to make intelligent management decisions that are valuable for their crop production (Pierce and Nowak, 1999). However, the most critical aspect of soil testing is actually obtaining representative soil samples (*i.e*. collected with adequate spatial density at the proper depth and at the appropriate time) and while conventional soil sampling techniques are laborious and time consuming, PSS allows for rapid and inexpensive collection of precise, quantitative, and fine-resolution data (Viscarra Rossel *et al.,* 2011).

The term PSS is used when field-based sensors are used to obtain signals from the soil when the sensor's detector is in contact with, or close to (within 2 m) the soil (Viscarra Rossel and McBratney, 1998; Viscarra Rossel *et al.,* 2010). The sensors provide soil information because the signals correspond to physical measures which can be related to soils and their properties. PSS data can be gathered by installing a single sensor or an array of soil sensors permanently in field locations or in an on-the-go manner by dragging them over field transects or in an on-the-spot or in a go-stop-measure-go manner. Each approach may have its own advantages and limitations and in many cases, an integrated approach is required.

According to a series of annual surveys (Holland *et al.,* 2013; Croplife/Purdue, 2005-2013), SSM technologies have been growing slowly but steadily since they were introduced in the late 1980's. The biggest growth in SSM technologies has been in the adoption of GPS guidance and automation systems such as light bar control, auto steer; planter row unit controls and sprayer section controls. Field mapping services are the next most popular SSM technology and are mostly intended for legal, billing or insurance purposes. According to the surveys, barriers to the adoption of technology include: local topography (limits the use by farmers); farm income pressure (limits use of precision services); lack of confidence in site-specific recommendations; variable soil types (limits profitability of precision); cost of precision services greater than benefits; interpreting data to make logical decisions is too time consuming.

At the same time, some of the major challenges encountered by the SSM equipment dealers were: seeking the right employees and cost of employees; demonstrating the value to growers; risk of the investment cost of equipment due to rapid technological changes; equipment complexity and incompatibility between equipment (Holland *et al.,* 2013; Croplife/Purdue, 2005-2013).

Soil data gathered using PSS and soil sampling with conventional laboratory testing can be used together to produce soil maps with greater accuracies (Hoeft *et al.,* 1996). PSS technologies can benefit the field mapping service providers by: 1) increasing the quality of maps, 2) overcoming the cost and time related challenges of conventional mapping and 3) maximising the mutual gains offered by service providers to producers.

At present, SSM seeks to use information technologies to improve the economic, environmental and social aspects of modern farming operations. Among an array of technological breakthroughs, the ability to accurately account for spatial variability in crop growing environments has been one of the most challenging to implement in a robust manner (Viscarra Rossel *et al.,* 2011). The uncertain nature of collecting economically feasible data which reveal the spatial distribution of key soil attributes has been the main concern of early adopters of new technologies (Dobermann *et al.,* 2004; McBratney *et al.,* 2005). This suggests the need for advanced equipment and methods to obtain high-density measurements of soil physical and

chemical properties at a relatively low cost. For example, an intelligent autonomous mobile robot, similar to, but simpler and more affordable than the Mars Rover (Volpe *et al*., 1996), could serve to obtain high-density measurements of important soil attributes, including N, P, C and H levels. However, such a system would have to be affordable to a North American farmer and sufficiently robust to operate in uneven field surface conditions.

The best instances of implementing such an approach are associated with a number of internationally-known agricultural robotics projects, including: the Autonomous Crop Treatment Vehicle, the Weedy Robot and the Weeding Robot, the Autonomous Tractor, the Autonomous Christmas Tree Weeder, the Supportive Autonomous Vehicle for AGriculturE, the Autonomous Platform and Information System, the Hortibot, and the SlugBot.[1] However, to the best knowledge of the author of this thesis, no automated soil mapping or soil sensing systems for agriculture have been reported. Therefore, it would be very interesting and equally valuable to design and develop such a sytem.

1.2 Statement of Rational Objectives of the Research

1.2.1 Statement of Rationale

It is clear from the above discussion, that SSM is an ongoing field of research. Advancements in SSM technologies and techniques are expected to optimize and improve the current benefits. PSS is a new and independently growing discipline. Current engineering research in PSS mainly focuses on the use of appropriate proximal soil sensors, their signal processing and data interpretation. Advancements in an automated mobile soil data acquisition platform seek to develop diverse sensor fusion and real-time control algorithms, where automation is expected to enhance the efficiency of the platforms used in collecting soil information. Unmanned mobile soil instrumentation is a potential solution where it is crucial to collect sensor measurements on time, where health of human labour is at risk and data misinterpretation and financial losses are undesirable. However, the challenge of choosing the most suitable soil data acquisition platform,

---

[1] Available at http://www.unibots.com/Agricultural_Robotics_Portal.htm

soil sensor fusion algorithm and strategies for soil information collection continue to be "hot" topics in this discipline.

1.2.2 Objectives of the Research

The overall objective of this research is to develop an on-the-spot soil analyser to enhance and advance proximal soil sensing technologies.

The specific objectives are as follows:

1. To develop and evaluate an on-the-spot soil analysing platform for *in situ* measurements of NO_3^- and H^+ ion activity (Chapter 3);

2. To analyse the capability of visible-near-infrared and mid-infrared spectral data to determine soil texture, soil C and soil total P (Chapters 4, 5, 6 and 7); and

3. To develop a methodology for the clustering of high-density proximal-sensing soil data (Chapter 8).

Chapter 2

GENERAL REVIEW OF LITERATURE

2.1 Site Specific Management (SSM)

At present SSM is practised using a group of technological tools *viz*: Global Navigation Satellite System (GNSS), Geographical Information System (GIS), Remote Sensing, Yield Monitoring and Mapping, Variable Rate Technology, Navigation, Auto-Guidance and Autonomous Systems, Soil Sampling and Mapping, and PSS (Blackmore, 2002; Viscarra Rossel and McBratney, 1998; Khosla, 2008; Viscarra Rossel *et al.,* 2010). These will be discussed briefly here.

2.1.1 Global Navigation Satellite System (GNSS)

Global Navigation Satellite System is one of the key technologies used in SSM. United States NAVSTAR Global Positioning System (GPS) and the Russian GLONASS are global operational GNSS's. GPS receivers with sufficient accuracy are available at moderate cost and are used for yield mapping, grid sampling, variable rate application, and other precision activities (Nemenyi *et al.,* 2003). Position accuracies to within centimetres are achievable by choosing advanced systems specially equipped with error correction hardware, such as Real Time Kinematic (RTK) or Differential (D) GPS.

2.1.2 Geographical Information System (GIS)

Geographical Information System is the soft platform used to save and handle spatial information electronically. The GIS makes it possible to generate a complex view of agricultural fields and to make valid agro-technological decisions (Pecze, 2001). GIS application software enables the farmer to assemble and organize different sources of site-specific and geographical information in various layers.

2.1.3 Remote Sensing

Remote sensing is a popular technology used for acquiring information about the earth's surface from far away. This is done by recording the electromagnetic (EM) energy, which is either reflected or emitted from the earth's surface. The information is then processed, analyzed, and used to develop prescription maps for variable rate applications (Grisso *et al.,* 2011). Satellite images have been beneficial for the acquisition of information on large areas; however, by placing optical sensors on agro-vehicles or beneath unmanned aerial vehicles (UAVs), it is possible to map relative crop reflectance's at small scales (Adamchuk *et al.,* 2003).

2.1.4 Yield Monitoring and Mapping

The crop itself is the best sensor of its environment (Khosla, 2008). Yield monitors are a logical first step for those who want to begin practicing SSM and yield maps form a very important element in PA technologies (Blackmore 2000 and 2003). A yield map in combination with other gathered soil and crop attributes is intended to give the user an accurate assessment of how yields vary within a field (Fridgen *et al.,* 2004).

2.1.5 Variable Rate Technology

Variable Rate Technology (VRT) involves equipment with the ability to modify the application rate as it moves across the field (Adamchuk and Mulliken, 2005). VRT adjusts pesticide, herbicide and seeding application rates, as necessary. Gandonou (2005) showed that variable rate fertilizer application results in increased yields. On fields with high spatial variability, Variable Rate Application (VRA) resulted in significantly lower environmental N losses when compared to uniform rate applications (URA) (Roberts *et al*., 2001).

The two basic technologies for VRA are map-based and sensor-based. Map-based VRA adjusts input application rates based on a prescription map. Sensor-based VRA requires neither a map nor positioning system. Sensors on the applicator measure soil properties or crop characteristics "on-the-go". A feedback system calculates the input needs of the soil or crops and transfers the

information to a controller, which delivers the input to the location measured by the sensor (Grisso *et al.,* 2011).

2.1.6 Navigation, Auto-Guidance and Autonomous Systems

Based on machine operator functionalities, guidance systems are classified as guidance aides or autonomous systems. Guidance aides refer to systems that provide guidance information to the operator without replacing the operator (Blackmore, 2003; Ima and Mann, 2004). One of the more recently developed and most broadly adopted methods to reduce guidance error is the light-bar technology, which assists the machine operator in driving while minimizing application errors of overlaps and skips and reducing operator fatigue.

Another rapidly expanding technology is the auto-guidance (also called auto-steering) which is based on the use of GNSS receivers to perform field operations in a strict geometrical relationship with a previous travel path or other predefined geographical coordinates, without direct inputs from an operator (Easterly *et al.,* 2010). Auto-guidance systems, when adopted by producers, can improve field operations and management practices. For example, the producers can reduce skips and overlaps, save on fuel and labor costs, better manage seed, fertilizer and chemical inputs, simplify operations, and even add hours to the day during critical operating windows (Abidine *et al.,* 2004; Griffin, 2009; Winstead, 2010). However, the auto-guidance systems available to producers have different sets of interfaces and sensor configurations and their performance is frequently associated with an anticipated level of auto-guidance error, usually referred to as cross-track error (XTE).

Human capacity is limited in its ability to handle multi-sourced, simultaneous information and the introduction of miniature robots to perform agricultural tasks autonomously may prove to be a realistic option for farmers of the future (Blackmore *et al.,* 2002; Blackmore *et al.,* 2005; Haapala *et al.,* 2006). These robots may perform tasks such as scouting for weeds and diseases, yield and field mapping, or plant specific operations like sowing and fertilizing (Shockley, 2010).

2.1.7 Soil Sampling and Mapping

Different levels of nutrients and a variety of soil types are common features in agricultural soils (Sparks, 2003). Conventionally, soil sampling and laboratory testing is the best way to identify these differences. Both grid and directed sampling are used to evaluate soil properties for the management of variable rates of fertility.

Gathering data from soil samples to create a soil map is a relatively expensive and time consuming task but provides valuable information for future fertilizer treatments and the design of decision support systems (Adamchuk *et al.,* 1999; McBratney *et al.,* 2003; Ping and Doberman, 2003; Dobermann *et al.,* 2004).

2.1.8 Proximal Soil Sensing

Proximal soil sensing is a set of technologies developed to measure the physical, chemical and biological properties of soil using a sensor placed in contact with, or at a proximal distance (less than 2 m), to the soil being characterised (Viscarra Rossel *et al.,* 2011). The technology uses sensors based on various working principles such as electrical, mechanical, optical, chemical, acoustic, etc. to detect or measure the target soil property. As illustrated in Fig. 2.1, PSS is classified into ten categories that are based on the mode of operation, (mobile/static), measurement type and environment (invasive/non-invasive/*in situ*/*ex situ*), use of external energy source (passive/active) and inference type (indirect/direct).

Among the categories, mobile operation can be advantageous only if sensor response time is faster then the speed of data collection. Direct inference sensors have advantages over the indirect ones due to the need to develop calibration models using laboratory analysed data. However, conventional soil sampling and laboratory analysis is necessary for developing and calibrating the calibration models, and the number of representative soil samples used in this step is expected to decline over time (Viscarra Rossel *et al.,* 2011).

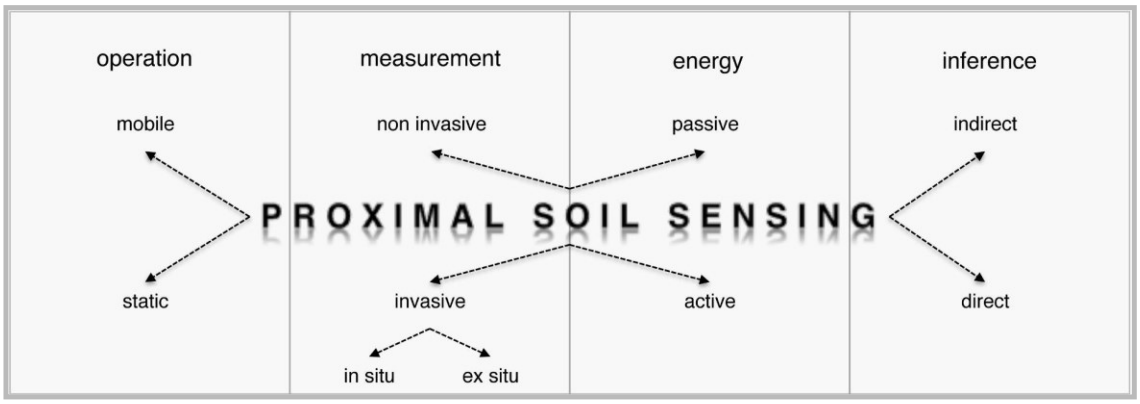


Fig. 2.1 Classification of PSS Systems (modified from Viscarra Rossel *et al.,* 2011)

From the above discussion, it is clear that the development of sensors and sensing platforms are expected to increase the efficiency and effectiveness of precision agriculture. Sensors developed for on-the-go measurements of soil properties have the potential to provide benefits due to the increased density of measurements collected at a relatively low cost and in a very short time.

2.2 PSS Sensors

The three sensing principles most widely used for PSS (Adamchuk *et al.*, 1999; Viscarra Rossel *et al.,* 2011) are as follows:

2.2.1 Potentiometry

Fig. 2.2 illustrates an ion-selective electrode (ISE) system used in a wide variety of applications for determining the concentrations of various ions in aqueous solutions. ISE's have proven to be the most advantageous, convenient and rapid field type measuring systems available. A wide range of ISEs measuring levels of soil anions and cations such as NO_3^-, K^+, Na^+ and pH are commercially available (Adamchuk *et al.,* 1999; Viscarra Rossel, 2001; Sethuramasamyraja, 2006). PSS using potentiometric sensors is currently an active area of research with particular focus on the development of mobile soil pH, lime requirements, and nutrient sensing (e.g. Viscarra Rossel and McBratney, 1997; Adamchuk *et al.,* 1999; Viscarra Rossel *et al.,* 2005). The development of phosphate ion selective electrodes still constitutes a part of many vigorous research projects being carried out around the world and the development of a phosphate

ion-selective probe using cobalt electrodes has shown promising results (Kim *et al.,* 2007; Kim *et al.,* 2009) and suggests the need for further research.

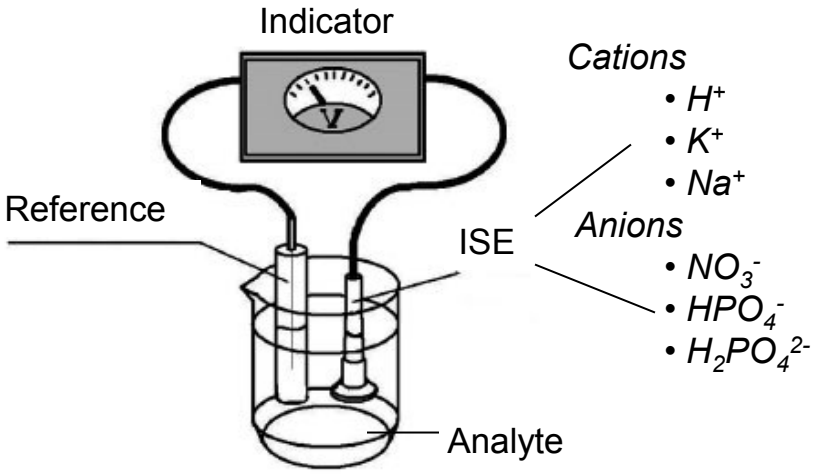


Fig. 2.2 Ion-selective electrode setup used for measuring ion activities in soil solutions

2.2.2 Soil Spectroscopy

A common technique which uses optical reflectance for PSS, relates to the use of soil visible (vis), near-infrared (NIR) and mid-infrared (mid-IR) spectroscopy (Viscarra Rossel *et al.,* 2011). Viscarra Rossel *et al.* (2006b) provide a summary of vis/NIR/Mid-IR spectroscopy measurements in soil as well as a discussion of the basic principles.

Interest in using reflectance spectroscopy (Fig. 2.3) to measure soil properties is widespread because the techniques are rapid, relatively inexpensive, require minimal sample preparation. In addition, they are non-destructive, require no hazardous chemicals, and several soil properties can be measured from a single scan (Viscarra Rossel *et al.,* 2006a). Vis/NIR/mid-IR spectroscopy has been used successfully to quantify soil texture, soil organic matter, moisture, total C, total N, available P and cation exchange capacity (Dalal and Henry 1986; Chang *et al., 2001;* Viscarra Rossel *et al.*, 2006b).

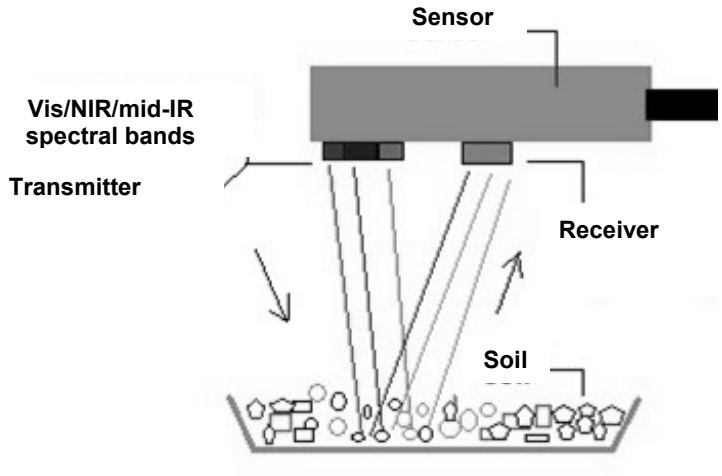


Fig. 2.3 A simple illustration of measuring of soil optical reflectance

2.2.3 Electrical Conductivity

Two popular methods for mapping soil electrical conductivity are based on the use of electromagnetic induction (EMI) (Fig. 2.4a) (Kitchen *et al*., 1996; Sudduth *et al*., 2001; Anderson-cook *et al*., 2002) or a contact electrode (Fig. 2.4b).

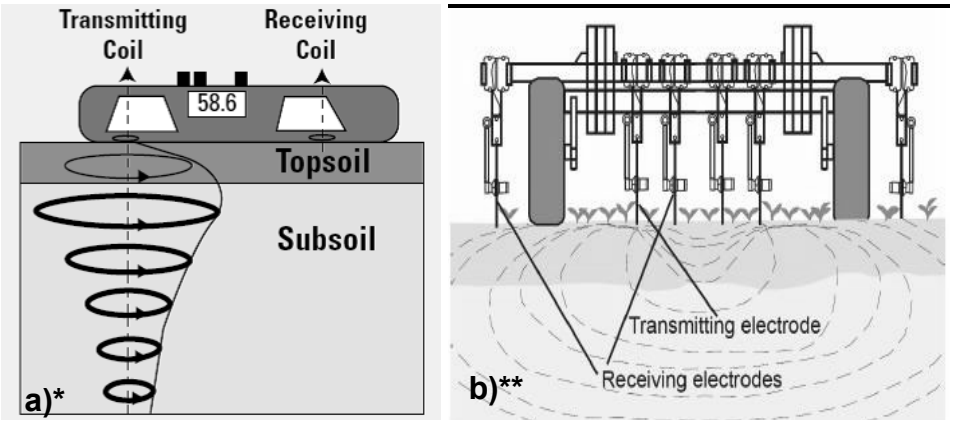


Fig. 2.4 Apparent soil electrical conductivity measurement (ECa) setups: (a) non-contact type, (b) contact type

* Image source: Davis *et al.* (1997) ; ** Image source: Grisso *et al.* (2009)

Both systems show similar results (Doerge *et al.,* 2002). Corwin and Lesch (2005) provide a comprehensive review of the historical development of EC measurements in agriculture as well as a discussion of basic principles, different geophysical techniques for measuring EC, mobile EC measurement equipment and applications to SSM.

2.3 PSS Platforms

There are two mobile-operated PSS sensing platforms that directly relate to this study: 'on-the-go' and 'on-the-spot' sensing.

2.3.1 On-the-Go Sensing

A prototype automated soil pH mapping system (Adamchuk *et al*., 1999) was developed to obtain high resolution measurements of soil pH at a relatively low cost. While traveling across the field, a soil sampling mechanism located in a toolbar-mounted shank scooped a sample of soil from a depth of approximately 0.10 m and brought it into firm contact with the sensitive membranes of two flat-surface, ion-selective electrodes. After stabilization of the electrode output (typically 5 to 15 s), the electrode surfaces were rinsed while a new soil sample was being obtained. Christy *et al*. (2004), modified the soil sampling mechanism by using a horizontal coring tube, which allowed reliable sampling in diverse conditions. A commercial system utilizing this method has been available since 2003 (Veris Technologies, Inc., Salina, Kansas, USA).

A prototype laboratory and field portable multi-ion measuring system (MIMS) for proximal sensing of soil NO_3^-, K^+ and Na^+ using ISEs is reported by Lobsey *et al.* (2010). The technique involves the use of a universal extracting solution for rapid estimation of soil extractable nutrient concentrations and provides in-field soil nutrient analysis in less than 30 seconds. The measurement process performs ISE calibration, temperature compensation, and soil analysis with nutrient estimation in a fully automated manner.

A commercially available on-the-go, multi-soil property mapping platform (Veris® MSP3[2], Veris Technologies, Inc., Salina, Kansas, USA) uses a combination of soil electrical conductivity sensors and optical sensors operating underneath crop residue and the soil surface and pH electrodes. Variability in soil texture, soil organic matter and pH can be mapped using a sensor fusion approach.

Equipment developers have used either, hydraulically-powered, or hybrid-powered systems and components that are both bulky and noisy and could limit the usage of smaller sized vehicles. The sensor detectors directly contact the soil surface and cause undesirable damage to the field across the travel path. The necessity for continuous engagement between soil and sensor can therefore be the limiting factor for its use especially in pasture. It would also be ideal if the operating mechanism could run using small, clean, electrically-powered systems and components.

2.3.2. On-the-Spot Sensing

Contrast to on-the-go, ruggedized mobile based platforms carrying vis-NIR spectroscopic instruments and integrated soil sensing probes for collecting point measurements in the field have been developed (Veris® P4000[3], Veris Technologies Inc., Salina, Kansas USA). The P4000 (Fig. 2.5) consists of a dual type spectrometer instrument, operating in the visible and near-infrared regions of the spectrum. The advantage of the P4000 system is its ability to simultaneously collect multi-sensor measurements of soil including vis-NIR soil reflectance. Soil electrical conductivity (EC) and soil penetrating resistance (forces) are the two other types of measurements that can be performed using the Veris P4000 in soil profiles up to depths of 100 cm. Numerous soil attributes can be predicted indirectly by using the sensor based measurements compared against conventional laboratory analysed data (Hodge and Sudduth, 2012; Dhawale *et al.,* 2013b; Wetterlind *et al*., 2013; Piikki *et al.,* 2014).

---

[2] Mention of a trade name, proprietary product, or company name is for presentation clarity and does not imply endorsement by the authors or McGill University, nor does it imply exclusion of other products that may also be suitable.

[3] Mention of a trade name, proprietary product, or company name is for presentation clarity and does not imply endorsement by the authors or McGill University, nor does it imply exclusion of other products that may also be suitable.

The P4000 is only available for research purposes. In its current modification it uses an array of sensors that indirectly infer the soil properties being measured. The equipment is operated using a heavy pick up truck that can be difficult to manoeuvre between transects in rough terrains and can cause compaction across the travel path during wet seasons. The system requires at least two skilled personnel: a driver and an operator which may increase the cost of data collection. A multi-soil property mapping platform, manoeuvered using an all terrain vehicle and equipped with a fully automated data collection procedure will advance PSS technologies.

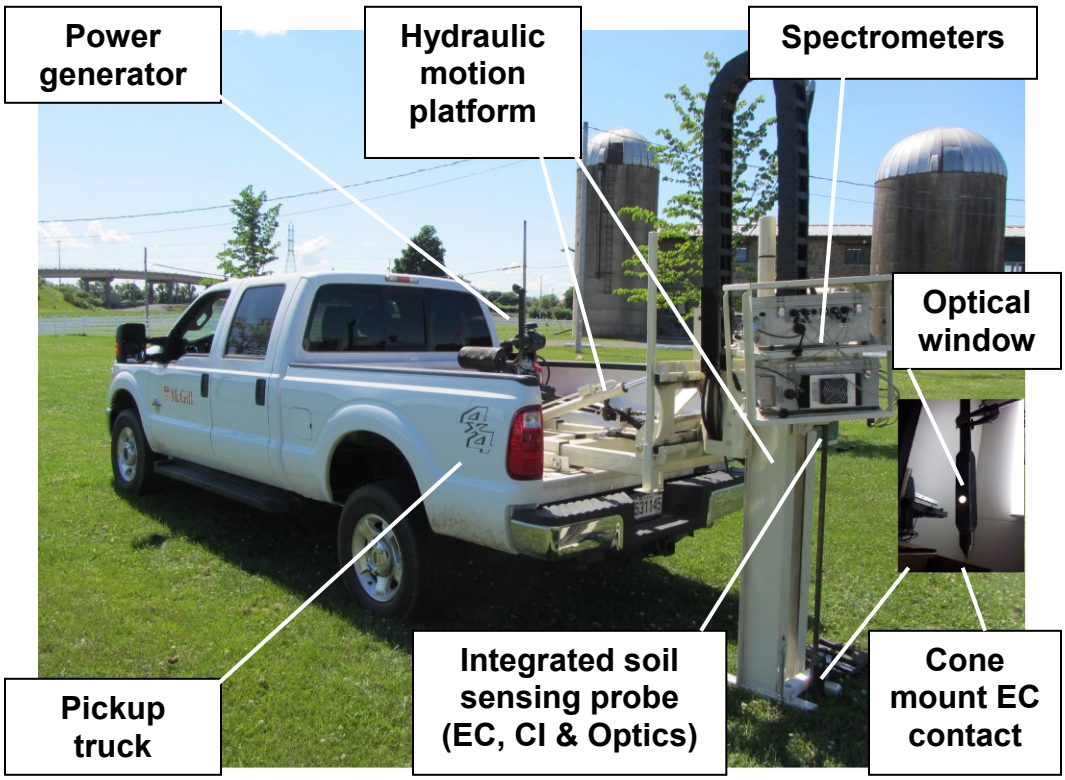


Fig. 2.5 The Veris® P4000, and its key components

2.4. Analysing PSS Data

The post-collection challenge of high-density PSS data is to group the information and delineate a chosen agricultural field into various management groups related to the measured soil properties. To pursue various SSM practices, spatial data is frequently split into groups (clusters

or zones) to represent significantly different growing conditions (Fraisse *et al.,* 2001; Ping and Dobermann, 2003). Geo-spatial data clustering is an important process (Li and Wang, 2010), which is widely used in remote sensing (Deng *et al.,* 2003), neuroanatomy analysis (Prodanov *et al.,* 2007), and other areas.

Several different spatial clustering algorithms have been developed to group geospatially dense PSS-based measurements of soil attributes into management zones. For example, Management Zone Analyst (MZA) (Fridgen *et al.,* 2004) represents a publicly available tool accepted by a number of practitioners. The algorithm is based on computing a distance matrix and performing clustering over this new distance matrix. It is closely related to the popular k-means clustering algorithm, where quality of the resulting clusters depends heavily on the selection of initial centroids and the results are not repeatable. However, this method requires cross-validation to select the best among several runs (Abdul-Nazeer and Sebastian, 2009). Although the method allows multidimensional data analysis, complexity and frequently occurring discontinuities of management zones make this technology non-robust for potential users (Shatar and McBratney, 2001; Kerby *et al.,* 2007).

From the above discussion, it is evident that a new method for analysing high density PSS data is necessary. The new method should be capable of grouping contiguous and homogeneous geo-spatially dense measurements into managements groups; since the majority of reported methods require a human analyser, introducing automation into the analysis procedure is also important.

2.5. Soil Sampling/Sensing Dilemma

Where to measure soil properties using proximal soil sensors depends on the type of the sensor (Viscarra Rossel *et al.,* 2011). If the sensor measures the soil property directly and on-the-spot, then sampling or sensing problems will be the same as conventional spatial soil sampling because it requires optimization of the geographical coverage of the measurements. In this case, the sampling problems might relate to the resolution of the measurements or to the response time

of the sensor in order to optimize the amount of information collected (Viscarra Rossel *et al.,* 2011).

If the measurements are made indirectly, a calibration step will be required to predict the soil property from the sensor measurements. For such a case, a calibration sampling design to optimize coverage of the feature space will be required and ideally, the sampling should cover the same geographic space as the calibrations (Adamchuk *et al.* 2008; Adamchuk *et al.* 2011a; Viscarra Rossel *et al.,* 2011).

de Gruijter *et al.* (2010) described geographical and feature space sampling with PSS for fine-resolution soil mapping. Christy (2008) proposed a sampling design for the calibration of vis-NIR spectra collected on-the-go and Adamchuk *et al. (*2008) and Adamchuk *et al. (*2011b) compared designs that consider both geographic and feature space, field boundaries and other transition zones.

In summary, if PSS is used to measure a soil property directly and using an on-the-spot technique then the sampling or sensing problems will be the same as with conventional soil sampling. If PSS is used to measure the soil property indirectly using an on-the-spot technique then further research is required to develop optimal calibration sampling designs.

Connecting Text for Chapter 3

The role of proximal soil sensing techniques for in-field soil nutrient analysis and the literature relating to the existing research in this area were thoroughly described in the previous chapters. Chapter 3 is related to the first objective listed in Chapter 1. This chapter discusses the principle used, development and field testing of a portable and automated On-the-spot Soil Analyser (OSA). The OSA uses an array of potentiometric sensors and can simultaneously determine ions such as H^+ and NO_3^- on-the-spot. The technology is expected to provide an opportunity to extend the suite of deployable sensors and thus, allow for sensor fusion algorithms and integrated data acquisition practices. Different parts of this study were presented at the conferences listed below and a manuscript has been prepared for submission to the "Computers and Electronics in Agriculture Journal".

Dhawale, N.M., F. Rene-Laforest, V.I. Adamchuk, & S.O. Prasher. 2014. Development and evaluation of an automated on-the-spot soil analyser (OSA) for proximal soil sensing. *Computers and Electronics in Agriculture Journal* (to be submitted).

Adamchuk, V. I, **Dhawale, N**.M., & F. Rene-Laforest. 2014. Development of on-the-spot soil analyser for measuring soil chemical properties. *In: proceedings of 12th International Conference on Precision Agriculture, July 20-23, 2014,* Sacramento, California, USA.

Dhawale, N.M., V.I Adamchuk, S.O. Prasher, J.K. Whalen, L. Pan, & A.S. Mat Su. 2013. Rapid measurement of nitrate ion activity using a direct soil sensing approach. *In: Poster of Scientific Program of CSSS/MSSS/CSAFM Joint Meeting*, Winnipeg, Manitoba, 21-25 July 2013, 99. Winnipeg, Winnipeg, MB, Canada.

Dhawale, N.M., V.I Adamchuk, & S.O. Prasher. 2012. A review on electro-chemical sensors for real-time mapping of agriculture soil nutrients on-the-go. *In: poster of Integrated Sensor Systems (ISS) Training program, Summer School Event*. University of Montreal, Montreal, QC. Canada.

Adamchuk, V.I., **Dhawale, N.M**., Rene-Laforest, F., Prasher S.O., & Pouliot, A. 2014. On the Spot Soil Analyser. *US Patent Application No. 14/494,719*.

Chapter 3

DEVELOPMENT AND EVALUATION OF AN AUTOMATED ON-THE-SPOT SOIL ANALYSER FOR PROXIMAL SOIL SENSING

Dhawale, N., V.I. Adamchuk, S.O. Prasher, F. Rene-Laforest

Abstract

The goal of this research is to develop an On-the-spot Soil Analyser (OSA) capable of simultaneously deploying several different sensors to measure soil properties at a predefined depth. The mechanism developed is able to rapidly remove topsoil, condition the soil surface and bring the designated sensors into direct contact with the soil. After the measurements are obtained and the geographical coordinates are recorded, the sensors are retrieved and the analyser is raised slightly above the ground surface. The distorted area is restored and the analyser is converted back to transportation mode making it ready for the next set of measurements. Unlike other systems, this mechanism allows for the deployment of multiple sensors at a given measurement depth and in a completely automated mode of operation. This technology should provide an opportunity to extend the suite of deployable sensors and to automate the process, thus, allowing for advanced sensor fusion algorithms and integrated data acquisition practices.

Keywords: Proximal soil sensing, ion-selective electrodes, autonomous vehicles, soil pH, nitrate

3.1 Introduction

Sensor systems have been widely used to obtain information on soil physical, chemical and biological attributes. Depending on the distance to the target, existing sensing platforms are remote (aerial and satellite-based systems) or proximal, (placed in contact with soil or at a distance of less than 2 m) (Viscarra Rossel *et al*., 2011). Unlike conventional soil sampling and laboratory analysis, many proximal soil sensing (PSS) systems have been developed to obtain data which reveals spatial (horizontal and vertical) as well as temporal changes in selected soil properties. Thus, PSS reduces the number of soil samples needed for follow-up measurements of soil attributes required by traditional techniques.

PSS-based data have been used in agriculture, construction, ecology, archaeology, and other important activities. PSS measurements can be combined with geographic coordinates to create geographical maps. With different instruments that are based on electrical, electromagnetic, optical, radiometric, magnetic, seismic, mechanical, acoustic, electrochemical and other measurement techniques, it is possible to better understand spatial and temporal soil heterogeneity. Due to the complex nature of soil, sensors respond to a variety of soil properties and this leads to the need for sensor fusion (Adamchuk *et al*., 2011b). By combining different data sources (including remote sensing and destructive soil sample analysis), it is possible to increase the quality of the information obtained, while minimizing costs.

One of the most attractive approaches is to create high-resolution thematic soil maps using sensor systems that are moved across landscapes (Hummel *et al*., 1996; Sudduth *et al*., 1997; Adamchuk *et al*., 2004; Shibusawa 2006). These on-the-go soil sensors make periodic measurements while traveling across the terrain and they associate each measurement with geographic coordinates. The main disadvantage of on-the-go sensing is the limited time allowed for each measurement and, when using contact techniques, soil distortion is created along the entire path. This is especially relevant to mapping soil chemical attributes using potentiometry, as shown in Fig. 3.1. According to Adamchuk *et al*. (1999), this method requires the collection of a small amount of soil from a predefined depth and bringing it into contact with a combination of ion-selective electrodes (ISEs). In a typical commercial operation, an average 12s per

measurement allows for more than 20 measurements per hectare while operating at 10 km h^{-1} in parallel passes 15 m apart.

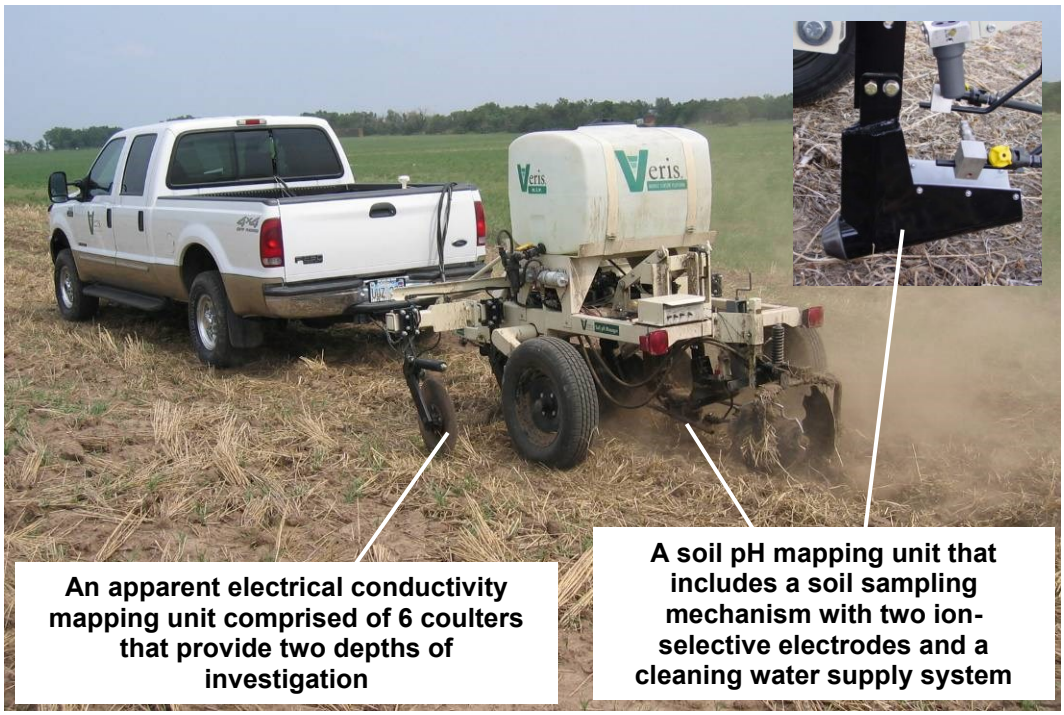


Fig. 3.1 The Veris® Mobile Sensor Platform (MSP)

As an alternative to on-the-go mapping techniques, on-the-spot measurements can be made in situations where spatially sporadic test locations are needed or soil coverage does not allow for the continuous engagement of soil and parts of the sensor system (e.g., pasture). For example, the instrumented manual probe developed by Adamchuk (2011) shown in Fig. 3.2 can be used to make real-time measurements of soil pH, soluble K, or residual NO_3^- (depending on the type of ISE) while walking the test area. An all-terrain vehicle (ATV) mounted with equipment (Fig. 3.3) can be used to make similar measurements in a consistent and more ergonomic manner. In both cases, equipment has been limited to only one electrode and needs an operator to conduct the test.

The aim of this paper is to report on the development of an on-the-spot soil analyser capable of simultaneously deploying several different sensors to measure soil properties at a predefined depth using preconditioned soil surface in contact with the sensors. This technology should

provide an opportunity to extend the suite of deployable sensors and to automate the process, thus, allowing for advanced sensor fusion algorithms and integrated data acquisition practices.

Fig. 3.2 Instrumented manual probe for on-the-spot soil ion measurements

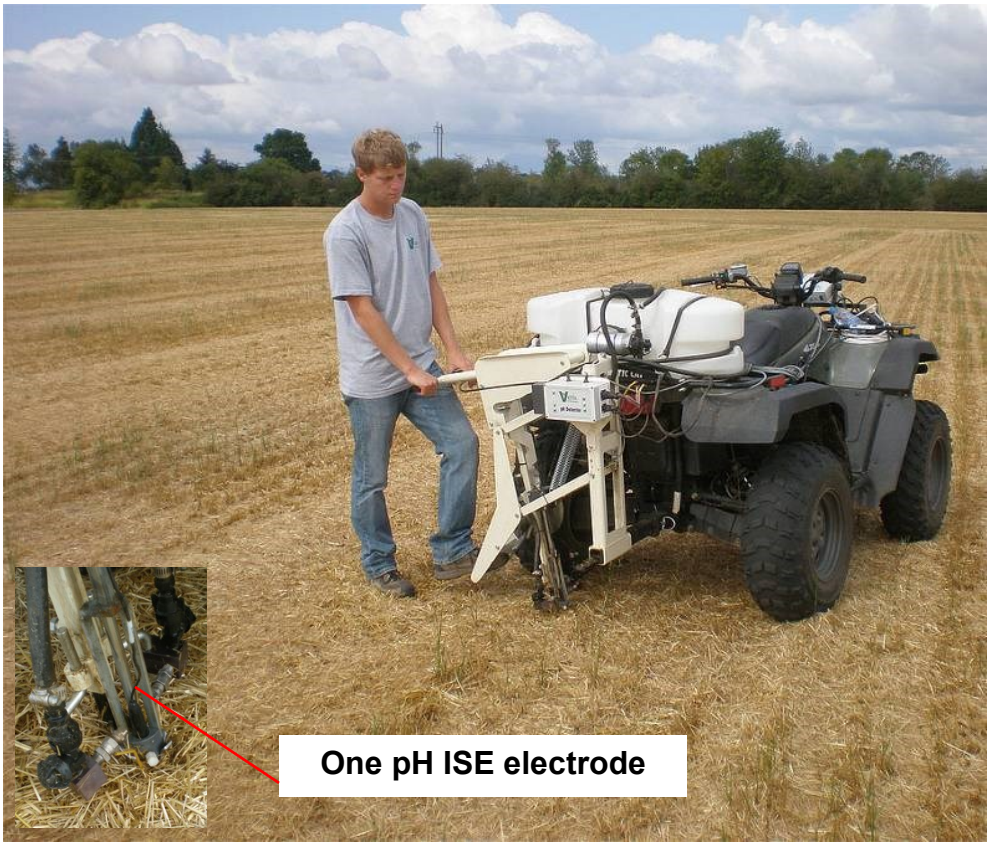


Fig. 3.3 The Veris® pH Detector (http://www.veristech.com)

3.2 Materials and Methods

3.2.1 Electrode Selection

Although the OSA should be suitable for different types of sensors, enhancing the electrochemical *in situ* measurement capabilities appears to be the most worthwhile. Compared to other PSS solutions, electrochemical sensors are direct when measuring the concentration of specific ions in a soil solution. The most common technique is to measure soil pH, although the use of ISE's for measuring other ions is also common in environmental monitoring applications. Various prototypes have been developed to use ISEs and ion-selective field effect transistors (ISFETs) to map soil pH, lime requirements and macro-nutrients (e.g. Adamchuk *et al*., 1999; Adsett and Zoerb 1991; Birrell and Hummel 2001; Sibley *et al*., 2009; Viscarra Rossel and McBratney 1997; Viscarra Rossel *et al*., 2005). Kim *et al*. (2009) provide a recent review of relevant innovations. ISEs are potentiometric sensors that use ion-selective membranes to determine the concentration of the target species in a solution surrounding the membrane. During measurements, an electromotive force is generated at the sensing surface proportional to the log of the ion activity. This electromotive force is measured using a reference electrode (liquid junction). Many ISEs (e.g., pH, NO_3^-, Na^+, K^+ and Ca^{2+}) are commercially available.

Alternatively, ISFETs combine ISE technology with that of the field effect transistor. The key advantages of pH ISFETs over standard glass pH electrodes are their small size, increased durability, fast response, and the ability to be mass produced using microelectronic manufacturing techniques (Viscarra Rossel *et al*., 2011). ISFETs can be chemically modified to produce CHEMFETs, which are selective for different ionic species. CHEMFETs selective for NO_3^-, K^+ and Ca^{2+} have been developed and evaluated for use in soil nutrient sensing (Artigas *et al*., 2001; Adamchuk *et al*., 1999; Adsett and Zoerb 1991; Birrell and Hummel 2001; Kim *et al*., 2009; Sibley *et al*., 2009; Viscarra Rossel and McBratney 1997; Viscarra Rossel *et al*., 2005). Finally, metal electrodes are also being explored for PSS applications to address a need for increased physical durability. Antimony electrodes are being researched as a durable alternative to glass when used in direct contact with soil pH measurements (Adamchuk and Lund 2008;

Viscarra Rossel and McBratney 1997). Cobalt rod-based ISEs for measuring soil phosphates have been explored by Kim *et al*. (2007).

3.2.2 System Development

To successfully deploy any electrochemical sensor, the OSA should: 1) remove cropping material and topsoil to expose soil at a given depth, 2) apply water or a salt solution when necessary, 3) bring all the electrodes into gentle, yet firm, contact with the wet soil surface, 4) analyze sensor response to predict a steady-state output, 5) record geographic coordinates along with the sensor output, 6) retrieve and wash the sensors, and 7) cover the exposed area. The entire process should be safe, automated and suitable for different vehicle platforms. This requires a compact design, light in weight and using 12V DC power supply.

Based on these requirements, a conceptual design was formulated as shown in Fig. 3.4. Two virtual circles C_1 (*radius = 14 cm*) and C_2 (*radius = 61 cm*) together represent the geometry of the soil removal and preparation unit while interacting on a given soil layer S. The radius for C_1 and C_2 were chosen based on the necessity of exposing soil depths of up to 12 cm and the potential constraints exhibited by the electro-mechanical components.

Smaller circles C_1 and C_1' represent the soil cutter tool with two blades mounted opposite each other and denominate the start and end positions, respectively. The larger circle (C_2) represents the feeder for the soil cutter tool. The center of C_1 is fixed eccentrically away from C_2 at a distance of 47 cm. This virtually allowed C_1 to roll freely inside C_2; however, the limits were set so it can roll only within angles < 90° (transportation mode and the sensor deployment mode). The angle between positions C_1 and C_1' varies between 0°-45° and depends on the soil depths to be exposed. For example, to remove soil material up to depths of 8 cm, C_1 needs to be moved through an angle ~ 30°. The shaded portion shown inside the polygon ABcaA, represents the material removed from the soil layer S. Since the area of the circular segment is equal to the area of the circular sector minus the area of the triangular portions, for a given geometry the total area of the shaded portion can be estimated, and the volume can be estimated later depending on the width of the cutter tool. Provided the soil bulk density in g cm^{-3} is known (Saxton *et al.,* 1986),

the total mass of the removed soil can be estimated, for a range of expected soil textures. A blade width of 9-cm was used in this case, and was based on the constraints exhibited by the size of present and future deployable sensors.

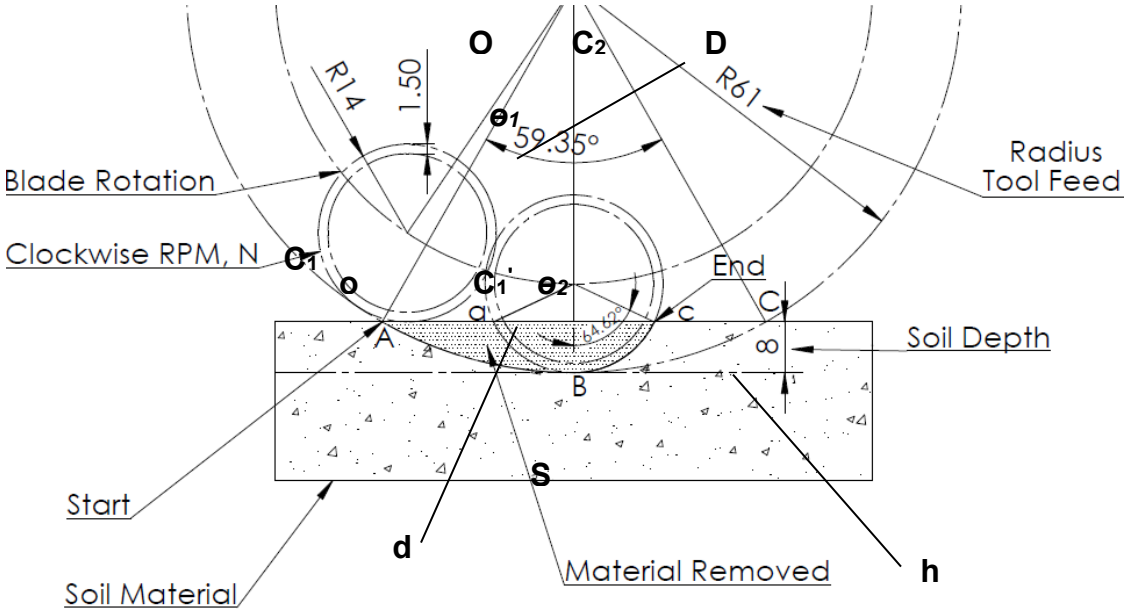


Fig. 3.4 The conceptual design of OSA

In concept C1 and C2 are the two virtual circles, but in practice some means of power is needed to drive those circles. To understand the power specifications of electrically driven geared motors, a small simulation was done to estimate the expected soil mass removed for the given geometry. As illustrated in Fig. 3.5, removal of dry soil mass was found to range between 3000-5000 g. However, in field conditions under varying soil texture and moisture levels it was found that removal of naturally moist soil may reach up to 13000 g. Considering the design requirements: compact design, light in weight and the need for 12V electrically powered components, removal of such large qualities of soil material all at once may not be a feasible option and therefore, a soil removing strategy was desired.

For example, if a 30 second time frame is allocated to expose soils up to 12 cm depths, then a minimum of 4 mm s^{-1} tool feed rate is required. Since only two cutting blades were used to allow space for sensor deployment, it was important to consider the cutting blade impact rate. On the

other hand soil is a complex material and based on its mineralogy, soil organic matter and moisture content, the physical properties may vary significantly in time and space.

Theoretically, it was very unpredictable to know the exact number of blade impacts required to break and slice a soil thickness of 4 mm and therefore, a trial and error approach was used to estimate the maximum RPM requirements of the cutter. In the initial design, a 150 RPM geared motor (5 impacts per turn) was found insufficient to expose saturated sandy and clay soils within the allocated time frame. A better performance was observed when using a 300 rpm geared motor (10 impacts per turn) and was further improved by selecting the cutter rotation counter clockwise to that of the feeder. Therefore, a strategy of feeding the tool in steps of 4 mm s^{-1} and by operating the cutter at up to 300 rpm was used.

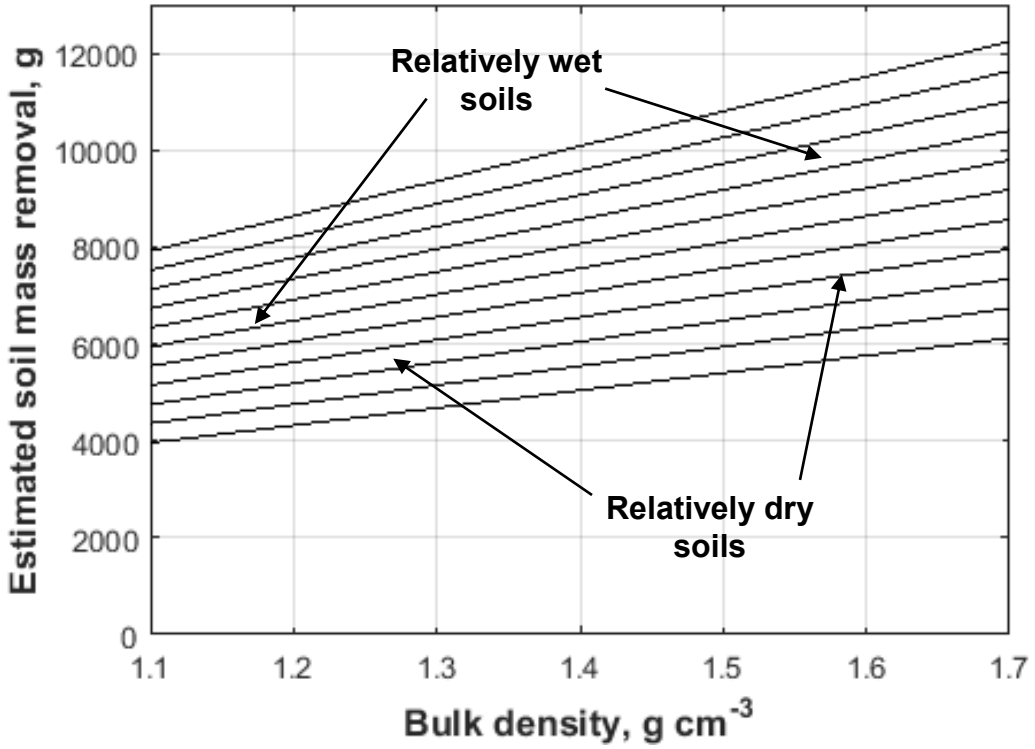


Fig. 3.5 Simulated results illustrating the potential ranges for soil material to be removed

This provided more time for the cutter to break, cut, and slice the soil before it could progress to the next depth. Finally, to provide optimum performance of the cutter in potentially varied soil moisture and textural environments, an additional torque was provided by adding another similar motor in series.

Using the conceptual design and trial and error experiments, the present prototype OSA system was developed as shown in Figures 3.6 and 3.7. It includes the following components: 1) soil preparation mechanism 2) sensor deployment mechanism, 3) electrode rinsing system, 4) sensors, 5) data acquisition and control system, 6) a guard and 7) supplementary and optional components. The system was designed for mounting on an ATV, SUV, or pickup truck using a standard hitch.

Fig. 3.8 illustrates the complete set of electro-mechanical devices used to operate the OSA mechanisms. The soil preparation mechanism was constructed using two cutting blades attached to opposite sides of two parallel discs rotated using a gear box (ToughBox Mini, AndyMark, Inc., Kokomo, Indiana, USA) driven by two 12 VDC powered motors (2655 rpm, 337 W, AndyMark, Inc., Kokomo, Indiana, USA). A gear ratio of 8.45:1 was adopted to produce a maximum torque of 20 N·m at 314 rpm. These discs were mounted using a special pivoted bracket. A 12 VDC powered linear actuator (45-cm stroke, Robotzone, LLC, Winfield, Kansas, USA) was used to move the pivoted part of the bracket to displace the discs up and down along a less than quarter-circle arch.

A hybrid control system consisting of relay drivers and relays, limit switches and servo-drivers, was used to match the speed of operation to different field conditions (Figures 3.9, 3.10 and 3.11).Due to the fact that the two discs were mounted using two separate short shafts, the space between the discs (width 10 cm and radius 5 cm.) could be used to deploy the sensors without retrieving the soil preparation mechanism.

To assure the blades were in their horizontal position when the rotation was stopped, an additional dedicated feedback control circuitry using two servo speed controllers (900 W, Talon-Cross the Road Electronics, AndyMark, Inc., Kokomo, Indiana, USA) and a continuous turn potentiometer (200 KΩ, Precision Sales and Equipment Inc., Newtown Square, Pennsylvania, USA) was implemented on a microcontroller platform (Arduino Uno, Interaction Design Institute, Ivrea, Turin, Italy). Figures 3.10 and 3.11 illustrate the schematics of the wiring diagrams used to control the two mechanisms.

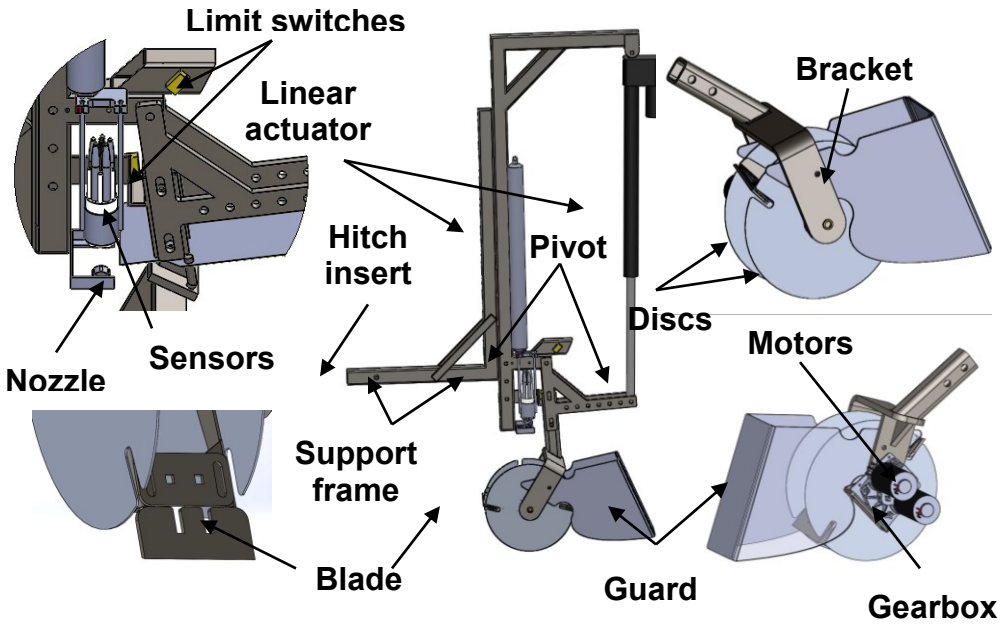


Fig. 3.6 On-the-spot Soil Analyser (OSA) in 3D

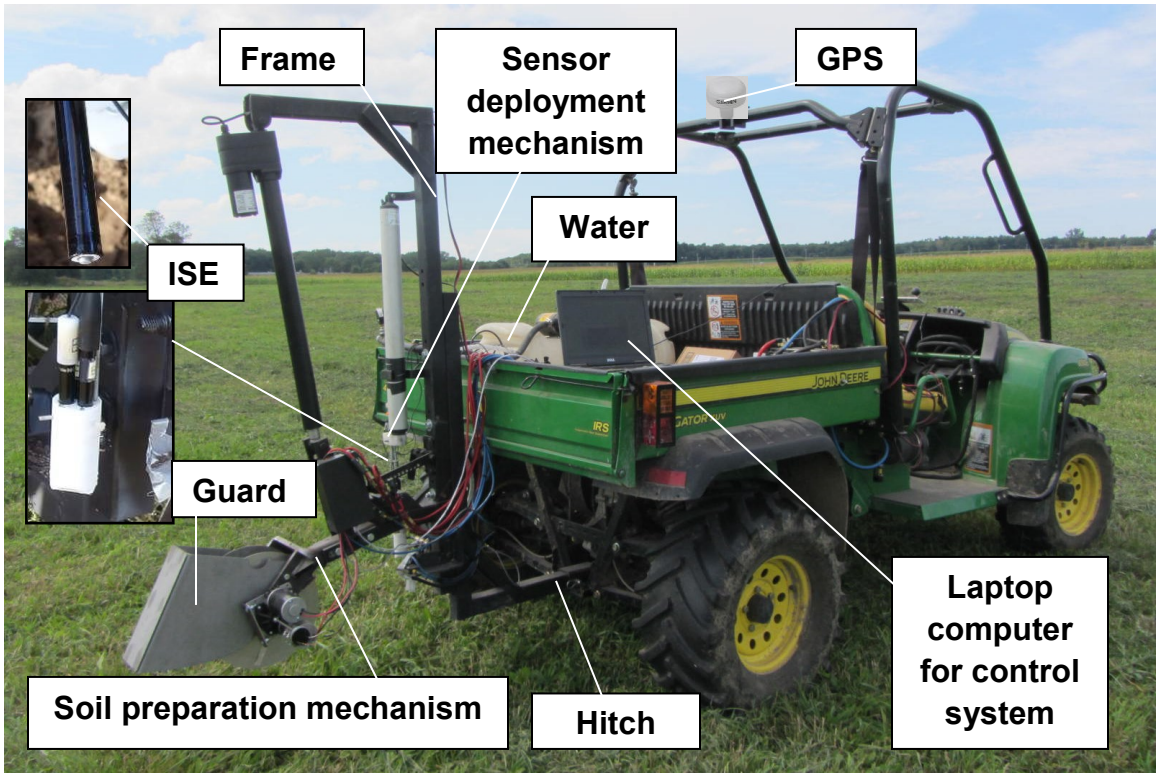


Fig. 3.7 Prototype On-the-spot Soil Analyser (OSA)

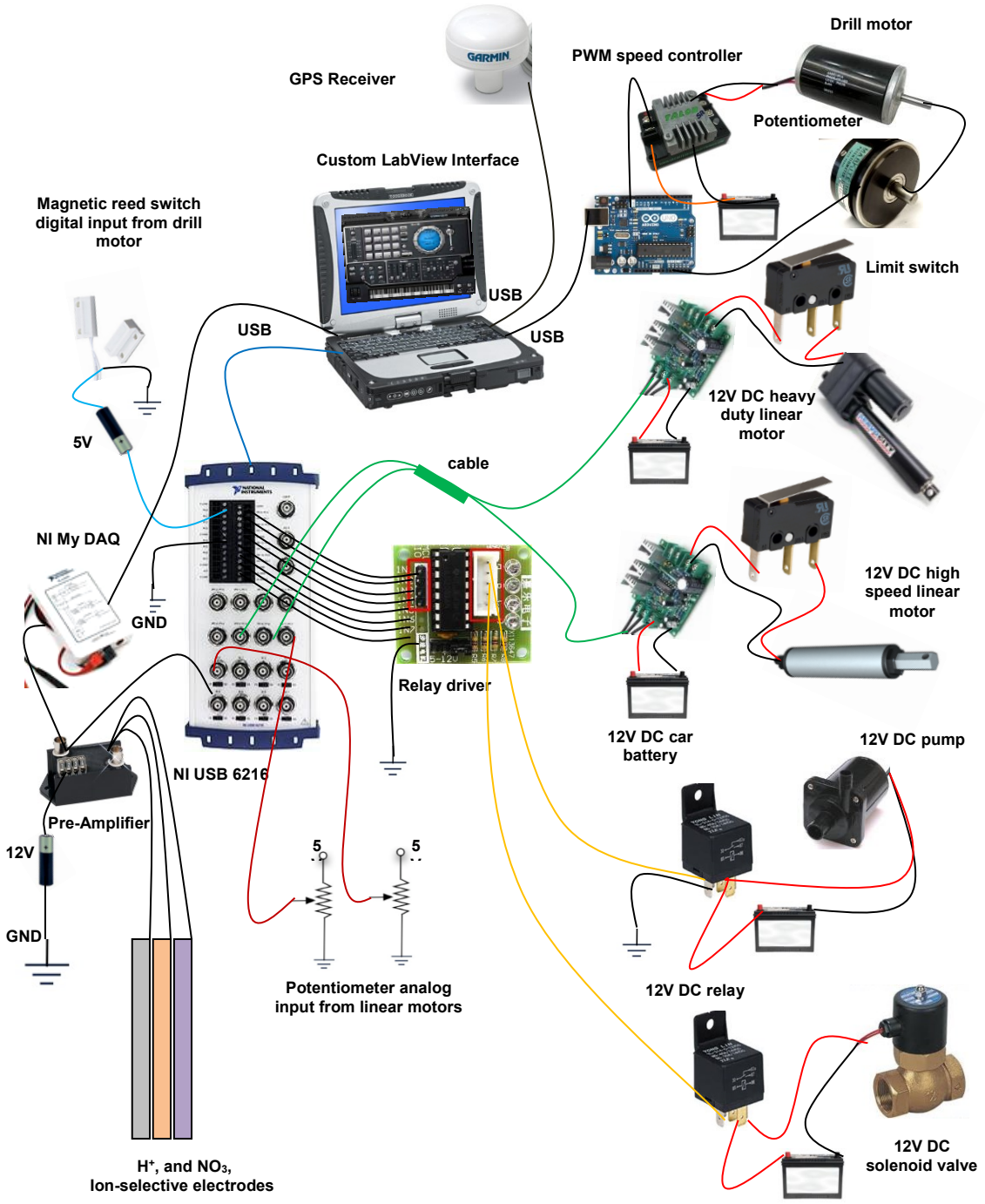


Fig. 3.8. The Key electronic components

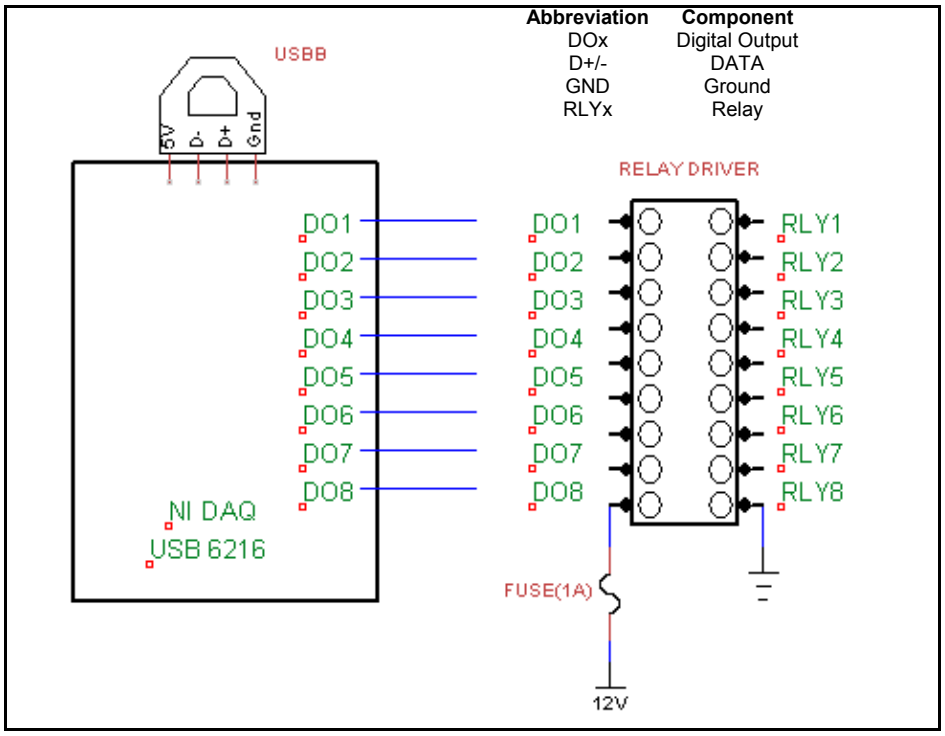


Fig. 3.9 DAQ card to relay-driver interfacing diagram

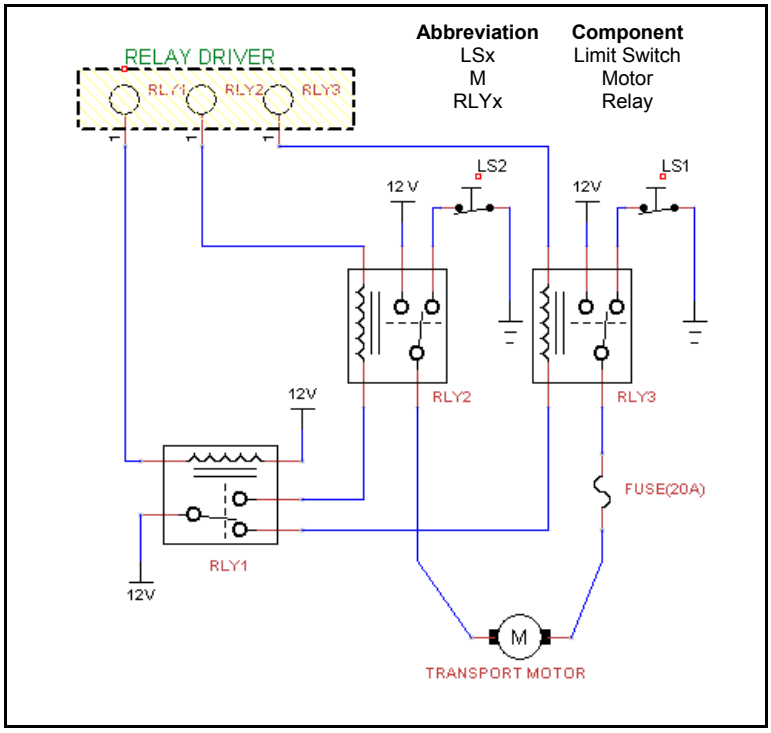


Fig. 3.10 Control wiring diagram of the transportation mechanism

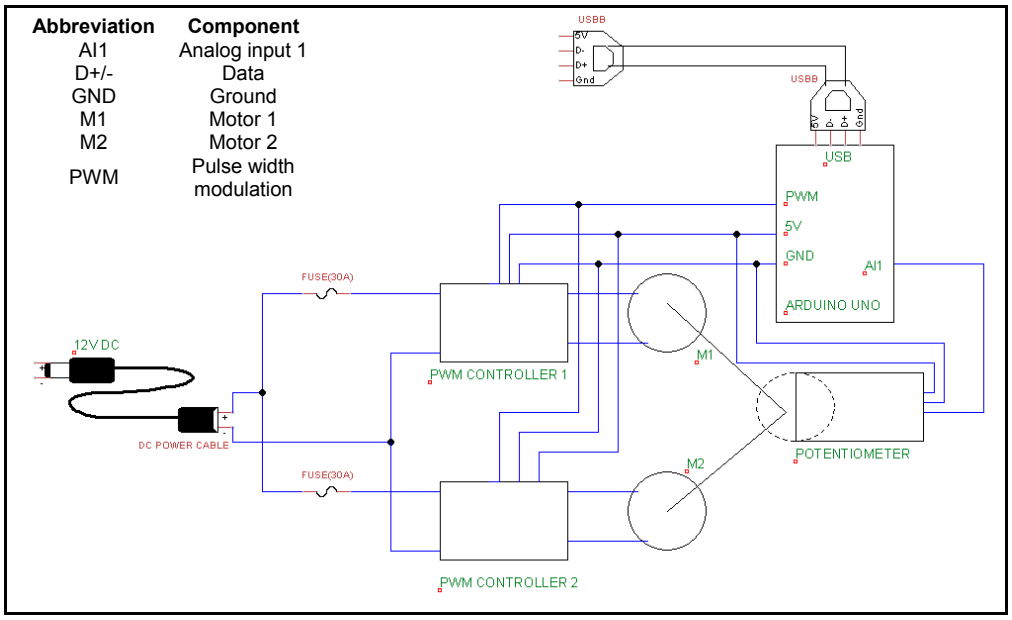


Fig. 3.11 Control wiring diagram of the soil preparation mechanism

Fig. 3.12 shows the schematics of the wiring used to control a high speed 12 VDC powered linear actuator (45 cm stroke, Progressive Automations, Richmond, British Columbia, Canada) which was used to vertically insert the sensor deployment unit between the two discs and retrieve it after the measurements (Fig. 3.13).

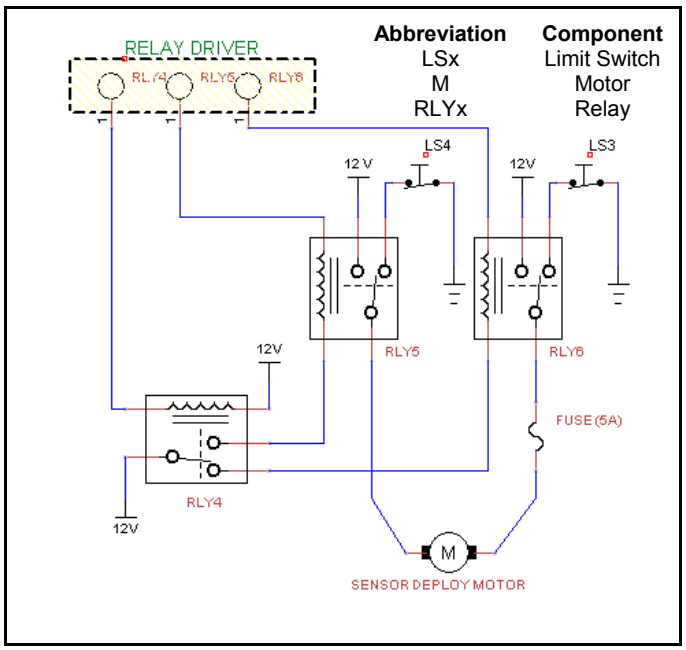


Fig. 3.12 Control wiring diagram of the sensor deployment mechanism

The sensor holding part of the mechanism matches the dimension of the sensor and in its current modification, it was designed to house three combination ion-selective electrodes with an outside diameter of 12 mm. The “gentle” action required to preserve the fragile sensor membranes was achieved using a spring mounting (< 30 g force).

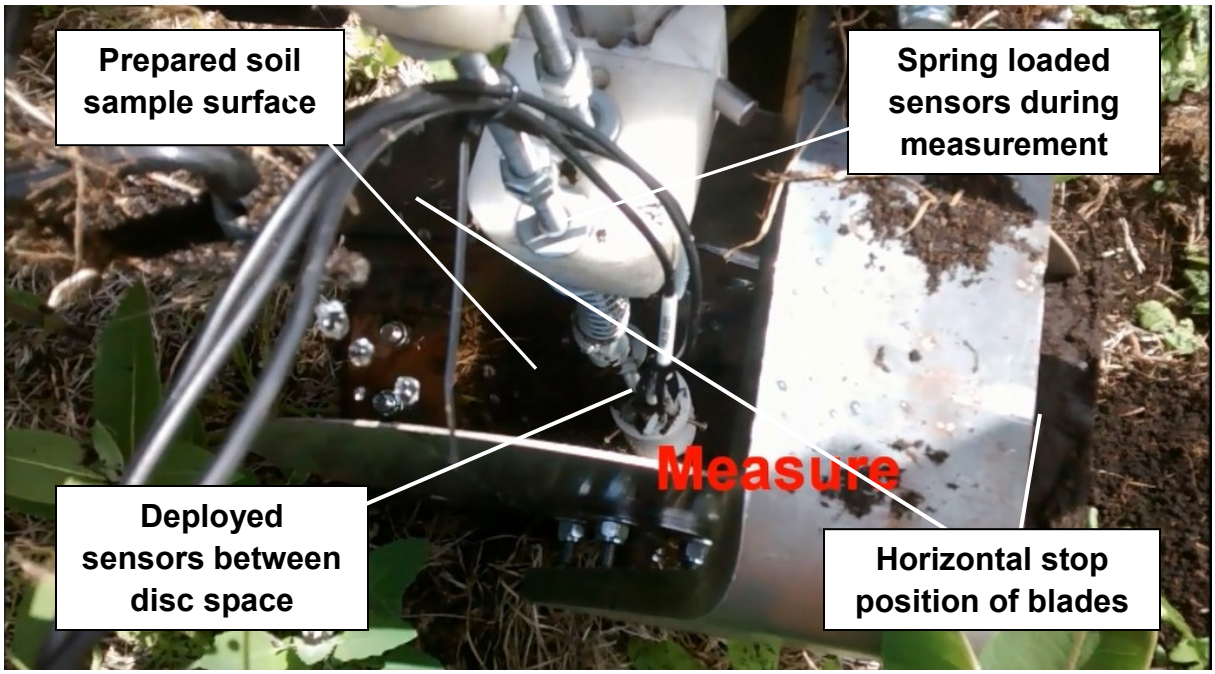


Fig. 3.13 Illustration of the deployed sensing array between the two discs

A flat surface-single junction glass pH electrode (Cole-Parmer, Chicago, Illinois, USA) and a antimony pH electrode (Veris Technologies, Inc., Salina, KS, USA) were used to measure soil acidity, while a nitrate ISE (detectIONTM 3021BN, Nico Scientific, Inc., Huntingdon Valley, Pennsylvania, USA) was used to measure soil NO_3^- activity. The sensor output signal impedances were matched to the input impedance specifications of the data acquisition hardware by connecting active unity gain pre-amplifier circuits (PHTX-22, OMEGA Engineering, Inc., Stamford, Connecticut, USA) in series.

To avoid sensor-reference cross interference induced by "ghosting", individual sensor analog outputs were measured using independent data acquisition cards. To avoid vehicle system induced noises, the pre-amplifiers were powered independently using a set of AA 12V battery packs. The detailed wiring diagram is illustrated in Fig. 3.14.

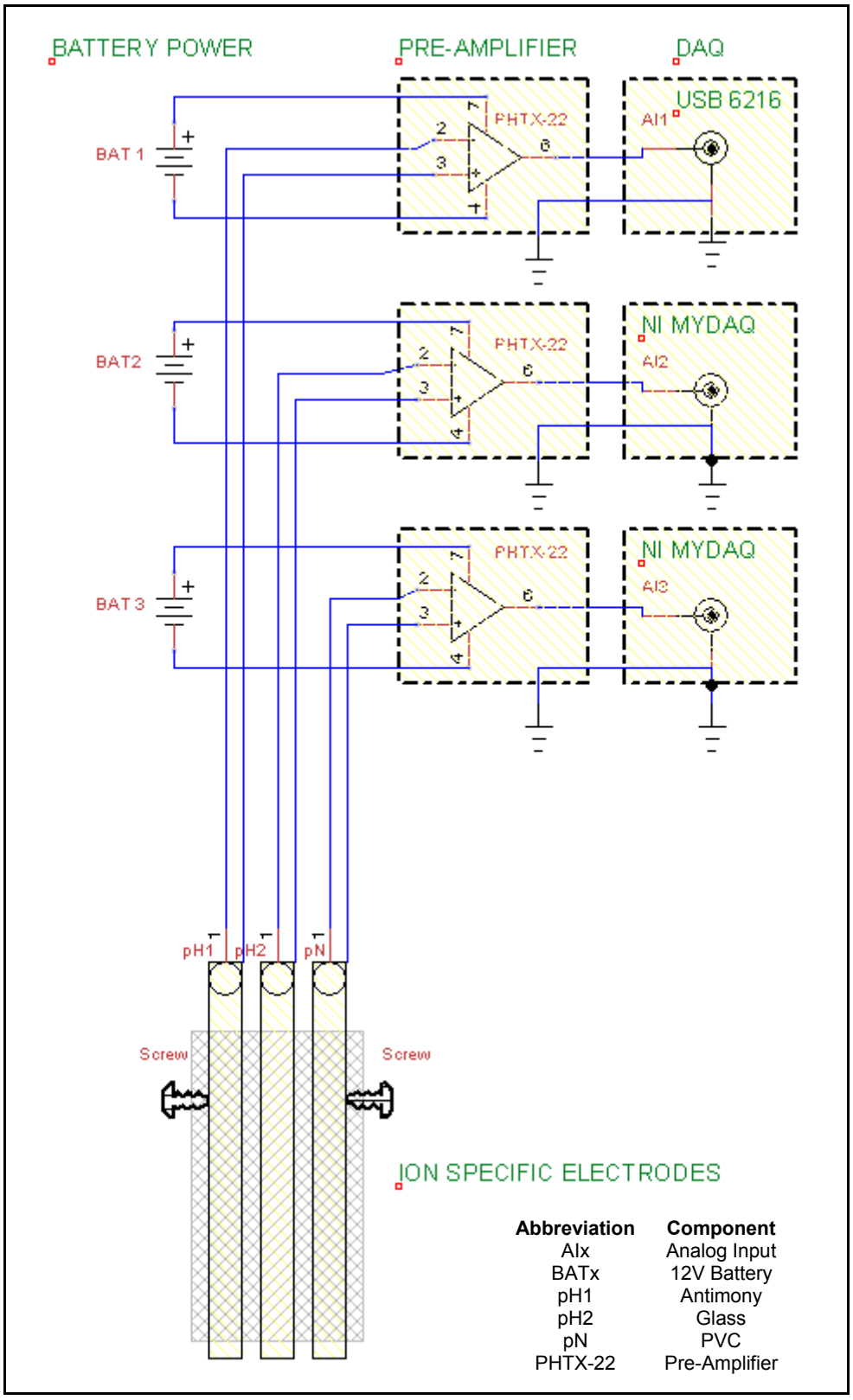


Fig. 3.14 Wiring diagram of the soil sensing array and the data acquisition unit

To restore the exposed surface area, a special guard as shown in Fig. 3.15 was designed. It pivoted freely and partially surrounded the side portions of the discs creating adequate space to accumulate the soil material that is removed during the soil preparation portion of the measurement cycle. The accumulated soil is displaced back to the exposed area when the measurement cycle is completed.

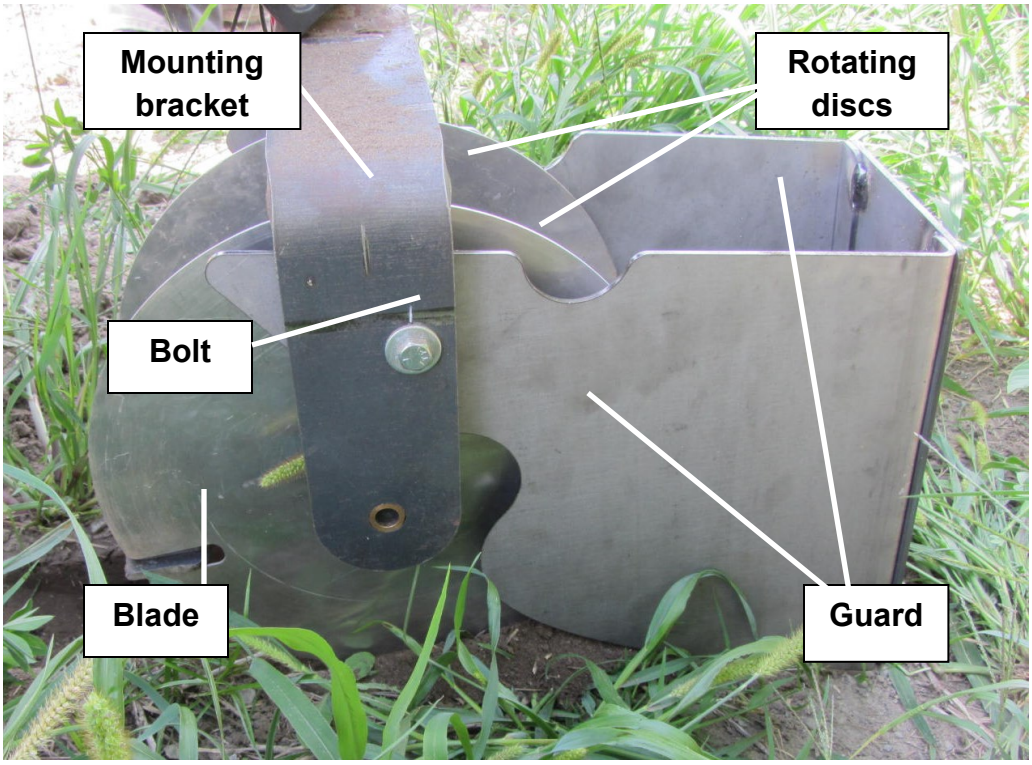


Fig. 3.15 Illustration of the guard mounted on the bracket

Two jet water nozzles were mounted on both sides of the system for washing the electrodes. A 50 L water tank was used to store the water and it was pumped using a 12 VDC power pump at designated times (to clean the electrodes and to moisten the soil when needed). The component wiring diagrams are shown in Figures 3.16 and 3.17.

The OSA operation software (Fig. 3.18) was developed using LabVIEW 12.0 software (National Instruments, Austin, Texas, USA). All data acquisition and control functions were performed through USB-6216 and MyDAQ data acquisition cards (National Instruments, Austin, Texas, USA).

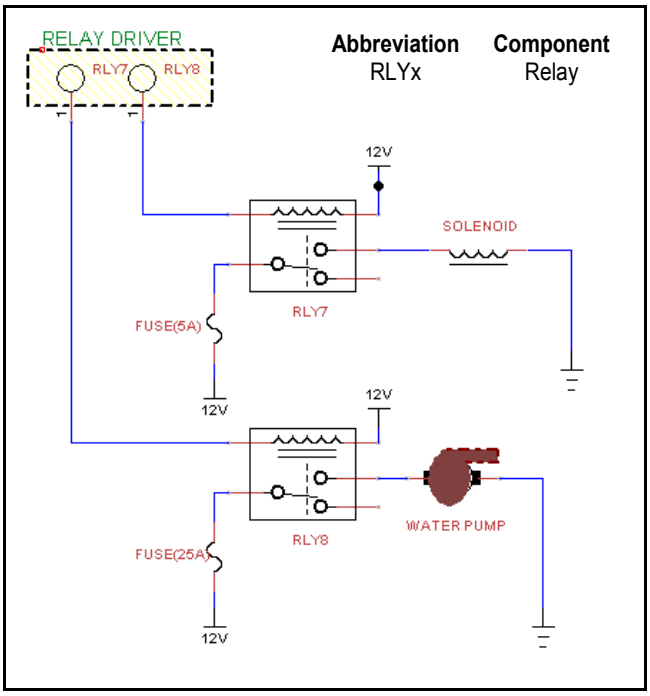


Fig. 3.16 Control wiring diagram of the sensor rinsing mechanism

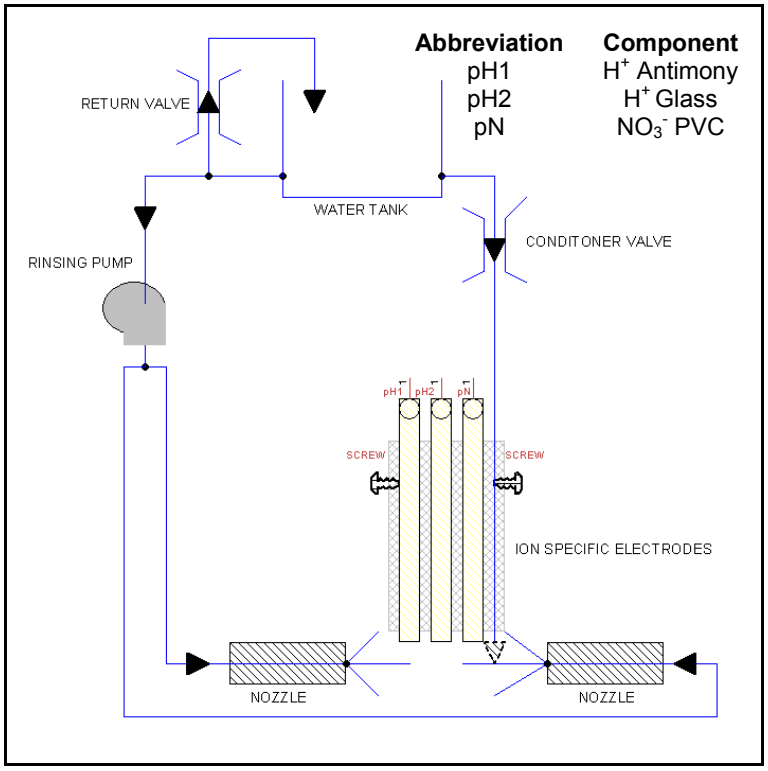


Fig. 3.17 The schematic of the water supplying unit

Based on the control sequence, when the OSA is brought to a new location and the cycle begins, the soil preparation unit is displaced downward while removing soil and crop residue until the discs with two blades are directly under the set of sensors. After the rotation action is stopped, the rinsing nozzles, or additional supply systems, spray water or a specific salt solution onto the exposed soil surface to be analyzed according to the needs of the test. When the soil surface is ready, all of the sensors are placed in contact with the soil and their responses are recorded and analyzed. Once the data acquisition system finds a strong predictability of the steady-state sensor output (Adamchuk *et al*., 1999; Adamchuk *et al*., 2005; Adamchuk *et al*., 2006) the resultant measurement is estimated, high-frequency signal noise is filtered out and the required parameters are calculated. The data are paired with geographic coordinates and time information obtained through a global navigation satellite system (GNSS) receiver and saved to a delimited text file. When the measurement data are secured, the sensors are retrieved and the soil preparation mechanism is raised slightly above the ground surface. The vehicle is moved forward to allow the guard to drag the accumulated soil material and refill the exposed area (Fig. 3.19). The system is then raised back to transportation mode and the electrodes are rinsed. The system is ready for the next measurement location.

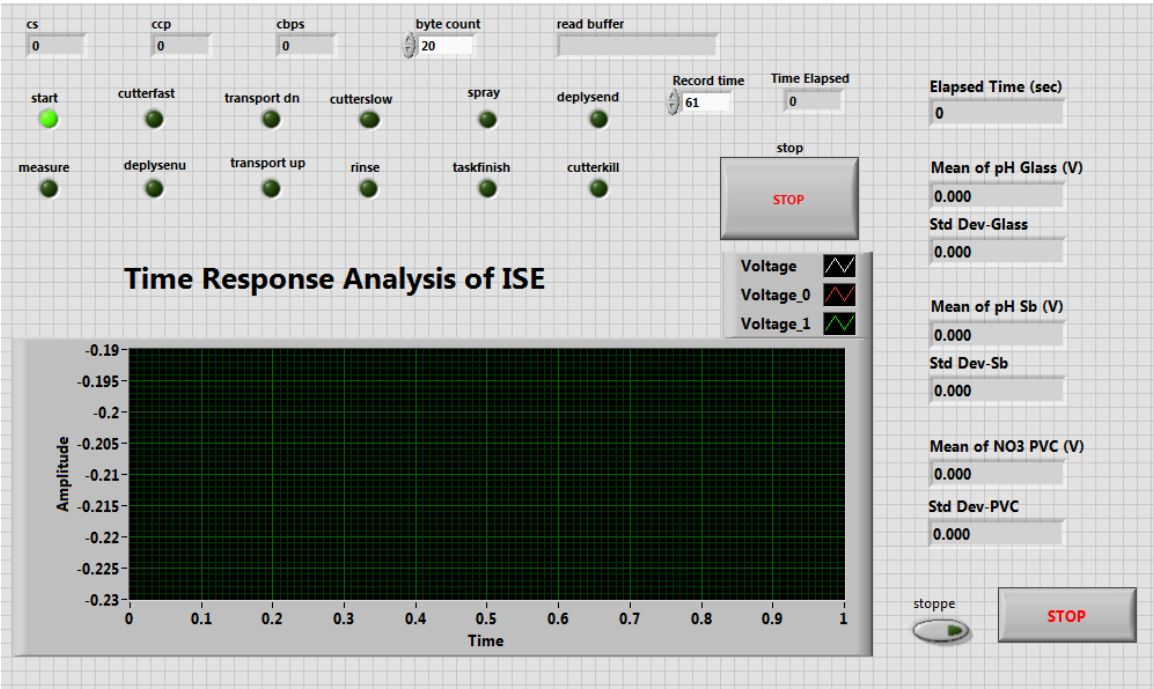


Fig. 3.18 Illustration of the data acquisition and control interface developed using LabVIEW

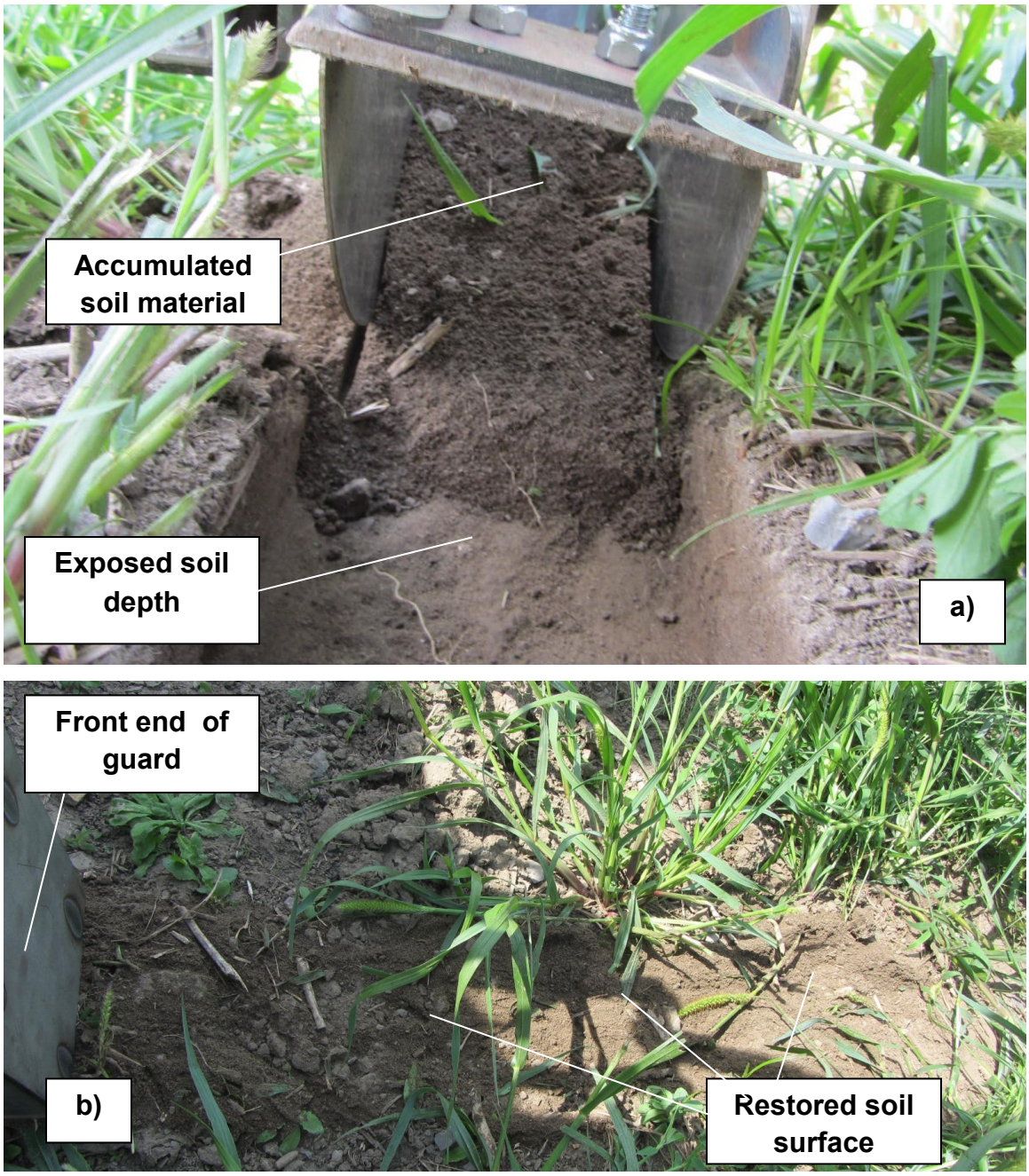


Fig. 3.19 Illustration of (a) the exposed soil surface and (b) its restoration using the guard

In its current modification, OSA can be adjusted for vehicles with hitches of varying heights and different measurement depths down to 12 cm (depending on the type of soil surface). Also, the space available for sensor deployment is not occupied completely, allowing for future expansion of the number and type of sensors, including optical soil reflectance, machine vision and capacitance sensors. We have recently tested a mid-infrared spectrometer and developed a digital

microscope-based sensor system and CO_2 gas analyser that could also be deployed using the OSA. This would allow for the simultaneous measurement of particle size distribution, soil organic carbon content, and the level of biological activity.

The sensor data flow structure is open to the capability to feed a future automatic navigation algorithm, which is the ultimate goal of this research. In the case of a completely automated vehicle (similar to the widely discussed Mars rover), the system would be able to use real-time estimates of soil attributes of interest to modify the travel path and optimize the number of measurements in order to better characterize the spatial distribution of these properties and to provide the best quality information in the shortest possible time.

3.2.3 System Evaluation

A preliminary system evaluation was conducted to assess the system's control functionalities. It was impossible to make quality comparisons of the measurements obtained with the OSA and in the laboratory due to a relatively early start of the subfreezing season. While constructing the prototype, two quick tests: one for pH and another for nitrate were performed using the instrumented manual probe (Fig. 3.2) equipped with a FieldScout pH 100 meter (Spectrum Technologies Inc., Aurora, Illinois, USA) at various research plots of the Lods Research Centre (McGill University, Ste-Anne-de-Bellevue, QC, Canada).

To assess performance of a nitrate detectION™ 3021BN (Nico Scientific, Inc., Huntingdon Valley, Pennsylvania, USA) ISE, measurements were performed in an experimental canola field divided into sixteen plots, treated with different levels of urea (0, 50, 100 and 150 kg N ha^{-1}). Two months after planting, random in situ measurements were taken 6-8 cm below the soil surface using the instrumented manual probe (Fig. 3.2). The electrode was calibrated using standard potassium nitrate (KNO_3) solutions with known NO_3^- concentrations. A simple linear regression method was applied to determine the relationship between the per plot averages of the electrode outputs and the NO_3^- concentrations measured using corresponding composite (five cores) 15-cm deep soil samples. These samples were analyzed to quantify NO_3^- concentrations in the laboratory, using a potassium chloride (KCl) extraction procedure (using a ratio, 1:10) and a

cadmium reduction colorimetric technique. Estimates of soil water content, sampling depth and soil bulk density were used to bring all the measurement to mg kg^{-1} units.

To asses the performance of an antimony pH electrode (Veris Technologies, Inc., Salina, KS, USA) measurements were performed on 10 random locations of seven different research plots with suspected variability in soil pH at the end of the crop harvesting season. The instrumented manual probe (Fig. 3.2) was used to make measurements 6-8 cm below the soil surface. After collecting the measurements, the electrode was calibrated using standard buffer solutions (pH 4 - SB101 and pH 7 - SB108) obtained from ThermoFisher Scientific (Waltham, Massachusetts, USA). In addition, physical soil samples were obtained as close to the manual probe measurements as possible. These physical soil samples were analyzed to quantify soil pH levels in the commercial laboratory using standard laboratory procedure (Watson and Brown, 1998) (measurement of soil pH in 1:1 weight-to-weight soil-water solution). A simple linear regression was performed to compare results with those from a commercial soil laboratory.

The system was designed in such a way that it can be operated from the standard hitch of a pickup, an all-terrain vehicle, or another platform. If such a platform is autonomous, measurements are conducted without an operator's involvement. Measurement locations and density can be either predefined or adjusted in real time to respond to the quality and variability of the measurements already obtained at a given site. The system can be used by individuals and organizations involved in small-scale soil mapping. Relatively inexpensive and easy to use, the system will be attractive to agribusinesses involved in soil fertility management. The low destruction approach will make this method applicable to specialty and permanent crop environments. An option to conduct measurements automatically can assist in obtaining data in hazardous areas without the need for personnel.

3.3 Results and Discussion

The system was able to remove approximately 8 cm (3.1 in) of topsoil under various field conditions (Fig 3.20), which provided a suitable surface for electrode deployment. The

electrodes could be brought into contact with the soil and rinsed well afterwards. The entire cycle could be performed within a 60 s timeframe. According to Fig. 3.21 both NO_3^- and pH electrodes revealed major differences in corresponding soil properties between different research plots and should be suitable for deployment using the developed OSA. Other sensors, including an array of portable optical and radiometric instruments, should also be appropriate to be used with this system.

Fig. 3.20 Test of the soil preparation mechanism

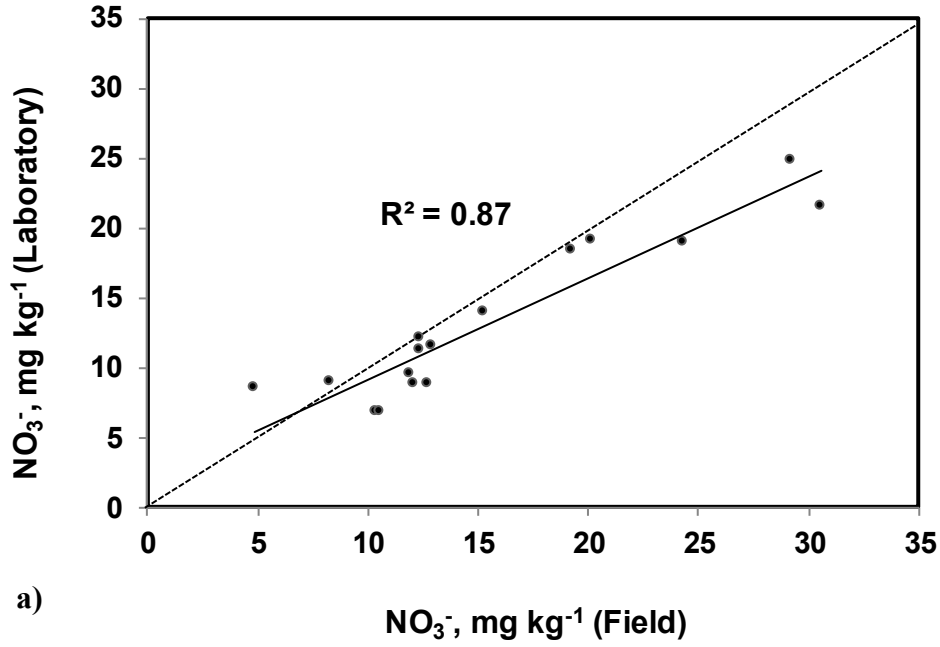


Fig. 3.21 Field tests results: (a) 2011- nitrate ISE

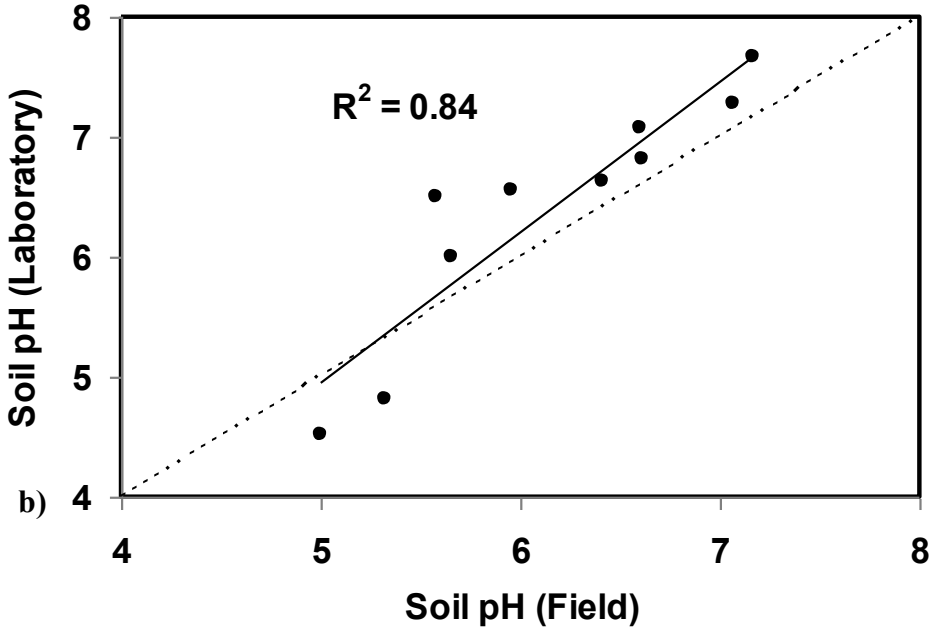


Fig. 3.21 Field tests results: 2013 - pH ISE

3. 4. Conclusions

Measurement of soil spatial variability provides important information used for soil cultivation, reclamation, remediation and other practices. Soil sampling and *ex situ* analysis is the most popular approach currently in use. Proximal soil sensing is an emerging technology resulting in on-the-go mapping of soil characteristics. Lower costs have been one of the main advantages for on-the-go sensing methods when compared to on-the-spot solutions. The latter approach has a number of technical advantages and the use of robots can make on-the-spot measurements even more attractive. Technology for mapping soil pH on-the-go and manually activated on-the-spot measurements of soil pH are both commercially available. The on-the-spot soil analyser presented in this study illustrates a new mechanism that enables rapid and reliable soil manipulation to facilitate the necessary interface between sensors and soil in an automated mode. Furthermore, when compared to existing solutions, this method is applicable in difficult soil conditions and enables sufficient opening of a subsurface soil area to engage an array of proximal soil sensors. For example, this system can carry several ion-selective electrodes (pH, nitrate, potassium, phosphorus, etc.), along with a miniature camera for machine vision analysis and a spectroradiomer for colorimetric analysis. During each measurement, this mechanism

removes a predefined layer of soil and crop residue and gently brings the sensors into contact with the prepared soil surface. When all measurements are stabilized, they are stored along with geographic coordinates and time of each measurement. After retrieving and rinsing the sensor surfaces, the system can be moved to a different location for new measurements.

Connecting Text to Chapter 4

Chapter 4 is related to the second objective of this study listed in Chapter 1. As shown in Chapter 3, the OSA platform in its current modification makes use of potentiometric sensors to simultaneously determine soil pH and NO_3^-. As discussed in Chapters 1 and 2, optical sensors based on the diffuse reflectance spectroscopy principle can be a complementary choice for extending the suite of deployable sensors on OSA. As an example, soil diffuse reflectance measurements collected in laboratory using two different field deployment ready vis/NIR/mid-IR hyperspectral instruments were evaluated for predicting percentages of sand, clay and soil organic carbon. The study was performed on a large archived data set consisting of 282 soils originating from agricultural sites in four Canadian provinces. Promising results for measuring soil texture in laboratory conditions indicated the potential usage of both systems. The results of this study were presented at conferences shown below and at the time this thesis was submitted, a manuscript has been submitted to the "Journal of Near Infrared Spectroscopy" as follows:

Dhawale, N.M., V.I. Adamchuk, S.O. Prasher, R.A. Viscarra Rossel, A.A. Ismail, J.K. Whalen, & M. Louargant. 2014. A comparative study between a field deployment ready vis/NIR and prototyped mid-IR hyperspectral instrument. *Journal or Near Infrared Spectroscopy* (under review).

Dhawale, N.M., V.I. Adamchuk, S.O. Prasher, R.A. Viscarra Rossel, A. A. Ismail, J.K. Whalen, & M. Louargant. 2014. Comparing vis/NIR/mid-IR hyperspectrometry for measuring soil physical properties. *Paper No. 141909453.* St. Joseph, Michigan: ASABE.

Herzellah, S.**, N**. **Dhawale**, H. He, J. Whalen, V. Adamchuk, S. Prasher, S. Rintoul, D. Pinchuk, J. Sedman, & A. Ismail. 2013. Comparative assessment of Visible/Near-Infrared and Mid-Infrared reflectance techniques for the rapid analysis of soil texture. *In: poster of Pittcon Conference and Expo*. 2013. Philadelphia., PA, USA.

Chapter 4

A COMPARATIVE STUDY BETWEEN A FIELD DEPLOYMENT READY VIS/NIR AND A PROTOTYPED MID-IR HYPERSPECTRAL INSTRUMENT

Dhawale, N., V. I. Adamchuk, S.O. Prasher, R.A. Viscarra Rossel, A. A. Ismail,
J. K. Whalen, M. Louargant

Abstract

Rapid developments in semiconductor technologies have improved the robustness and reduced the cost, size and complexity of hyper-spectral instruments; hence, they are now suitable for deployment in the field for the proximal sensing of various soil physical and chemical attributes. The goal of this research was to compare two field deployment-ready spectrometers. The first was a prototyped, mid-infrared (mid-IR) spectrometer operating between 5,520 and 11,126 nm. The second was a commercial, dual spectrometer system operating in both visible (vis) and near-infrared (NIR) regions between 342 and 2,220 nm. A large archived set of 282 soil samples (collected on contrasting fields from four Canadian provinces) was used to represent an extensive variety of soil textures, varying from sand to clay loam soils with a substantial range of soil organic carbon (SOC). Conventional soil analyses were performed on these samples prior to taking diffuse reflectance spectral measurements using both instruments. The spectral data were transformed into optical density measurements and then pre-treated by mean centering (MC). Both data sets were randomly partitioned into calibration (70%) and validation (30%) sets. Partial least squares regression (PLSR) was used to develop spectral calibrations for predicting the percentage of sand, clay and SOC. Both vis-NIR and mid-IR spectra revealed similar results. However, clay was better predicted using vis-NIR and sand using mid-IR spectra. The highest coefficients of determination (R^2) were found for sand (0.82) and clay (0.82). The corresponding root mean squared error (RMSE) was 10% and 7%, respectively. The ability to accurately predict

SOC was not as well supported for the set of soils in this experiment with a root mean squared error of approximately 0.4%. Neither sensor demonstrated substantially better performance; nevertheless, the tested methods prove the usefulness of both portable vis-NIR and mid-IR spectrometers for predicting soil texture.

Keywords: proximal soil sensing, on-the-spot soil analyser, visible and infrared soil spectroscopy

4.1 **Introduction**

Soil reflectance spectroscopy (SRS) has been used for the rapid assessment of various physical and chemical soil properties, both in laboratory conditions and directly in the field. A spectrometric method employs the interactions of vis, NIR and mid-IR radiation with the sample under investigation. In general, any vis, NIR, and mid-IR system is based on the sample's absorption of electromagnetic radiation at wavelengths in the range of 400-25,000 nm (25,000-400 cm^{-1}), where intense fundamental molecular frequencies related to soil components occur between wavenumber 4,000 and 400 cm^{-1}. Weak overtones and combinations of these fundamental vibrations due to the stretching and bending of NH, OH and CH groups dominate the NIR (700-2500 nm) and electronic transitions the vis (400-700 nm) portions of the electromagnetic (EM) spectrum (Viscarra Rossel *et al.,* 2006b).

A number of studies to determine soil attributes have resulted in the successful use of vis and NIR diffuse reflectance spectroscopy (DRS) to quantify particle size distribution, soil moisture, SOC, total carbon (TC) and total nitrogen (TN) simultaneously (Dalal and Henry, 1986; Chang *et al.,* 2001; Christy, 2008). Among various soil physical and chemical properties; particle size distribution, soil water content (SWC), soil organic matter (SOM), soil organic carbon (SOC) and TN, are the primary properties for which vis-NIR predictions have a theoretical basis, and cation-exchange capacity (CEC), electrical conductivity (EC), pH, potentially mineralizable N, specific surface area, wet aggregate stability, enzyme activities, microbial respiration and microbial biomass, are the secondary properties that are NIR spectroscopic predictable because of their correlation with certain primary properties (Chang *et al.,* 2001).

Infrared (IR) spectroscopy is more advantageous than vis and NIR since it is more sensitive to both organic and inorganic phases of the soil, thus, making their use promising in the agricultural and environmental sciences (Haberhaue *et al*., 1999; Janik *et al.,* 1995; Janik and Skjemstad, 1995, McCarty *et al.,* 2002; Masserschmidt *et al.,* 1999; Nguyen *et al.,* 1991). Soil analysis using mid-IR techniques was reviewed by Janik *et al.,* (1998) and the published applications of vis, NIR and mid-IR spectroscopy in soil analysis for the determination of primary and secondary properties were summarized by Viscarra Rossel *et al.* (2006b).

Numerous studies support the use of either type of spectroscopy for the simultaneous assessment of primary and secondary soil properties on-the-go (Christy, 2008; Reeves III, 2010). As an alternative to on-the-go technologies, on-the-spot measurements can be made where the field surface coverage does not allow for continuous engagement between soil and the moving parts of the sensor system (e.g. pasture). To explore this approach, a ruggedized platform carrying vis-NIR spectroscopic instruments to the field has been developed (Veris®P4000, Veris Technologies Inc., Salina, Kansas, USA). The advantage of the P4000 system is its ability to measure vis-NIR spectra, soil electrical conductivity (EC) and cone index (CI) for soil profiles down to a 1 m depth.

Four recent studies reported the successful use of the vis-NIR instrument of P4000 for predicting soil texture, SOM, SOC and plant available phosphorus ($P_{Mehlich-3}$), (Hodge and Sudduth, 2012; Dhawale, *et al.*, 2013b; Wetterlind *et al.,* 2013; Pikki *et al.,* 2014). In the first study, comparable results for SOC predictions were found while using *in situ* vis-NIR obtained using P4000, versus *ex situ* spectra obtained using a laboratory-grade vis-NIR instrument. In the second study, $P_{Mehlich-3}$ was predicted using *ex situ* vis-NIR spectra obtained using P4000 against conventional laboratory measurements with an R^2 of 0.85 and a standard error of prediction between 28 and 31 mg kg^{-1}. The third and fouth study reported of predictions with root mean squared errors of 5.5% clay, 6.6% sand and 0.31% SOM.

Recently, an all-terrain vehicle (ATV) mounted On-the-spot Soil Analyser (OSA) was developed to deploy a suite of sensors for subsurface measurements of chemical and physical soil characteristics in a consistent and ergonomic manner (Adamchuk *et al.,* 2014). In the current system modification, an array of ion-selective electrodes to measure soil pH and NO_3^- content is deployed. A prototype mid-IR spectrometer is a great addition to the electrodes that will allow integrating sensors based on different measurement principles (Dhawale *et al.,* 2013a). However, applicability of this instrument to predict physical soil characteristics in comparison with the earlier mentioned P4000 unit needs to be assessed on a large number of diverse soil samples.

The objective of this research was to compare the results of percent sand, clay, and SOC predictions using a mid-IR and a Vis-NIR spectrometer deployable for *in situ* operation by

testing 282 archived soil samples collected in contrasting agricultural fields across four Canadian provinces.

4.2 Materials and Methods

4.2.1 Data Collection

A prototype portable mid-IR variable-filter-array (VFA) spectrometer manufactured by Wilks Enterprise Inc. (East Norwalk, Connecticut, USA) was used to obtain mid-IR diffuse reflectance spectra (Fig. 4.1). A linear variable filter (LVF) detector was mounted on a 128 pixel array above an optical window made of zinc selenide (ZnSe) and placed on the top centre inside the enclosure. The spectrometer consisted of eight pulsing sources and was capable of maintaining a constant distance between the measured soil surface and the detector. The average of eight pulsed reflectance measurements was recorded per scan.

In addition, the LVF array was hardware configurable to provide for flexibility of operation in either NIR or mid-IR ranges of the electromagnetic spectrum. For this project, the VFA was configured to the short mid-IR range (1811-893 cm^{-1}) (8 cm^{-1} resolution). The electronics, enclosed in the upper part of the spectrometer, was powered using a standard 12 VDC adapter and it was interfaced with a standard laptop computer through a USB 2.0 cable. The C^1 data acquisition software (CogniSolve, Inc., Montreal, Québec, Canada) was used to record and store soil spectral reflectance measurements. The data collection process involved spreading soil samples in Petri-dishes and placing the mid-IR instrument over the sample with area ~707 mm^2 and for approximately 32 s. To minimize instrument noise, each spectrum was recorded as an average of 32 scans. The spectrometer was re-calibrated every 5 samples using a standard reference (a sheet of paper with a golden shade color) provided by the manufacturer.

The second instrument (part of the P4000) was a dual type spectrometer system, operating in the visible and near-infrared regions of the electromagnetic spectrum (Fig. 4.2). The first spectrometer USB2000 (OceanOptics, Dunedin, Florida, USA) was used to collect soil reflectance data between 342 and 1023 nm (resolution 5 nm) and the other spectrometer

Mini-Spectrometer (Model No. C9914GB, Hammatsu Photonics. K.K., Tokyo, Japan) measured between 1070 and 2220 nm (resolution 7 nm). The instrument included its own light source and was capable of maintaining a constant distance between the measured soil and the fibre optic probes using a designated sapphire window.

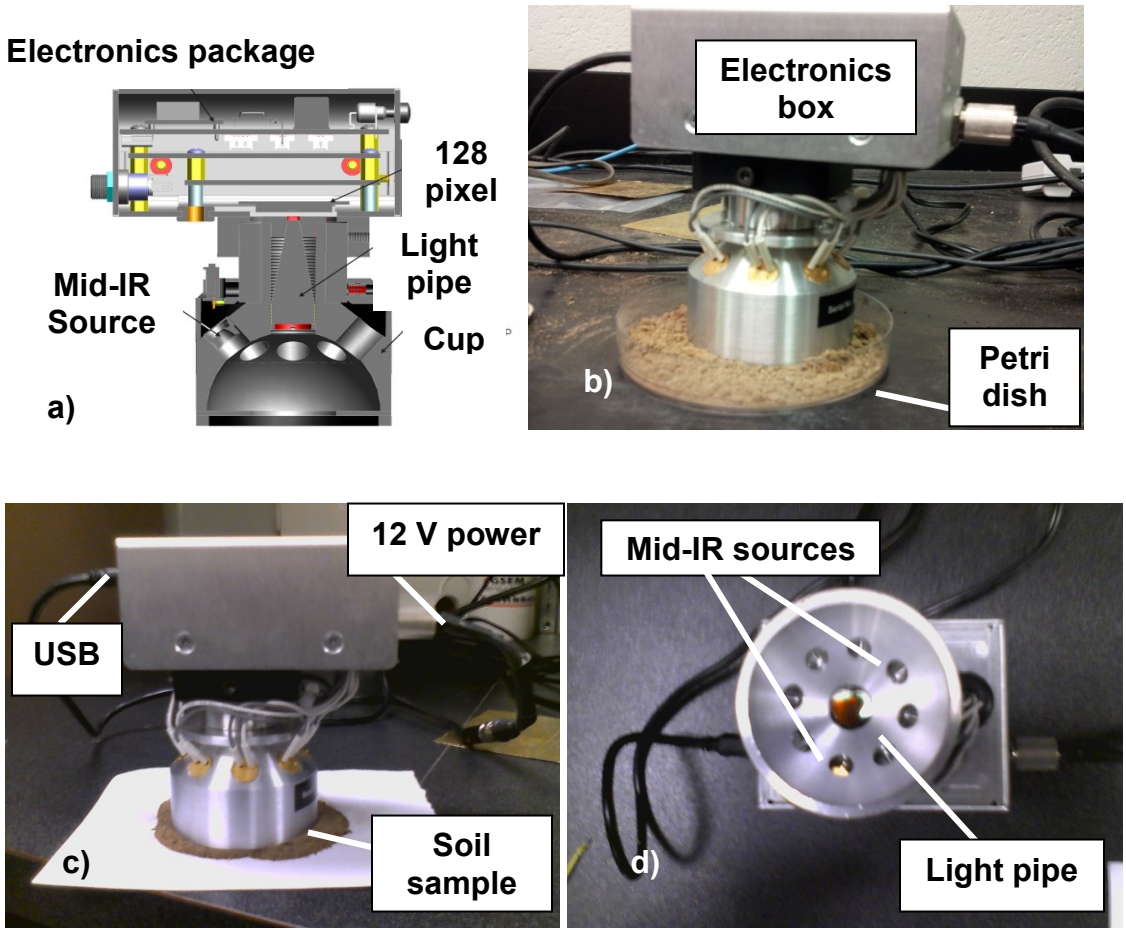


Fig. 4.1 Picture illustrating the portable mid-IR spectrometer in laboratory configuration

The data collection process involved filling <1 g of soil into a customized sample holder (radius of 5 mm and thickness of 5 mm) and placing it in contact with an optical window of area ~78 mm^2. At the beginning of each spectral measurement, the instrument was optimized and calibrated by measuring dark current followed by white reference measurements using the specially provided reference blocks (Fig. 4.2b). The instrument was re-calibrated every 20 samples. Soil spectra were interpolated to about 5 nm of spectral resolution, yielding a total of 380 data points (wavelengths) per spectrum. To minimize the instrument noise, each spectrum was recorded as an average of 30-32 scans.

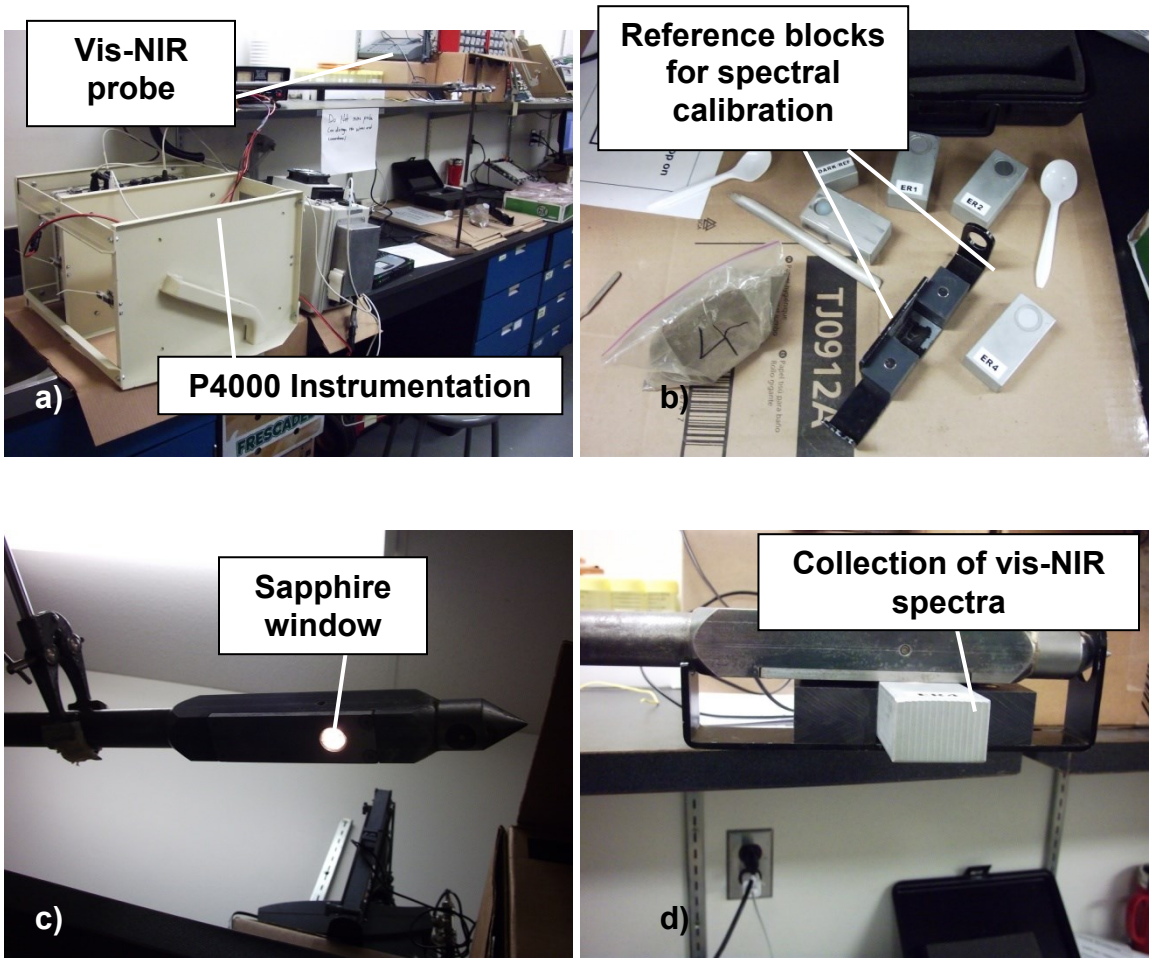


Fig. 4.2 Picture illustrating the portable vis-NIR spectrometer in laboratory configuration

A total of 282 soil samples were archived during two studies conducted from 2000 to 2007 and these soil samples were taken from experimental plots in four Canadian provinces with humid soil moisture regimes (St.Luce *et al.*, 2012). Sixty-nine percent (n = 195) of the samples were from a three year study (2007-2009) conducted on 43 sites located on research farms and in farmers' fields in the major corn growing regions of Québec (Nyiraneza *et al.,* 2010). Soil samples were collected from 23, 23, and 15 sites in 2007, 2008, and 2009, respectively. The other soil samples (n = 87) were collected from trials having four or five replicates with corn production in four Canadian provinces: five sites in Québec, two in New Brunswick, one in Ontario and one in British Columbia. All soil samples were aggregates of 15-cm long cores collected in the spring before the application of fertilizer. They were air-dried and ground to pass a 0.25-mm sieve for SOC analyses. SOC for all soils was determined by dry combustion using a

CNS-1000 (Leco Corp., St. Joseph, Michigan, USA). Soil texture was determined using the hydrometer method (Gee and Bauder, 1986).

4.2.2 Data Processing

From both instruments, the raw spectral data were imported in ParLeS software (Viscarra Rossel 2008) (version 3.1, 2007, University of Sydney, Sydney, Australia) and pre-processed for further data analysis. The obtained mid-IR soil reflectance spectra had a short range with a coarse resolution of about 7 cm^{-1} (*i.e.* ~ 40 nm) between each wavenumber and consisted of a total of 128 diffuse reflectance measurements per spectra (Fig. 4.3).

The vis-NIR soil absorbance spectra (Fig. 4.4) exhibited a small step discontinuity between 1023 and 1070 nm caused by the transition from one spectrometer to the other. Spectra noise was filtered out by removing the noisy "tails" of each spectrum (342-409, 1014-1075, and 2206-2220 nm). This produced 363 original data points per spectrum. The resulting vis-NIR spectra were corrected for offset. Finally, when needed, the data pre-processing steps included transforming reflectance measurements into optical density measurements as well as applying a mean centering technique (MC) for scaling and centering.

Both vis-NIR and mid-IR spectra were randomly partitioned into two sets: a calibration set (70%) and a validation set (30%). As a result, soils with different textures were split between calibration and validation samples as shown in Fig. 4.5. A decent variety was also observed for percentage sand, clay and SOC contents within the examples as shown by Figures 4.6-4.8. A partial least squared regression (PLSR) with leave-one-out cross validation was used for spectral calibration against laboratory measurements.

The orthogonalised PLSR-1 algorithm (Martens and Naes, 1989) implemented in ParLeS was used. This is a bilinear regression technique that extracts a small number of latent factors, which are a combination of the independent variables of reflectance or absorbance (at spectral wavelengths or wavenumbers), and uses these factors to produce a regression for the dependent variables (de Jong and Kiers, 1992; Geladi and Kowalski, 1986).

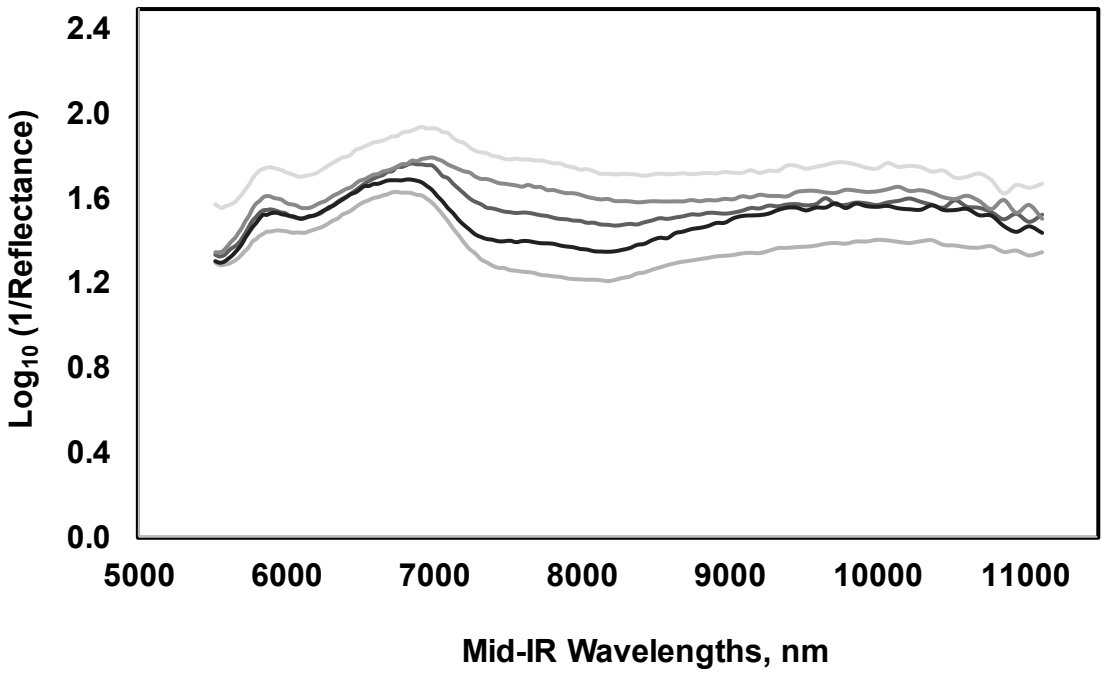


Fig. 4.3 Sample data illustrating potable mid-IR soil absorbance measurements

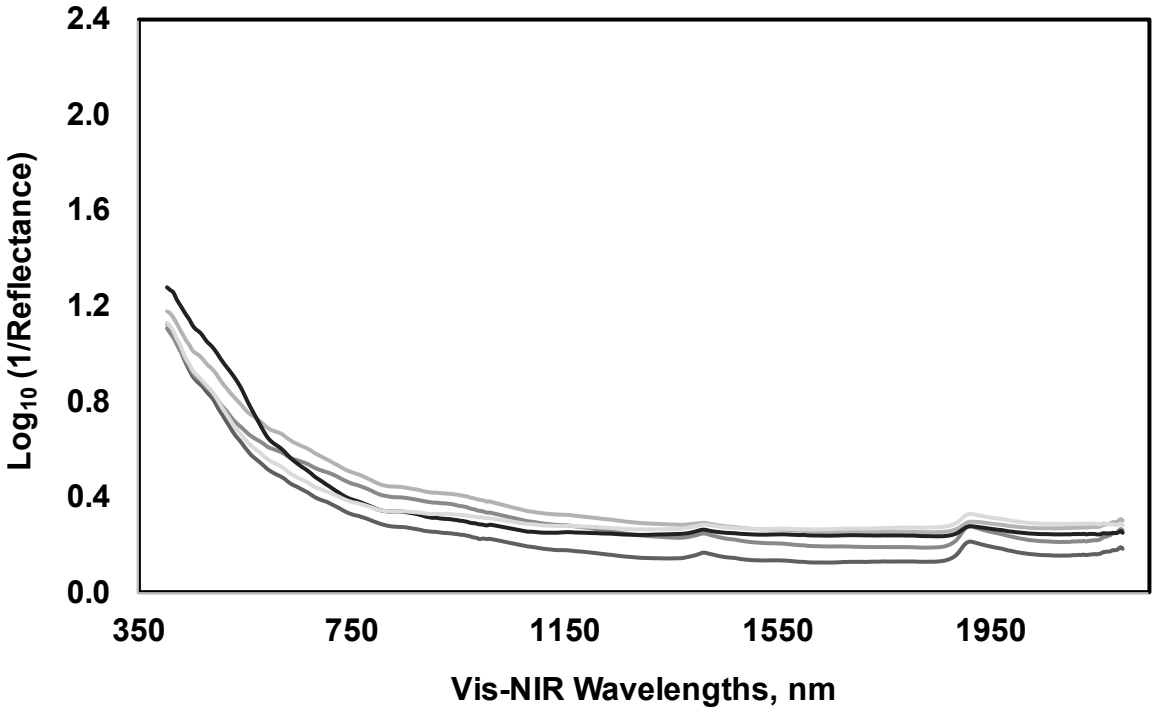


Fig. 4.4 Sample data illustrating vis-NIR soil absorbance measurements

The number of factors to use in each model was selected using leave-one-out cross validation on the calibration data. Then the selected model was applied to the validation set of soils. The performance of each model was evaluated using the coefficient of determination (R^2), root mean squared error (RMSE), the mean error (ME) and the standard deviation of the error distributions (SDE). The RMSE is a combination of both the SDE and the ME, where SDE represents a random error and ME indicates the bias, as follows:

$$R^2 = 1 - \frac{SSE}{SS_{yy}} \tag{4.1}$$

$$SS_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{4.2}$$

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y})^2 \tag{4.3}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2}{N}} \tag{4.4}$$

$$ME = \sum_{i=1}^{N} (\hat{y}_i - y_i) \tag{4.5}$$

$$SDE = \sqrt{\frac{\sum_{i=1}^{N} (\hat{y}_i - y_i - ME)^2}{N-1}} \tag{4.6}$$

where, SSE is the deviation of observation from their predicted values, SS_{yy} is the deviation of the observation from the mean, y_i is the laboratory unknown observed value of the interested soil property, $\hat{y}_i$ is the predicted value of the interested soil property, $\bar{y}$ is the mean of observation values and N is the sample size.

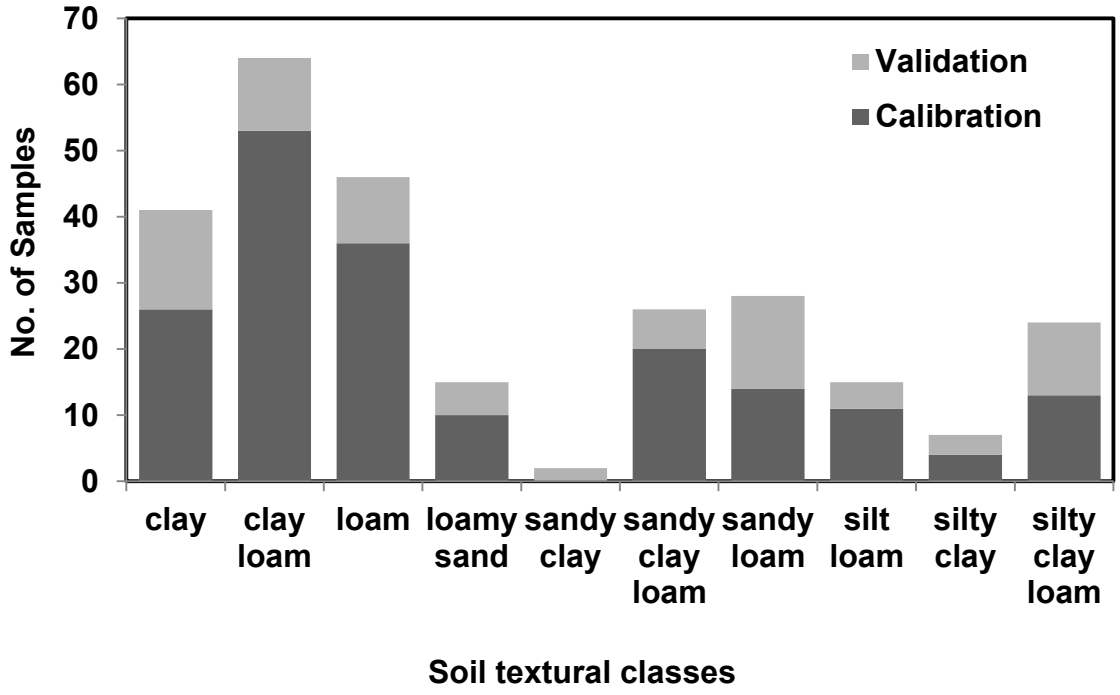


Fig. 4.5 Distribution of textural soil characteristics

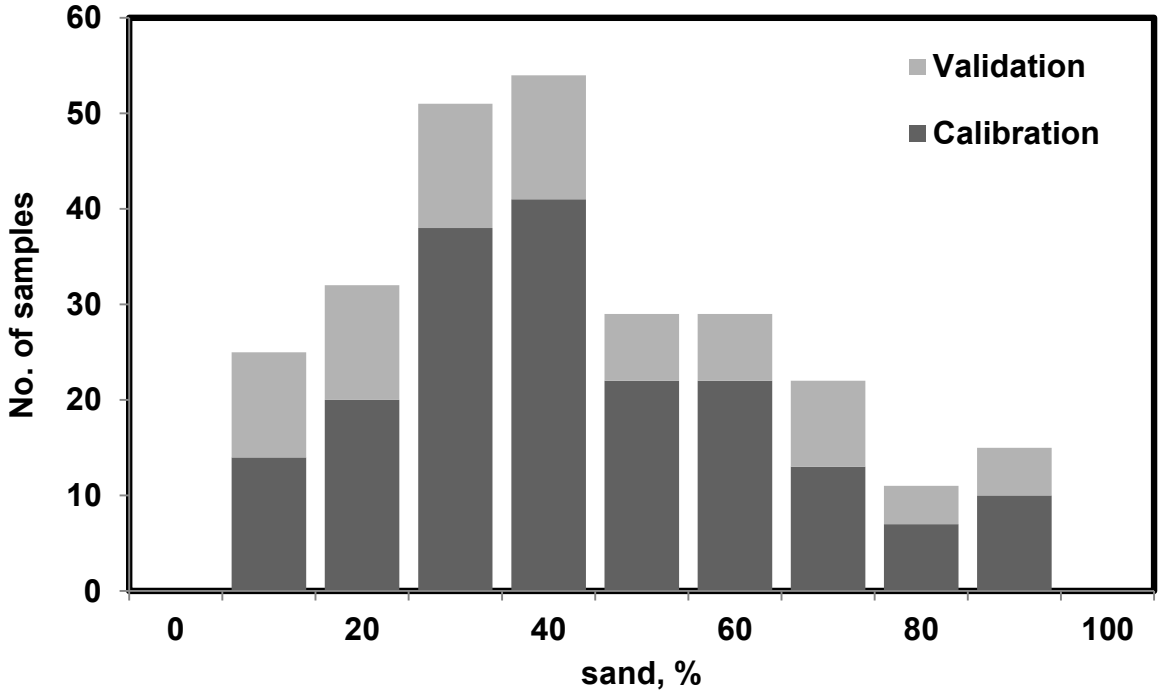


Fig. 4.6 Distribution of % sand content

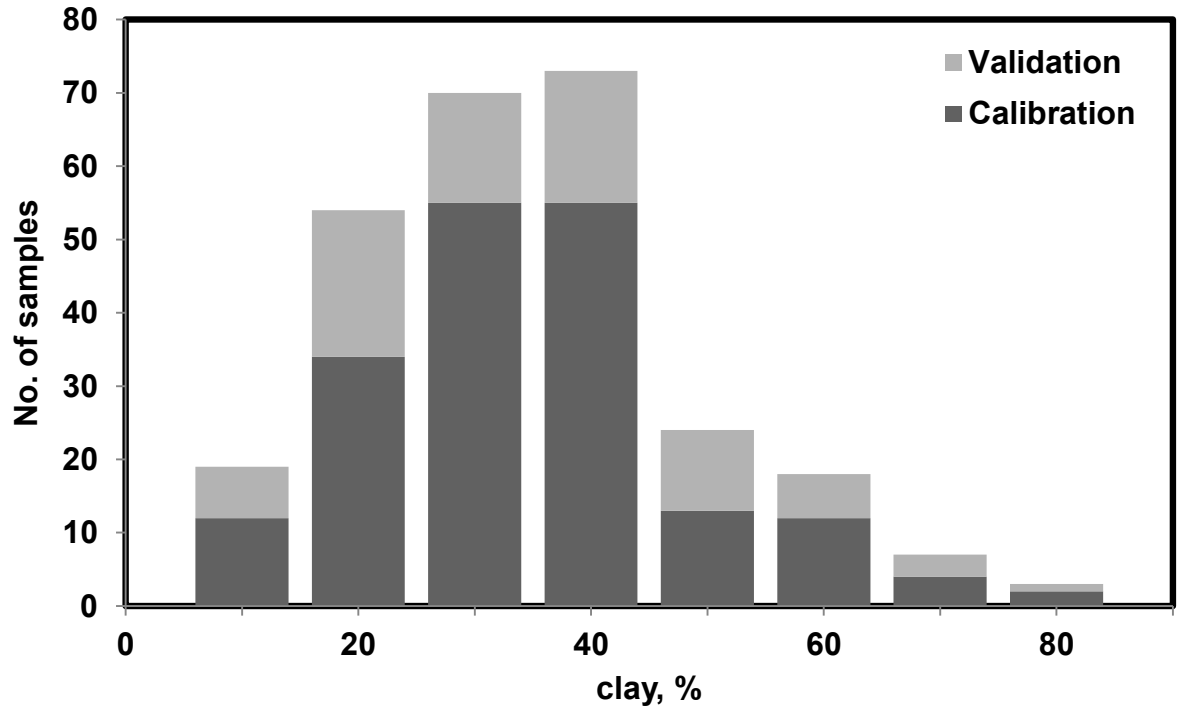


Fig. 4.7 Distribution of % clay content

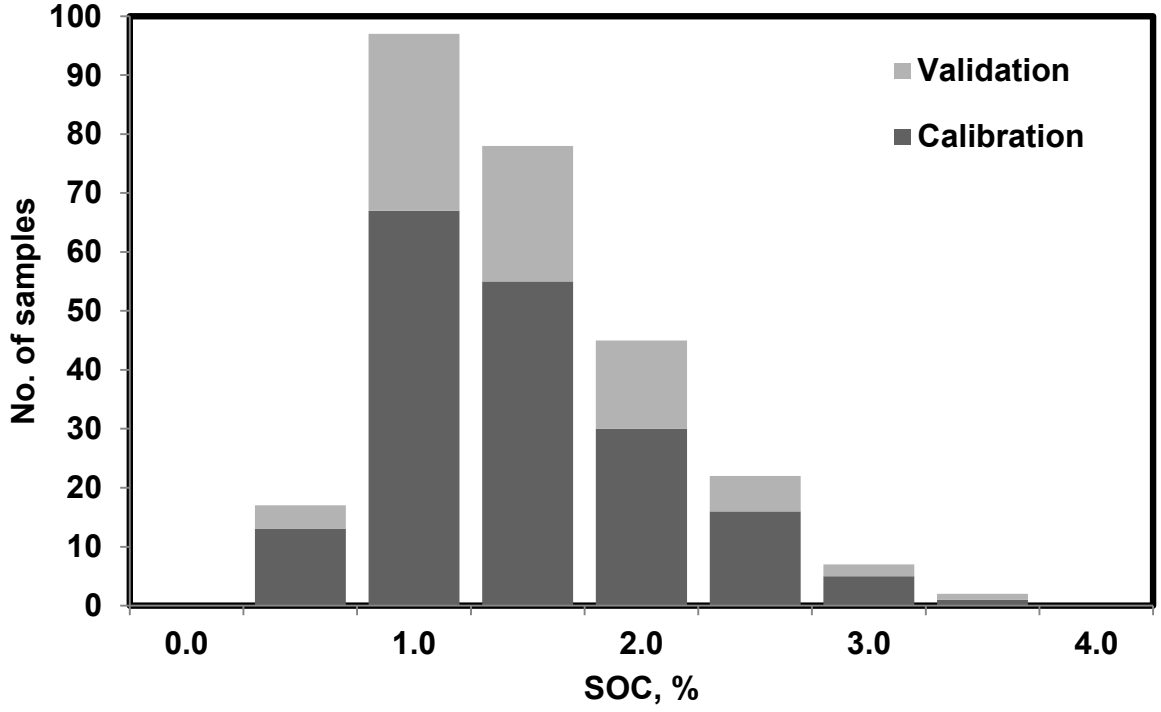


Fig. 4.8 Distribution of % SOC content

4.3 Results and Discussion

Percent sand and clay ranged from 0 to 86 % and from 4 to 76 %, respectively. A substantial range between 0.5 and 3.9 % was observed in SOC content (Table 4.1). As shown in Table 4.2, with mid-IR soil spectral calibration examples, three models using a relatively small number of latent factors (5, 4 and 5) were determined to be suitable, with an R^2 between 0.61 and 0.64.

For the validation sets of soil, RMSE was equal to 10.3% sand, 7.8% clay and 0.41% SOC. Small ME values indicate a fair split between calibration and validation sets of soil resulting in the absence of bias. The difference between the obtained results using calibration and validation datasets was also minor.

As shown in Table 4.3, with the vis-NIR soil spectral calibration, relatively large numbers of latent factors (15, 7 and 12) were found suitable to predict % sand, % clay and % SOC with R^2 ranging between 0.62 and 0.79. The RMSE was 12.7% sand, 7.2% clay, and 0.45% SOC. The remaining observations were similar to the mid-IR spectra. Relationships between measured and predicted percentages of sand, clay and SOC, using both spectrometer techniques are illustrated in Figures 4.9-4.11. As expressed by the RMSE (Tables 4.2 and 4.3), there was no obvious difference in terms of the instrument performance and similarly to the results reported earlier (Viscarra Rossel *et al.,* 2006b); % sand was slightly better predicted using mid-IR and % clay using vis-NIR. However, it should be noted that soil sample area brought in contact with the detectors of both systems was significantly different and therefore, the averaged absorbance values reported by the vis-NIR detector was collected from a sample area that was ~11 times smaller than the sample area used during mid-IR based data collection.

Similar to the results reported in an earlier study (Viscarra Rossel *et al.,* 2006b), % sand was slightly better predicted using mid-IR and % clay using vis-NIR. It was found that the hyperspectral analysis of dry, crushed, sieved and milled soil samples can reveal percent clay and sand with a standard error less than 12% and for SOC, less than 0.5%. Because % silt can be estimated using % clay and % sand predictions and its distribution in the dataset failed to

uniformly represent the entire range (required to find an optimal PLRS model), predicting this property was omitted from the current objectives.

Table 4.1 Statistical data of soil properties observed in calibration and validation sets

Data set	Statistics	% sand	% clay	% SOC
Calibration	Min	0	4	0.54
	Median	34	28	1.60
	Max	86	74	3.91
	Mean	38	29	1.71
	SD	20	14	0.60
Validation	Min	0	5	0.97
	Median	33	28	1.59
	Max	86	75	3.75
	Mean	37	30	1.76
	SD	24	16	0.61

Table 4.2 Statistical results of PLSR model calibration and validation, developed and tested using the mid-IR soil spectral data

Property	Data Set	No. of PLSR factors*	R^2	RMSE	SDE	ME
% sand	Calibration*	5	0.64	12.14	12.17	0.05
	Validation		0.82	10.33	10.32	1.20
% clay	Calibration	4	0.61	8.86	8.88	-0.03
	Validation		0.78	7.79	7.83	-0.25
% SOC	Calibration	6	0.63	0.37	0.37	0.00
	Validation		0.54	0.41	0.42	-0.02

*The PLSR number of factors were chosen by using leave-one-out cross validation technique

Table 4.3 Statistical results of PLSR model calibration and validation, developed and tested using the vis-NIR soil spectral data

Property	Data Set	No. of PLSR factors*	R^2	RMSE	SDE	ME
% sand	Calibration	15	0.74	10.40	10.40	0.14
	Validation		0.72	12.73	12.48	-2.85
% clay	Calibration	7	0.79	6.57	6.58	-0.02
	Validation		0.82	7.17	7.19	-0.58
% SOC	Calibration	12	0.62	0.38	0.38	-0.01
	Validation		0.49	0.45	0.45	0.09

*The PLSR number of factors were chosen by using leave-one-out cross validation technique

There was no obvious benefit of vis-NIR system over the mid-IR, or vice versa. The stated error estimates reported in this study and in earlier studies (Janik and Skjemstad, 1995; Janik *et al*., 1998; Masserschmidt *et al*., 1999; Janik *et al*., 1998; Chang *et al*., 2001; Chang and Laird, 2002; Shepherd and Walsh. 2002; Cozzolino and Moron, 2003; Viscarra Rossel *et al*., 2006b; Reeves III, 2010; Viscarra Rossel *et al*., 2011) are higher than those corresponding to traditional analysis, where the laboratory method has a detection limit of 2% for sand and clay (dry basis), and generally, is reproducible to within ± 8% (Gee and Bauder, 2986).

In terms of SOC, the standard error is 0.25-0.5% (Miller, 2006), which is compatible to the RMSE estimated in this study. However, these results do not necessarily represent the potential predictability of soil texture and SOC in situ under different levels of soil water content and bulk density. However, an increased number of less accurate yet unbiased measurements could significantly increase the overall accuracy of thematic soil maps (Adamchuk *et al.,* 2004, Gebbers *et al.,* 2009), such as SOM and particle size distribution.

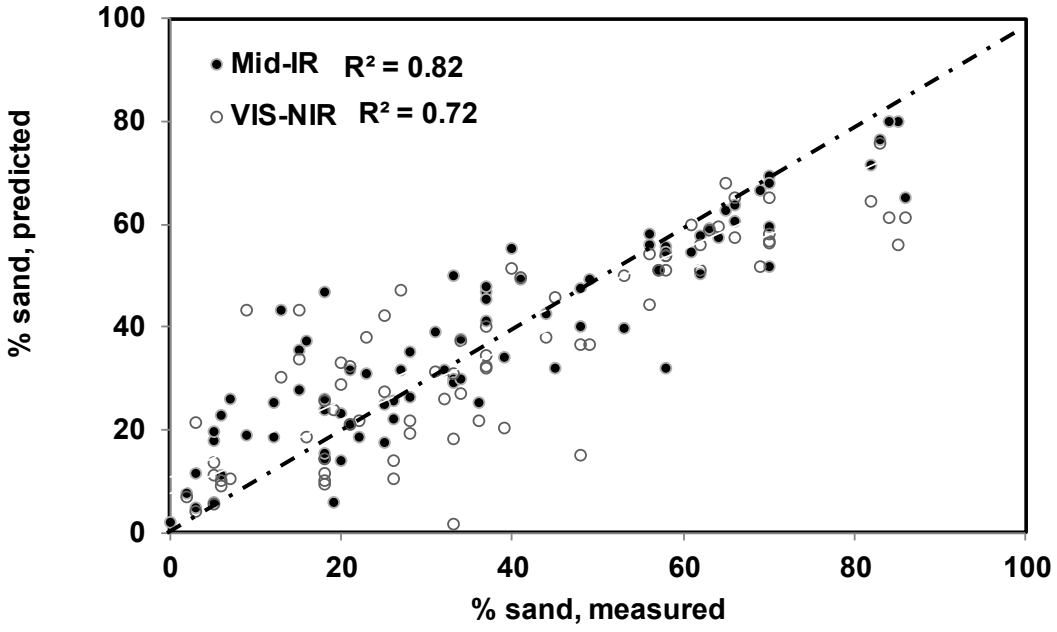


Fig. 4.9 Relationships between the measured and predicted % sand, with dots representing validation examples in accordance with Tables 4.2 and 4.3

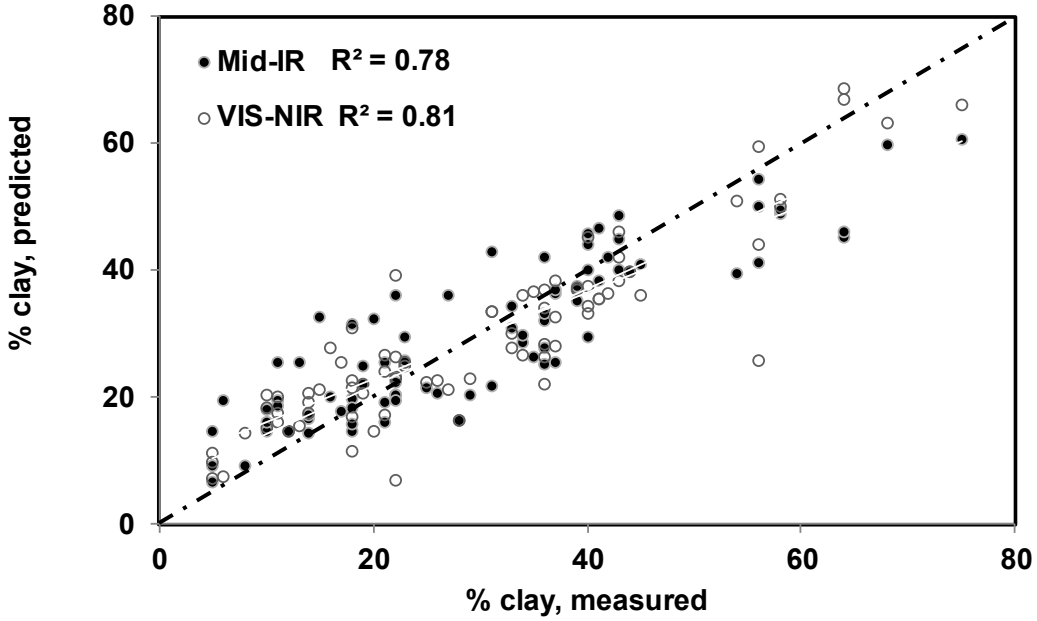


Fig. 4.10 Relationships between the measured and predicted % clay, with dots representing validation examples in accordance with Tables 4.2 and 4.3

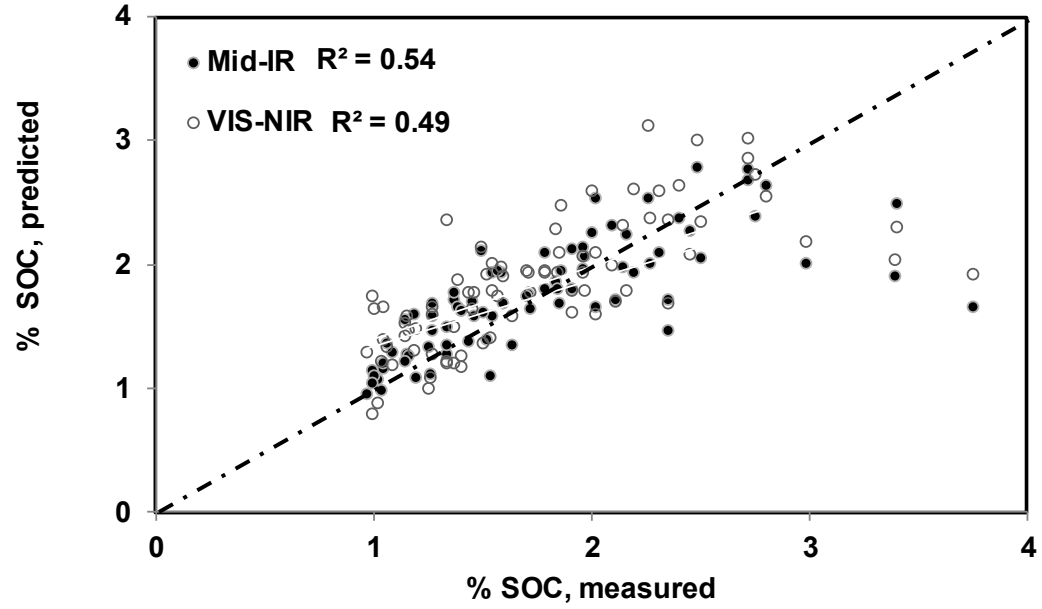


Fig. 4.11 Relationships between the measured and predicted % SOC, with dots representing validation examples in accordance with Tables 4.2 and 4.3

4.4 Conclusions

A prototype portable mid-IR spectrometer was compared to a vis-NIR system for predictions of percent sand, clay, and SOC in soil. PLSR models established on mid-IR calibration samples using leave-one-out cross validation needed fewer latent factors as compared to the vis-NIR spectra. Otherwise, there were no significant differences between the performance indicators for both systems. Relatively minor differences indicated that sand was better predicted using mid-IR and clay was better predicted using vis-NIR spectra. The highest coefficients of determination were found to be similar for both sand (0.82) and clay (0.82). The corresponding root mean squared errors were 10.3% and 7.2%, respectively. The ability to accurately predict SOC was not as well supported for the set of soils in this experiment with a root mean squared error of approximately 0.4%. The tested method proves the usefulness of both vis-NIR and a portable mid-IR for predicting soil physical properties.

Connecting Text to Chapter 5

Chapter 5 is also related to the second objective of this study listed in Chapter 1. As shown in Chapter 4, spectral data obtained in nicely prepared and milled soil samples in laboratories using a portable prototyped mid-IR instrument showed promising results for predicting soil texture. In order to examine the advantages of mid-IR over vis-NIR and to re-evaluate the instrument capability for measuring soil texture and soil organic matter in naturally moist field conditions, soil spectra were collected in naturally moist soil environments. Promising results for measuring soil texture indicated its potential use for OSA. The main results of this study were submitted to the "European Journal of Soil Science" as follows:

Dhawale, N.M., V.I Adamchuk, S.O. Prasher, R.A. Viscarra Rossel, A.A. Ismail, & K. Jasmeen. 2014. Proximal soil sensing of soil texture and organic matter with a prototype portable mid-infrared spectrometer. *European Journal of Soil Science* (in press).

Chapter 5

PROXIMAL SOIL SENSING OF SOIL TEXTURE AND ORGANIC MATTER WITH A PROTOTYPE PORTABLE MID-INFRARED SPECTROMETER

Dhawale, N., V. I. Adamchuk, S. O. Prasher, R.A. Viscarra Rossel, A.A. Ismail, J. Kaur

Abstract

Recent advances in semiconductor technologies have given rise to the development of mid infrared (mid-IR) spectrometers that are compact, relatively inexpensive, robust and suitable for *in situ* proximal soil sensing. The objectives of this research were to evaluate a prototype portable mid-IR spectrometer for direct measurements of soil reflectance and to model the spectra to predict sand, clay and soil organic matter (SOM) contents under a range of field soil water conditions. Soil samples were collected from 23 locations at different depths in four agricultural fields to represent a range of soil textures, from sands to clay loams. The particle size distribution and SOM content of 48 soil samples were measured in the laboratory by conventional analytical methods. In addition to air-dry soil, each sample was wetted with two different amounts of water before the spectroscopic measurements were made. The prototype spectrometer was used to measure reflectance (R) in the range between 1811 and 898 cm^{-1} (approximately 5522 to 11 136 nm). The spectroscopic measurements were recorded randomly and in triplicate, resulting in a total of 432 reflectance spectra (48 samples x three soil water contents x three replicates). The spectra were transformed to $\log_{10}$ (1/R) and mean centred for the multivariate statistical analyses. The 48 samples were split randomly into a calibration (70%) and a validation (30%) set. A partial least squares regression (PLSR) was used to develop spectroscopic calibrations to predict sand, clay and SOM contents. Results show that the portable spectrometer can be used with PLSR to predict clay and sand contents of either wet or dry soil samples with a root mean square error (RMSE) of around 10%. Predictions of SOM content resulted in RMSE values that ranged between 0.76 and 2.24%.

Keywords: proximal soil sensing, diffuse reflectance spectroscopy, visible and near-infrared spectra, naturally moist soil

5.1 Introduction

Soil texture and soil organic matter (SOM) affect plant nutrient availability; measuring them by conventional laboratory methods is laborious and time-consuming. Proximal soil sensing using diffuse reflectance spectroscopy (DRS) has been suggested as a possible alternative because measurements are rapid and inexpensive (Viscarra Rossel *et al.,* 2011). Furthermore, the technique can produce simultaneous assessment of different soil properties, including soil texture and SOM. Diffuse reflectance spectroscopy uses the interactions of visible (vis: 400-700 nm), near infrared (NIR: 700-2500 nm) and or mid-infrared (mid-IR: 2500-5000 nm) radiation to investigate the characteristics of the soil sample. Intense fundamental molecular frequencies related to soil components occur in the mid-IR part of the electromagnetic spectrum. Weak overtones and combinations of these fundamental vibrations from the stretching and bending of NH-, OH- and CH-groups dominate the NIR, and electronic transitions affect the vis portions of the electromagnetic spectrum.

There are many studies on the use of mid-IR spectroscopy in soil science (Nguyen *et al*., 1991; Janik *et al*., 1995; Viscarra Rossel *et al*., 2008). The mid-IR spectra measure the fundamental molecular vibrations of bonds in soil minerals and organic matter (Nguyen *et al*., 1991). Janik *et al*. (1998) wrote a review of soil analysis using mid-IR techniques and Viscarra Rossel *et al*. (2006b) reviewed the predictability of soil properties using vis, vis-NIR, NIR and mid-IR spectroscopy.

The choice of spectral region depends on the accuracy of the predictions, the cost of the technology and the amount of sample preparation required (Viscarra Rossel *et al*., 2006b). For instance, the gain in prediction accuracy achieved with a mid-IR instead of a vis-NIR spectrometer might not outweigh the greater cost of the mid-IR technology. In addition, mid-IR instruments are relatively large and fragile, and sample preparation for mid-IR analysis is time consuming as it requires the milling of the samples to particle sizes of around 80-100 μm. Field operation with these instruments is also thought to be less useful because the spectra are affected by soil water. These issues prevent the use of mid-IR spectroscopy in the field (Janik *et al*., 1998).

Reeves III (2010) and Reeves III *et al*. (2010) demonstrated the use of portable mid-IR spectrometers to measure soil carbon and nitrogen in naturally moist soil samples. They concluded that the mid-IR calibrations were not as robust as calibrations obtained with a NIR spectrometer. Linker (2008) demonstrated the use of an attenuated total reflectance (ATR) mid-IR spectrometer configured with a linearly variable filter (LVF) array (800-1550 cm^{-1}) to classify 202 wet soil samples at field capacity. They suggested that depending on the amount of clay, calcium carbonate (CaCO3) and SOM, the samples were classified accurately into five soil types.

There are few examples in the literature on the development and use of small, robust, inexpensive spectrometers that use new semiconductor technologies. Coates (2000) reported on a vis-NIR instrument that was constructed using semiconductor technology. The author discussed its working principles, applications, advantages and limitations compared to bench-top instruments constructed with conventional optics. The instrument described consisted of a multi pixel linear array detector coupled with a LVF and an electronically controlled vis-NIR source.

The LVF is a continuous multi wavelength filter that is commonly used to modify the intensity distribution in a radiation beam. It is used in various optical sensors where wavelength separation is required and does not typically involve moving parts. Electronically controlled sources do not require any special mechanism for cooling and are thought to provide stable operation. The author also suggested that compactness, miniaturization and portability might provide logical extensions of spectroscopy to applications beyond the laboratory for on the spot use. Similar technologies can help potentially to transfer soil mid-IR spectroscopy to the field.

The aim of this research was to evaluate the use of a prototype portable mid-IR spectrometer based on advanced semiconductor technology and with a spectral range of 1811-898 cm^{-1} to predict clay, sand and SOM contents of soil samples with different textures and wetted to different water contents.

5.2 Materials and Methods

5.2.1 Instrumentation and Data Acquisition

A prototype, portable mid-IR variable-filter-array (VFA) diffuse reflectance infrared Fourier transform (DRIFT) spectrometer from Wilks Enterprise, Inc. (East Norwalk, Connecticut, USA), was used to obtain the reflectance spectra of soil samples (Fig. 5.1).

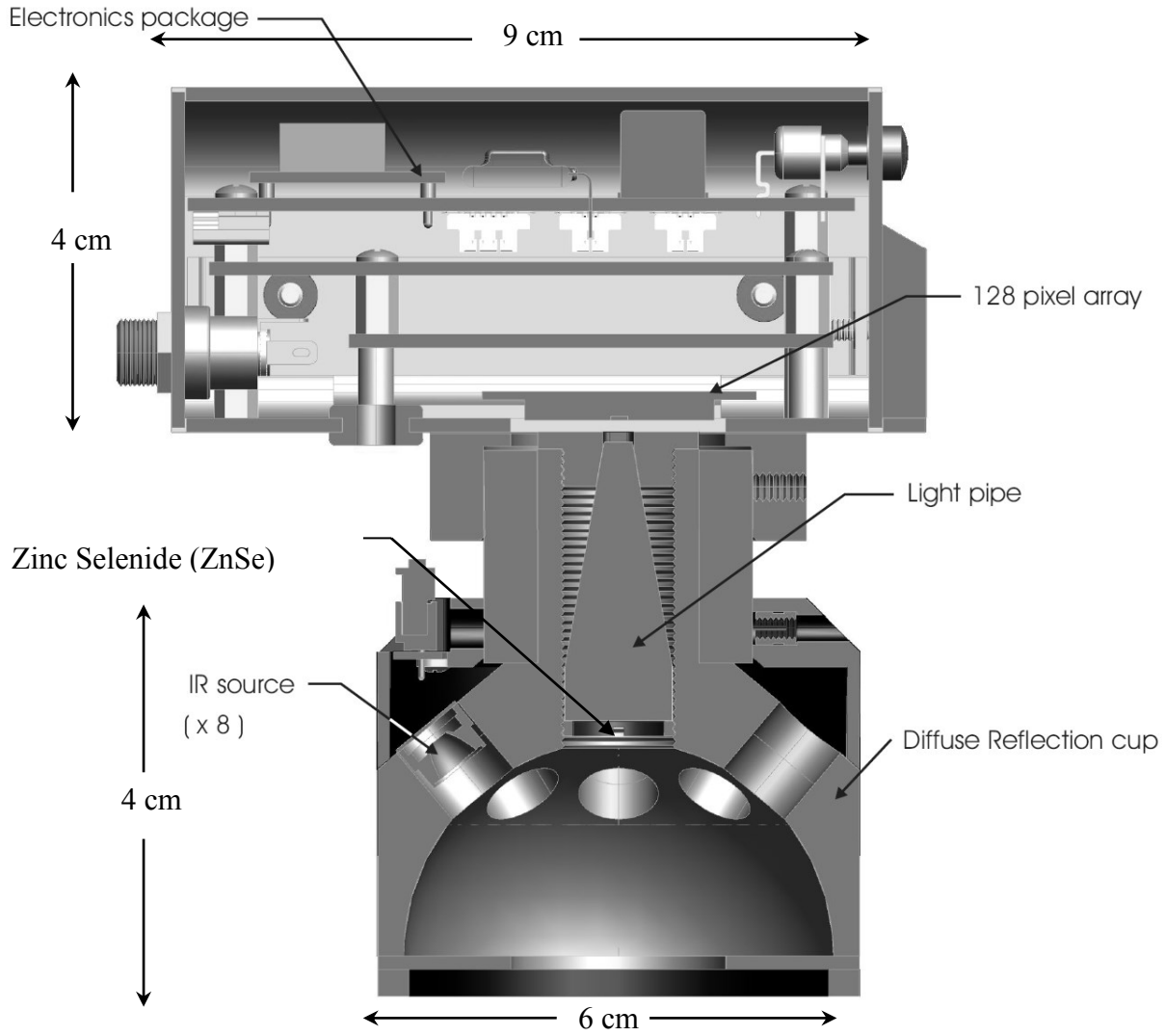


Fig. 5.1 Prototype mid-IR spectrometer (Wilks Enterprise, Inc., East Norwalk, Connecticut, USA)

This instrument is composed of five fundamental elements: (i) an integrated imaging sphere, (ii) eight electronically modulated sources of IR light, (iii) a LVF coupled with a 128 linear pyroelectric detector array, (iv) a USB port for serial communication and (v) signal conditioning and processing electronics. It can be operated under laboratory or field conditions in ambient temperatures between 15-60° C and up to 98% relative humidity.

A LVF made of zinc selenide (ZnSe) was mounted on a 128-pixel array above the optical window. A distinct feature of the spectrometer's design was to maintain a constant distance between the source of light, the object being measured and the detector. It is possible to mount one of several different LVF arrays to allow measurement in both the NIR and mid-IR ranges. In this study, a short mid-IR VFA (1811-898 cm^{-1}) was used. The instrument can be powered by a standard 12 V DC adapter and data rea recorded to a standard laptop. The C[1] spectral data analysis and management software (CogniSolve, Inc., Montreal, Québec Canada) was used to transform the measured digital numbers to percentage soil diffuse reflectance.

About 10-g of soil were spread in Petri dishes and the mid-IR instrument was placed over the 5 -mm deep sample with an area of about 707 mm^2 (Fig. 5.2). As well as the air-dry soil samples,

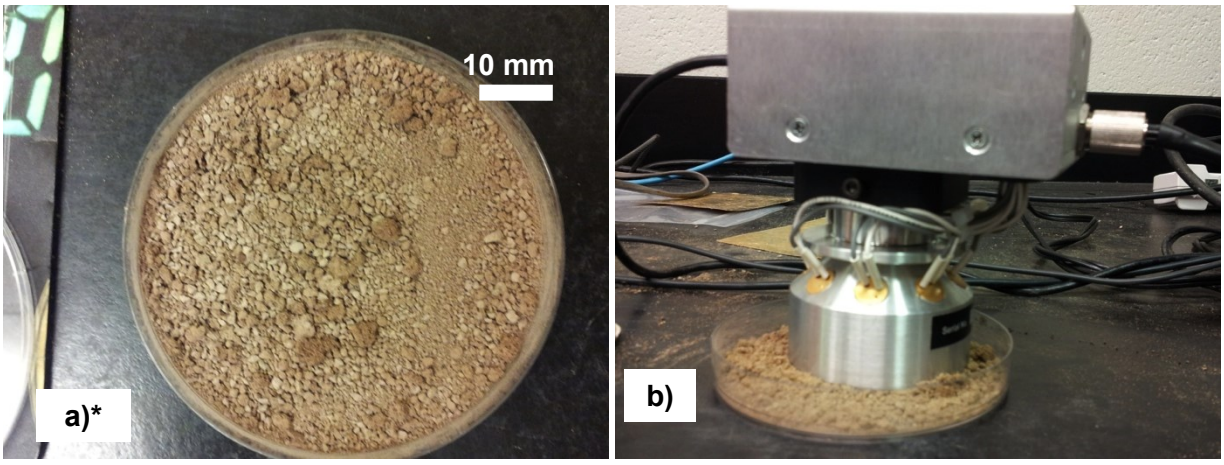


Fig. 5.2 Experimental setup: (a) sample preparation and (b) measurements

*Over-sized particles appear due to the fact that water was applied and aggregates were formed after drying.

wet samples were prepared by mixing soil with water (according to the experimental design described below) to represent natural soil moisture content during *in situ* measurements with the instrument. To minimize instrumental noise, each spectrum was the average of 32 consecutive measurements taken at a rate of approximately one measurement per second. The spectrometer was re-calibrated after every five samples using a gold-colour standard reference sheet provided by the manufacturer.

5.2.2 Experimental Design

Twenty three locations in four agricultural fields on the Macdonald campus research farm, Ste-Anne-de-Bellevue, QC, Canada (45°24' N, 73°56' W) (Fig. 5.3), were identified as

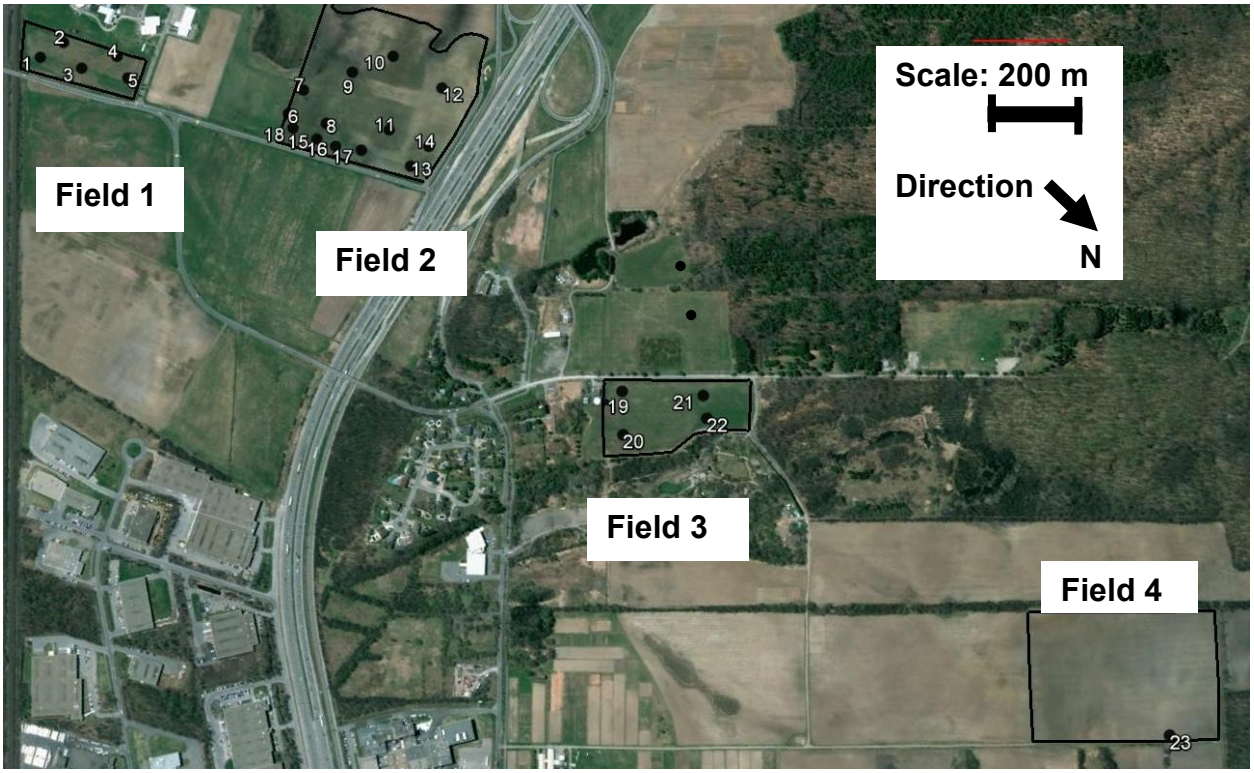


Fig. 5.3 Soil sampling locations distributed over four experimental fields (Ste-Anne-de-Bellevue, QC, Canada)

representing soil types with different textures, varying from sand to clay loam. At each location, soil cores were taken in plastic tubes to a depth of 100 cm with a mechanized soil sampler (Veris® P4000, Veris Technologies Inc., Salina, KS, USA). The depth of each sample was limited

by the depth to bedrock or by reaching an impenetrable soil layer. When possible, each core was partitioned according to depth (<15, 15-30 and >30 cm). This resulted in 21, 17 and 10 samples per depth layer, respectively (Table 5.1). In addition to the impenetrable depths, seven samples with a SOM greater than 10% were not included in this analysis in order to focus on mineral soil samples only. Therefore, a total of 48 soil samples were used in this experiment.

After air-drying, each soil sample was split into two sub-samples. The first sub-sample was used for the conventional analysis of particle size and SOM. Particle size was determined by the hydrometer method (Gee & Bauder, 1986). Soil organic matter was measured by loss-on-ignition (350° C) (Schulte & Hopkins, 1996). The descriptive statistics of the soil analyses results are summarized in Table 5.2. The second set of sub-samples was transferred to Petri dishes and used for the spectroscopic measurements. Conventional mid-IR measurements are often made on soil milled to around 100-80 μm (Janik *et al.*, 1998). However, for this study, spectral measurements were recorded on ≤ 2 mm sieved soil and without any additional sample preparation.
To prepare soil samples that represented field conditions, 8-10 g of each air-dry soil sample were mixed with 0.5, 1.0, 1.5 and or 2.0 g of water. Table 5.3 summarizes the amounts of water as a percentage added to the air-dry samples according to the texture class to represent moist soil samples that were relatively dry (5-10% of soil gravimetric moisture) and relatively wet (15-20% of soil gravimetric moisture). Nine soil spectra were recorded from each of the 48 soil samples; that is three replicated spectra were recorded from the air-dried soil and three replicates from each of the wetted sample sets. This resulted in 144 air-dry and 288 moist soil spectra obtained randomly for each level of added soil water in three, non-consecutive replicates.

5.2.3 Data Processing and Analysis

The spectra were transformed from reflectance measurements into optical density values ($\log_{10}$ 1/reflectance) and were mean centered. The spectra were calibrated using a PLSR algorithm (Martens & Naes, 1989). Partial least squares regression is a well-known chemometric technique that is based on bilinear regression that extracts a smaller number of latent factors from large multi-dimensional data-sets. These factors represent linear combinations of the independent spectroscopic variables and the soil properties (Geladi & Kowalski, 1986).

Table 5.1 Laboratory measured soil properties of 48 soil samples collected from 23 locations and 3 depths (when available) used for mid-IR spectra calibration and validation

Field #	Sampling location	Sampling depth		
		< 15 cm	15 - 30 cm	> 30 cm
1	1	***56/34/10(SL) & %5.9****	***32/40/28(CL) & %1.6***	***38/29/33(CL) & %3.2*** **
	2	53/38/9(SL) & %4.1	50/21/29(SCL) ***& %***/1.4	N/A***
	3	***69/22/9(SL) & %2.5***	***56/21/23(SCL) & %0.5***	N/A
	4	70/22/8(SL) & %1.0	74/16/10(SL) ***& %***0.9	N/A
	5	65/31/4(SL) & %6.0	56/31/13(SL) ***& %***7.5	11/39/50(C) ***& %***4.5
2	6	***34/21/45(C) & %1.0***	N/A	N/A
	7	***75/17/8(SL) & %1.9***	N/A	N/A
	8	73/18/9(SL) ***& %***1.2	62/30/8(SL) ***& %***4.4	N/A
	9	53/31/16(SL) ***& %***3.5	N/A	N/A
	10	63/31/6(SL) ***& %***4.0	56/36/8(SL) ***& %***7.3	57/19/24(SCL) ***& %***4.4
	11	55/35/10(SL) ***& %***6.6	74/15/11(SL) ***& %***0.9	36/46/18(L) ***& %***1.1
	12	29/52/19(SiL) ***& %***5.1	26/42/32(CL) ***& %***5.2	N/A
	13	36/32/32(CL) ***& %***2.1	46/35/29(L) ***& %***1.2	N/A
	14	42/36/22(L) ***& %***0.5	12/55/33(SiCL) ***& %***1.0	14/47/39(SiCL) ***& %***5.6
	15	69/26/5(SL) ***& %***4.7	62/30/8(SL) ***& %***3.2	N/A
	16	65/26/9(SL) ***& %***4.6	55/37/8(SL) ***& %***3.6	N/A
	17	N/A	N/A	27/47/26(L) ***& %***5.9
	18	N/A	***70/13/17(SL) & %2.1***	N/A
3	19	***91/7/2 (S) & %3.8***	***36/28/36(C) & %2.1***	N/A
	20	96/2/2(S) ***& %***1.5	96/2/2(S) ***& %***1.5	93/5/2(S) ***& %***3.6
	21	60/30/10(SL) ***& %***3.9	50/25/25(SCL) ***& %***1.2	3/16/81(C) ***& %***0.8
	22	***38/29/33 (CL) & %3.2***	N/A	***18/22/60(C) & %1.0***
4	23	***7/26/67 (C)& %1.0***	N/A	***2/26/72(C) & %0.8***

* % sand/% silt/% clay (texture class) & % SOM with texture classes: C = clay, CL = clay loam, L = loam, S = sand, SCL = sandy clay loam, SL = sandy loam, SiCL = silty clay loam, SiL = silty loam.

**Bold italic font marks soil samples dedicated to test the prediction model developed.

***N/A: data/sample not available.

To validate the model with a separate set of data not used for calibration, the 48 soil samples were split randomly into 34 (70%) calibration samples and 14 (30%) validation samples (Table 5.1). Fig. 5.4 illustrates the distribution of soil texture and SOM between the two sets of soil samples. To evaluate the performance of the air-dry and wet spectroscopic models (Table 5.4) (i) a leave-one-out cross-validation and (ii) the designated independent validation sets were used. These validations illustrate the expected performance of the modelling for different potential applications such as; using (i) air-dry calibration to predict soil properties for new air-dry samples, (ii) moist calibration to predict soil properties for new moist samples, (iii) moist calibration to predict soil properties for new air-dry samples and (iv) air-dry calibration to predict soil properties for new moist samples.

Table 5.2 Descriptive statistics on soil properties of interest

Soil property	Minimum	Median	Maximum	Mean	Standard deviation
% sand	2	55	96	51	24
% silt	2	29	55	28	12
% clay	2	17	81	22	19
% SOM	0.5	2.9	7.5	3.0	2.0

Table 5.3 The amount of water added in percentages added to 8-10 g of air-dried soil sample

Soil texture	Added water %	
	Relatively dry soil	Relatively wet soil
Clay(C)	10	20
Clay Loam(CL)	10	20
Loam(L)	10	20
Sand(S)	5	15
Sandy Clay Loam(SCL)	10	15
Sandy Loam(SL)	5	15
Silty Clay Loam(SiCL)	10	20
Silty Loam(SiL)	10	15

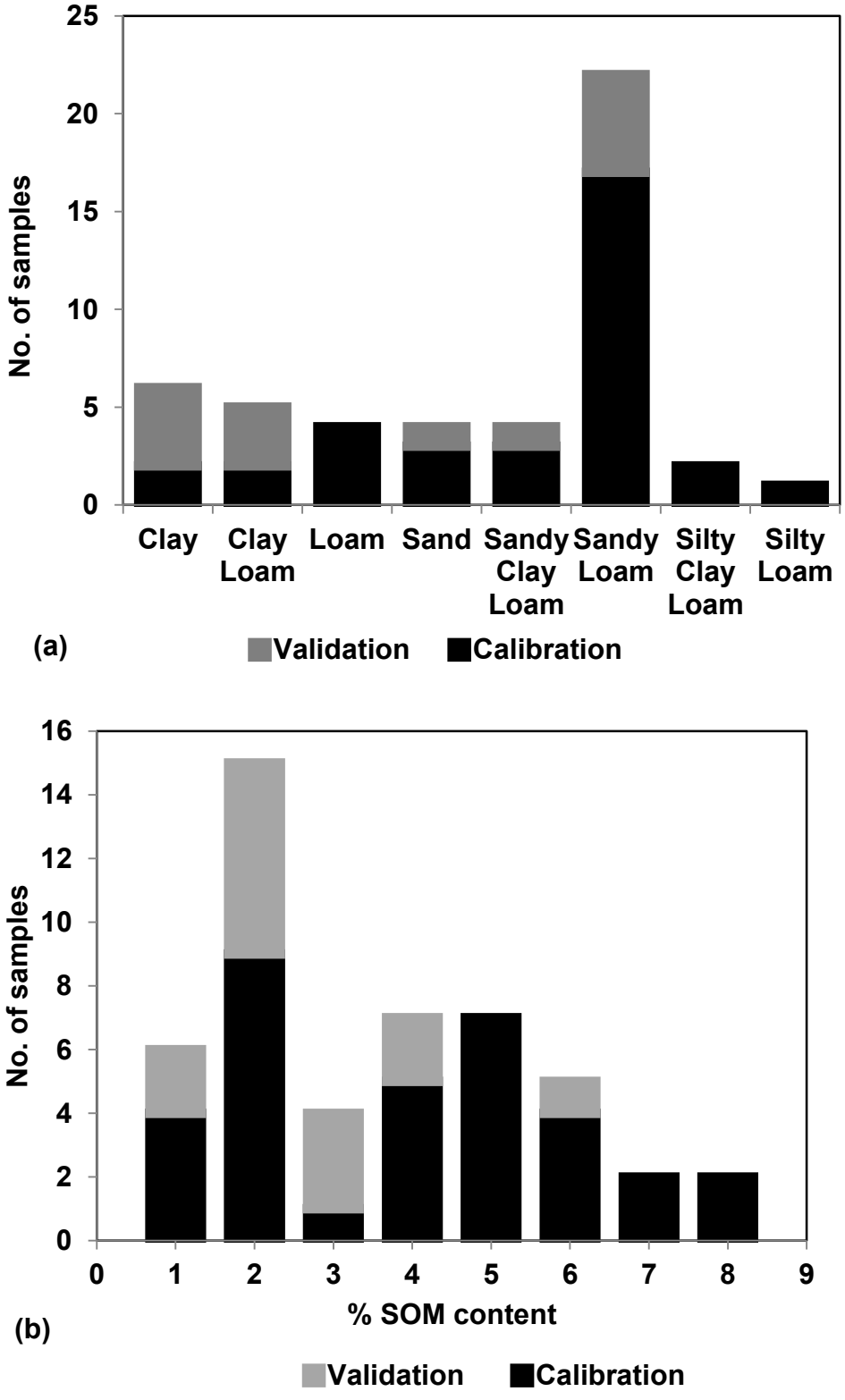


Fig. 5.4 Distribution of the soil samples according to: (a) soil texture and (b) % SOM content

Table 5.4 Statistical results demonstrating performances of various calibration models using PLSR

Soil sets used for		Performance Indicators				
*Model calibration**	*Model validation*	*PLSR Factors*	R^2	*RMSE*	*ME*	*SDE*
% sand						
Air-dry calibration	Cross validation	3	0.74	11.81	-0.03	11.87
Moist calibration	Cross validation	5	0.81	10.12	-0.04	10.15
Air-dry calibration	Air-dry test set validation	3	0.90	8.58	-2.10	8.41
Moist calibration	Moist test set validation	5	0.82	11.65	-4.30	10.89
Moist calibration	Air-dry test set validation	5	0.80	11.98	-3.70	11.53
Air-dry calibration	Moist test set validation	3	0.88	10.26	2.34	10.05
% clay						
Air-dry calibration	Cross validation	3	0.65	9.48	-0.93	9.52
Moist calibration	Cross validation	5	0.79	7.27	-0.03	7.29
Air-dry calibration	Air-dry test set validation	3	0.88	10.50	-5.35	9.14
Moist calibration	Moist test set validation	5	0.91	7.82	-2.57	7.43
Moist calibration	Air-dry test set validation	5	0.84	9.32	-2.63	9.05
Air-dry calibration	Moist test set validation	3	0.89	10.08	-3.37	9.56
% SOM						
Air-dry calibration	Cross validation	6	0.54	1.42	0.02	1.43
Moist calibration	Cross validation	6	0.49	1.48	0.01	1.49
Air-dry calibration	Air-dry test set validation	6	0.58	1.17	0.72	0.93
Moist calibration	Moist test set validation	6	0.62	1.21	0.84	0.87
Moist calibration	Air-dry test set validation	6	0.82	0.76	-0.46	0.61
Air-dry calibration	Moist test set validation	6	0.43	2.24	1.97	1.07

*Model Calibration - data set used in model calibration, Model Validation - data set and method used for model validation, PLSR - Partial least squared regression factors, R^2 - coefficient of determination, RMSE - root mean squared error, ME - mean error, SDE - standard deviation of error distribution.

The performance indicators included (i) the coefficient of determination (R^2) for the linear regression between the measured and predicted soil properties, (ii) the root mean squared error (RMSE), (iii) the mean error (ME) and (iv) the standard deviation of the error distributions (SDE). While R^2 indicates the overall model performance for a given set of samples, RMSE estimates the measurement error in physical units. The ME and SDE split this error between bias and imprecision. The following equations were used:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \quad (5.1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{N}} \quad (5.2)$$

$$ME = \frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)}{N} \quad (5.3)$$

$$SDE = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i - y_i - ME)^2}{N-1}} \quad (5.4)$$

where, y_i is the measured soil property, $\hat{y}_i$ is the predicted soil property, $\bar{y}$ is the mean of measured soil properties and N is the number of soil samples used for validation.

Further, the PLS regression coefficients (betas) and variable importance for projection (VIMP) (Wold, 2001) were used to identify the mid-IR frequencies (waveband numbers) used to predict sand, clay and SOM contents by both air-dry and wet soil calibrations.

The spectroscopic and chemometric analyses were performed with the software ParLeS (Viscarra Rossel, 2008).

5.3 Results and Discussion

Fig. 5.5 shows six average spectra from two soil samples with different textures and three levels of soil water contents. Samples with sandy soil had more SOM than samples with more clay. The spectra of samples with more clay reflected less than samples with more sand. The spectra of sandy soil had distinct local minima near 1100 cm^{-1} whereas that of clayey soil had one near 1300 cm^{-1}. The reflectance spectra of these two types of soil, with three different water contents, were markedly different between 1000 and 1500 cm^{-1}. In Nguyen *et al*. (1991), mid-IR

absorbance spectra measured using DRIFT showed similar characteristic near 1000 cm^{-1} and the authors named them 'inversion bands', which result from severe distortion of strongly absorbed bands at large sample concentrations of various soil minerals (Nguyen *et al*., 1991). They are prominent generally at frequencies below 1100 cm^{-1}.

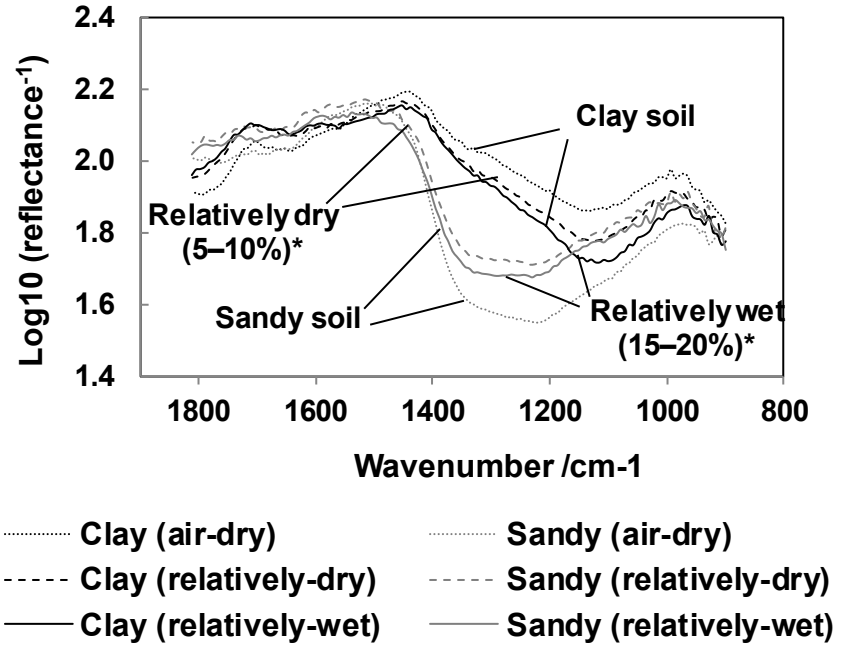


Fig. 5.5 Examples of the mid-IR spectra

*An approximated range of soil moisture content in accordance to Table 5.3, and based on the assumption that readily available moisture contents in air-dried sandy soils should be much lesser than the same in air dried clay soils (Wäldchen *et al.* 2012).

Leave-one-out cross validation showed that the best spectroscopic models to predict clay, sand and SOM had between three to six PLSR factors. When predicting soil texture, the RMSE for both sand and clay contents was around 10%. However, the samples had a wider range of sand content and so the R^2 value was greater for sand than for clay. The RMSE was smaller by more than 1% for predictions of both sand and clay when using moist soil than with air-dry soil. Differences in soil water did not affect predictions of SOM. For both dry and wet samples, RMSE estimates were almost 1.5% SOM and R^2 was around 0.5, indicating that the spectroscopic models for SOM were poor. As one might expect, there was no bias (ME) when evaluating the models by cross-validation.

When the spectroscopic models were validated with the independent set of samples models, bias in the predictions was more apparent. However, this did not affect the RMSE values markedly; they were within 1-2% of those obtained from cross-validation. The RMSE values were not always smaller when using the same soil water content during calibration and validation. In fact, the RMSE was 9.3% when wet soil was used to predict the clay content of the air-dry soil samples. Therefore, in these experiments soil water content did not affect greatly the performance of the spectroscopic models. Fig. 5.6 shows the relationship between the measured and the predicted sand and clay contents for the air-dry calibration data; in each case the RMSE values were around 10%.

The quality of SOM predictions was poor. The smallest RMSE of 0.76% SOM was obtained when wet soil calibration was used to predict the validation set of air-dry soils. However, when the air-dry calibration was used to predict SOM in moist soil in the validation set the results were poor with an RMSE value of 2.24% SOM (Fig. 5.7). The reasons for this might be that absorptions for water masked absorptions related to SOM and that the soil samples were not finely ground, which is typically done for mid-IR spectroscopy. The addition of water to the air-dry sample would affect light scattering that might affect the spectral preservation favourably of SOM related information.

Fig. 5.8 shows the regression coefficients and the VIMP. As assessed by the sizes of their negative or positive peaks in air-dry soil, waveband numbers near 1100 and 1300 cm^{-1} were important for predictions of sand; near 1600 cm^{-1} for predictions of clay and those near 900, 1650, 1750 and 1800 cm^{-1} were most important for predictions of SOM. Similarly, in the wet soil samples waveband numbers near 900, 1100 and 1500 cm^{-1} were useful for predicting sand, those near 900, 1200 and 1400 cm^{-1} for predicting clay and near 900, 1000, 1400 and 1450 cm^{-1} for predicting SOM.

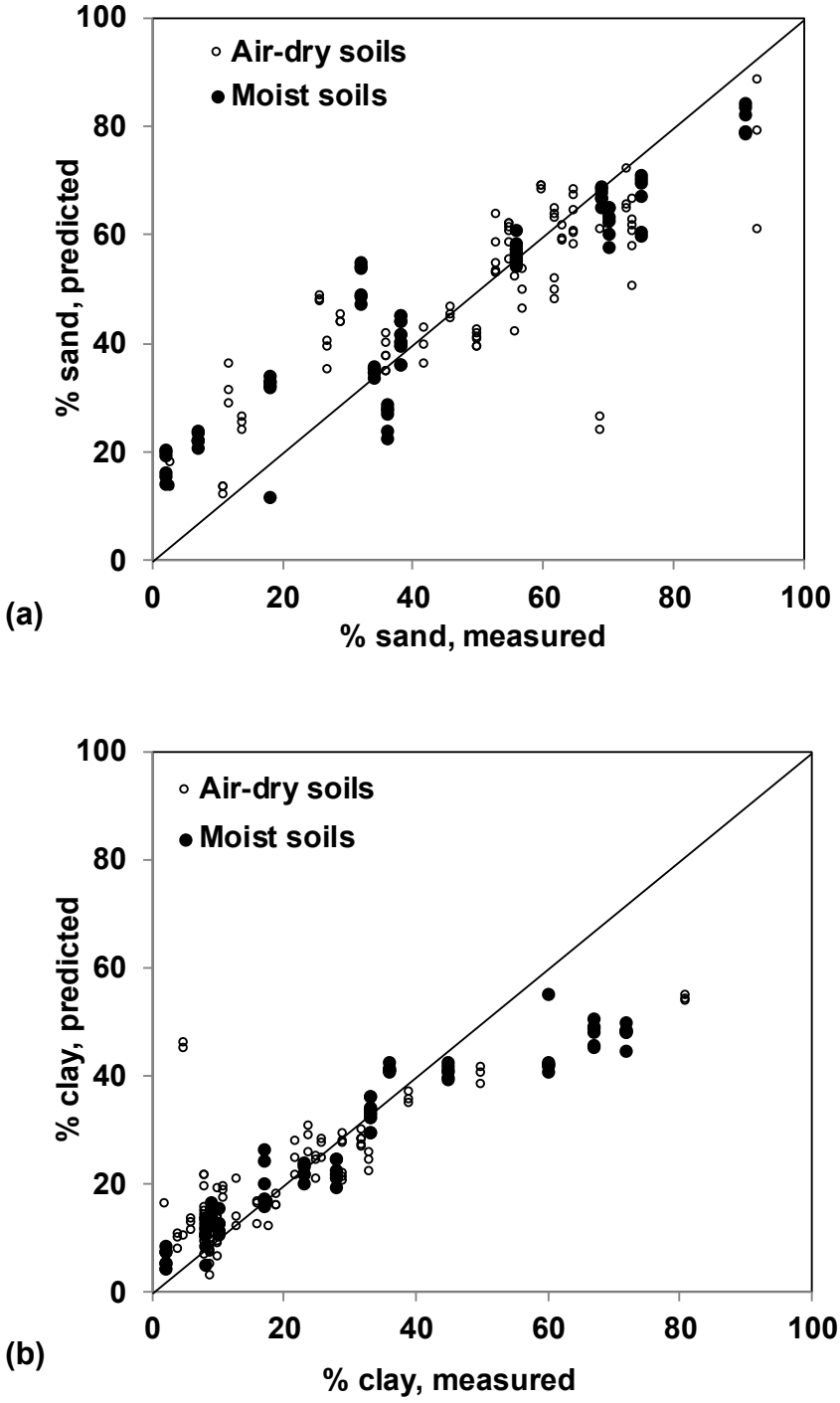


Fig. 5.6 Relationships between the measured and predicted (a) % sand, (b) % clay; where the open dots represent air-dry calibration with leave-one-out cross validation and closed dots represent air dry calibration on moist validation in accordance with Table 5.4

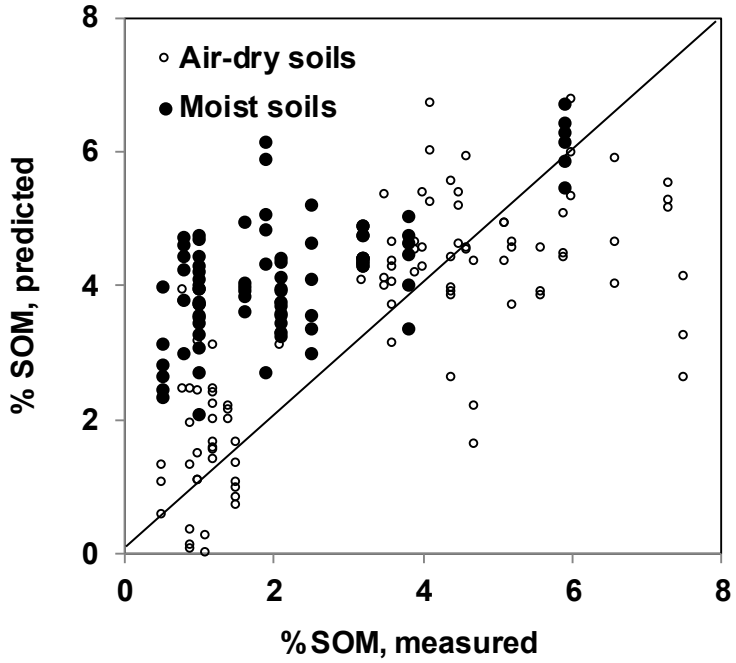


Fig. 5.7 Relationships between the measured and predicted % SOM, where the open dots represent air-dry calibration with leave-one-out cross validation & closed dots represents air dry calibration on moist validation, in accordance with table 5.4

Diffuse reflectance spectroscopy with vis-NIR and mid-IR spectra coupled with multivariate statistics has been proposed as a means to complement conventional laboratory analysis because it can provide inexpensive and rapid measurements of soil properties, including texture and SOM. Portable vis-NIR spectrometers are commercially available, for example: USB2000 (Ocean Optics, Inc., Dunedin, Florida, USA), FieldSpec 4 Standard-Res Spectroradiometer (Analytical Spectral Devices, Ltd., Boulder, Colorado, USA), Mini-Spectrometer C9914GB (Hammatsu Photonics, K.K., Tokyo, Japan). They have been used in different configurations to measure soil properties *in situ* (Christy, 2008; Viscarra Rossel *et al*., 2009; Hodge & Sudduth, 2012).

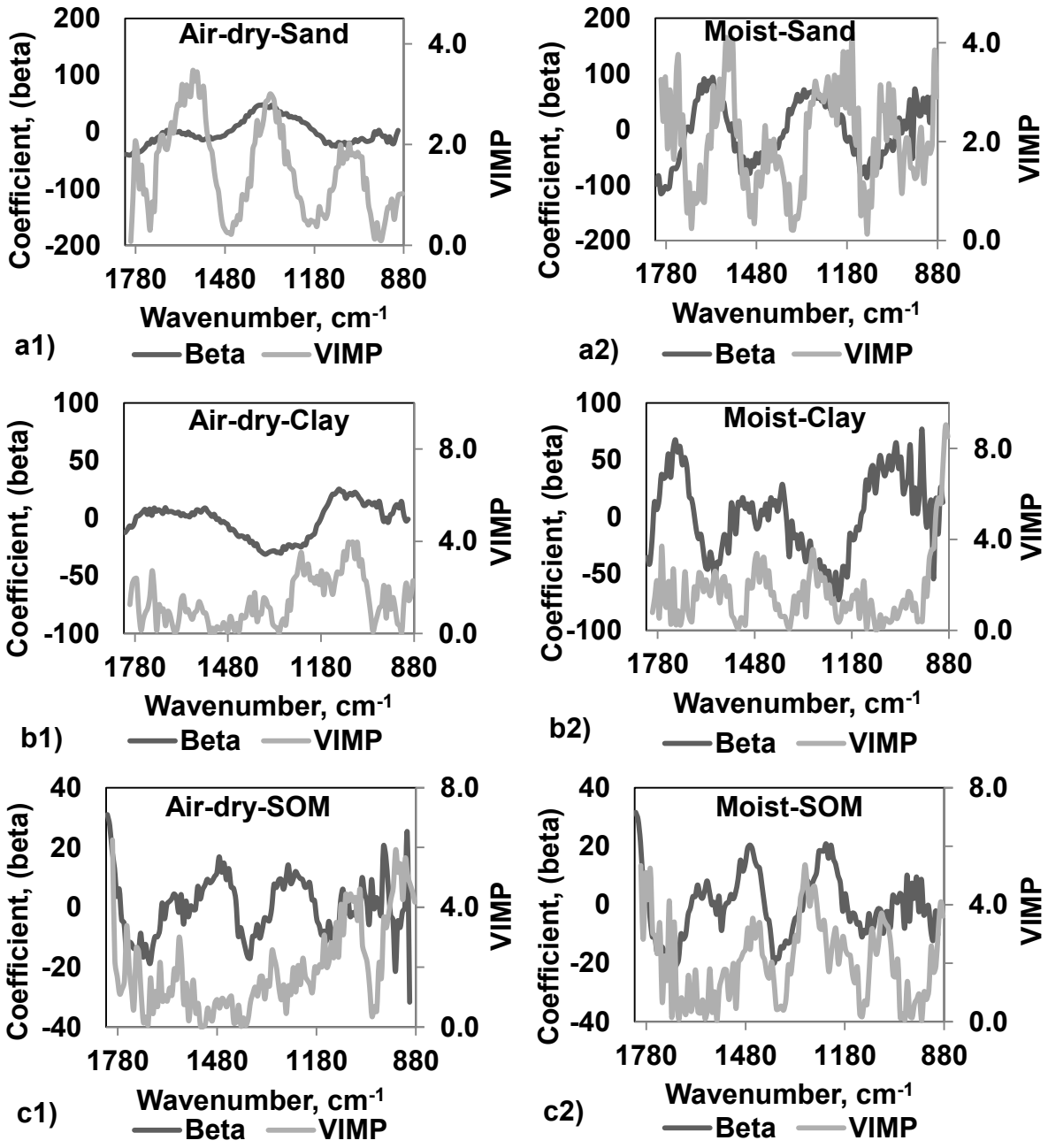


Fig. 5.8 The PLS regression coefficients (betas) and variable importance of projections (VIMP) index related to the prediction of percentage sand, clay and SOM contents using portable mid-IR spectra for both air-dry and moist calibration sets of soil samples

There are a few portable mid-IR spectrometers that are also commercially available, for example: Pyreos (Pvt. Ltd., Edinburgh, Scotland, U.K), Allied Scientific Pro (Pvt. Ltd., Gatineau, Québec, Canada), SOC 400 portable FTS (Surface Optics Inc., San Diego, California, USA), 4100

Exo-Scan FTIR Diffuse Reflectance (Agilent Technologies, Inc., Santa Clara, California, USA). With some exceptions, these 'portable' instruments are often large, fragile and expensive so that they cannot be used practically in the field.

The advantage of mid-IR is that this range is more sensitive than the NIR to both the organic and inorganic composition of soil. Comparing the spectra from the prototype portable mid-IR instrument to spectra reported in other studies (Janik *et al.,* 1998; Viscarra Rossel *et al*., 2008) absorptions near 1800-1500 cm^{-1} are generally caused by the overtones and combination bands of quartz and clay minerals, peaks near 1700 cm^{-1} and 1400 cm^{-1} might result from amides, aromatics, aliphatic acids and alkyl groups of organic materials in soil. Absorptions in the region of 1500-800 cm^{-1} occur mainly from the fundamental O-Si-O stretching and bending vibrations in soil materials. However, absorptions near 1100-1050 cm^{-1} might also be attributed to carbohydrates present in soil, whereas those near 1600 and 1800 cm^{-1} might result from bending of H-O-H and attributed to water (Etzion *et al.,* 2004). The absorption near 1450 cm^{-1} might be associated with the presence of calcium carbonate (Linker, 2008).

The main disadvantage of mid-IR analysis reported previously (Janik *et al*., 1998; Viscarra Rossel *et al*., 2006b; Linker, 2008; Reeves, 2010) is the need for more sample preparation than for vis-NIR, for example, as well as spectroscopic distortions caused by the presence of water in soil. The results presented in this study also show more overlaps and distortion in the spectra of the wet soil samples than those for air-dry samples (Fig. 5). The effects of soil moisture and other environmental factors that affect soil spectra recorded under field conditions could be accounted for in a similar way to Ji *et al*. (2015).

The spectroscopic models could predict sand and clay contents reasonably accurately with RMSEs of 10%. Predictions of SOM were less reliable with RMSEs between 0.76 and 2.24%. These predictions did not differ greatly between spectroscopic models that used spectra from either air-dry or wet soil samples. In part, this might be due to random variation in the soil spectra (as assessed from the replicated measurements) that is similar in magnitude to the differences attributed to the water content. Thus, *in situ* measurements made with the prototype spectrometer can result in similar measurement errors to those estimated under laboratory

conditions. A comparison of the prediction errors against those from an earlier reported study (Reeves III *et al.,* 2010) and several other published results summarized by Viscarra Rossel *et al.* (2006b) showed that the prediction errors of percent sand, clay and SOM content in this study were generally greater. The errors reported in these studies were typically below 10% for sand and clay and 0.5% for SOM. As well as the instrument itself, differences might result from the relatively small, yet diverse, set of samples that were not milled to 80-100 μm. The reason for not milling the samples in these experiments was to test the potential of the spectrometer for use *in situ* under field conditions. This was also the reason for wetting the samples to different water contents.

Practically, the hydrometer method for measuring soil clay, sand and silt contents is reproducible to within 8% (Gee & Bauder, 1986) and for SOM by loss on ignition it is around 0.25-0.5% (Miller, 2006). Although *in situ* measurements with spectrometers tend to produce larger errors than those from conventional laboratory analyses, the quality of the overall analysis, especially if it is for soil mapping, could be improved by increasing the number of less accurate unbiased soil measurements (Adamchuk *et al.,* 2004; Gebbers *et al.,* 2009). Thus, the prototype mid-IR spectrometer described in this study might be used to enhance the overall quality of soil analysis for such purposes.

5.4 Conclusions

A prototype Mid-IR spectrometer was evaluated in terms of its ability to predict the contents of sand, clay and SOM. Based on 34 calibration and 14 validation soil samples analyzed in air-dry and moist conditions, it was found that clay and sand could be predicted with MSE values of around 10%. Wetting the soil did not significantly affect the results and similar measurement errors were obtained for both air-dried and wet soil. The predictability of SOM was not as encouraging. Further testing is needed to investigate potential optimizations of the soil/sensor interface needed to facilitate the deployment of Mid-IR spectroscopy for proximal soil sensing.

Connecting Text to Chapter 6

Chapter 6 is also related to the second objective of this study listed in Chapter 1. As shown in Chapter 4, spectral data obtained in laboratories using commercially available portable vis-NIR equipment showed promising results for predicting soil texture. This chapter discusses a novel study performed on vis-NIR spectral measurements collected in both; *ex situ* and *in situ* environments, in terms of both reproducibility (precision) and predictability (accuracy) of the key physical soil properties, including percent sand, clay and soil organic matter. Based on the results of this study, it can be concluded that vis-NIR spectral measurements obtained using commercially available equipment are compatible in terms of precision and accuracy when using in laboratory conditions and in the field, when soil water content varies substantially. The results of this study were presented at conferences shown below and at the time this thesis was submitted, a manuscript has been submitted to "Computers and Electronics and Agriculture Journal" as follows:

Dhawale, N.M., V.I. Adamchuk, S.O. Prasher, & R.A.Viscarra Rossel. 2014. Precision and accuracy of vis-NIR hyperspectral soil reflectance measurements. *Computers and Electronics and Agriculture Journal* (under review).

Dhawale, N.M., V.I. Adamchuk, S.O. Prasher, R.A. Viscarra Rossel, & A.A. Ismail. 2013. Analysis of the repeatability of soil spectral data obtained using different measurement techniques. *In: Proceedings of the 3rd Global Workshop on Proximal Soil Sensing*, Potsdam, Germany, 26-29 May 2013, eds. R. Gebbers, E. Luck, and J. Ruhlmann, 161-165. Potsdam, Germany: ATB Leibniz-Institut fur Agratechnik Potsdam-Bornim.

Dhawale, N.M., V.I. Adamchuk, S.O. Prasher, & A.A. Ismail. 2013. Analysis of repeatability of potentiometric and optical soil sensors used for measuring agricultural soil properties. *In: poster of Integrated Sensor Systems (ISS) Training program, Summer School Event*. University of Montreal. Montreal. QC. Canada.

Chapter 6

PRECISION AND ACCURACY OF VIS-NIR HYPERSPECTRAL SOIL REFLECTANCE MEASURMENTS

Dhawale, N., V.I. Adamchuk, S. O. Prasher, R. A. Viscarra Rossel

Abstract

Measuring soil texture and soil organic matter (SOM) is essential due to the way they affect the availability of crop nutrients and water during a growing season. Among different proximal soil sensing (PSS) technologies, diffuse reflectance spectroscopy (DRS) have been deployed to conduct rapid soil measurements *in situ*. This technique is indirect and, therefore, requires site specific calibration. The quality of soil spectra is affected by the level of soil preparation and can be assessed through repeatability (precision) and predictability (accuracy) of unbiased measurements and their combinations. The aim of this research was to assess both precision and accuracy of visible and near-infrared (vis-NIR) hyperspectral soil reflectance measurements obtained using a commercial soil profiling tool when predicting SOM, sand and clay. The vis-NIR measurements were conducted in both field and laboratory conditions. Nineteen locations in three agricultural fields were identified to represent an extensive range of soils, varying from sand to clay loam. All the measurements were repeated three times and a ratio spread over error (RSE) was used as the main indicator of the ability of each spectral parameter to distinguish among field location with different soil attributes. Both simple linear regression (SLR) and partial least squares regression (PLSR) models were used to define predictability of SOM, sand and clay. A relatively small number of test locations limited the ability to develop widely applicable calibration models. However, the methodology of evaluating both precision and accuracy of DRS measurements will help future developers evaluate both robustness and applicability of any PSS instrument.

Keywords: proximal soil sensing, diffuse reflectance spectroscopy, visible and near-infrared spectra, measurement precision and accuracy

6.1. Introduction

Soil texture and soil organic matter (SOM) affect plant nutrient availability. Measuring these properties using conventional laboratory methods is laborious and time consuming (Gee and Bauder, 1986; Schulte and Hopkins, 1996; Soil Survey Staff, 2004). Diffuse reflectance spectroscopy is an indirect analytical technique (Couteaux *et al.,* 2003) that has been proposed as an alternative to laboratory tests that would allow for relatively inexpensive and rapid measurements of soil properties. Numerous studies support the use of diffuse reflectance spectroscopy to simultaneous assess various soil properties in laboratories as well as directly in the field (Krishnan *et al.,* 1980; Dalal and Henry, 1986; Clark *et al.,* 1990; Morra *et al.,* 1991; Henderson *et al.,* 1992; Sudduth and Hummel 1993a; Sudduth and Hummel 1993b; Ben-Dor and Banin, 1994; Bishop *et. al.,* 1994; Ben-Dor and Banin, 1995; Clark, 1999; Chang *et al.,* 2001; Chang and Laird, 2002; Martin *et al.,* 2002; Shepherd and Walsh, 2002; Cozzolino and Moron, 2003; McBratney *et al.,* 2003; Malley *et al.*, 2004; Viscarra Rossel *et al.*, 2006b; Christy, 2008; Viscarra Rossel, 2009; Stenberg, 2010). Veris®P4000 (Veris Technologies, Inc., Salina, Kansas, USA)[4] is an example or a ruggedized hyperspectral instrument developed to obtained vis-NIR spectra at multiple depths within soil profile.

Since this technique is indirect, it is important to develop calibration models using data specific pedo-transformation models against measurements obtained according to standardized laboratory protocols (Viscarra Rossel *et al.,* 2011). However the quality of soil spectra affected by the level of soil preparation and data acquisition process may influence the performance of these calibrations models (Chodak *et al.*, 2002; Couteaux *et al.*, 2003; Udelhoven *et al.*, 2003; Madari *et al.*, 2006; Nduwamungu *et al.,* 2009).

According to Holman (2001) *"Accuracy and precision are considered the very important characteristics of a given measurement system". A*ccuracy and precision both can be described as; the difference of instrument reading and the primary standard and the degree of random variations in the instruments output while measuring a constant quantity respectively. Therefore,

[4] Mention of a trade name, proprietary product, or company name is for presentation clarity and does not imply endorsement by the authors, McGill University and CSIRO, nor does it imply exclusion of other products that may also be suitable.

in precedence to a calibrating process, the performance of a soil optical reflectances measured by the DRS instrument can be assessed by means of evaluating both; accuracy and precision component of spectral measurement error.

As it is illustrated in Fig. 6.1, repeated measurements of soil spectra for a given soil sample can significantly vary from one measurement to another. These differences can be associated with 1) random noise within the data acquisition process that includes inconsistencies of soil-detector interface and 2) soil micro-variability that can occur at very small separation distances.

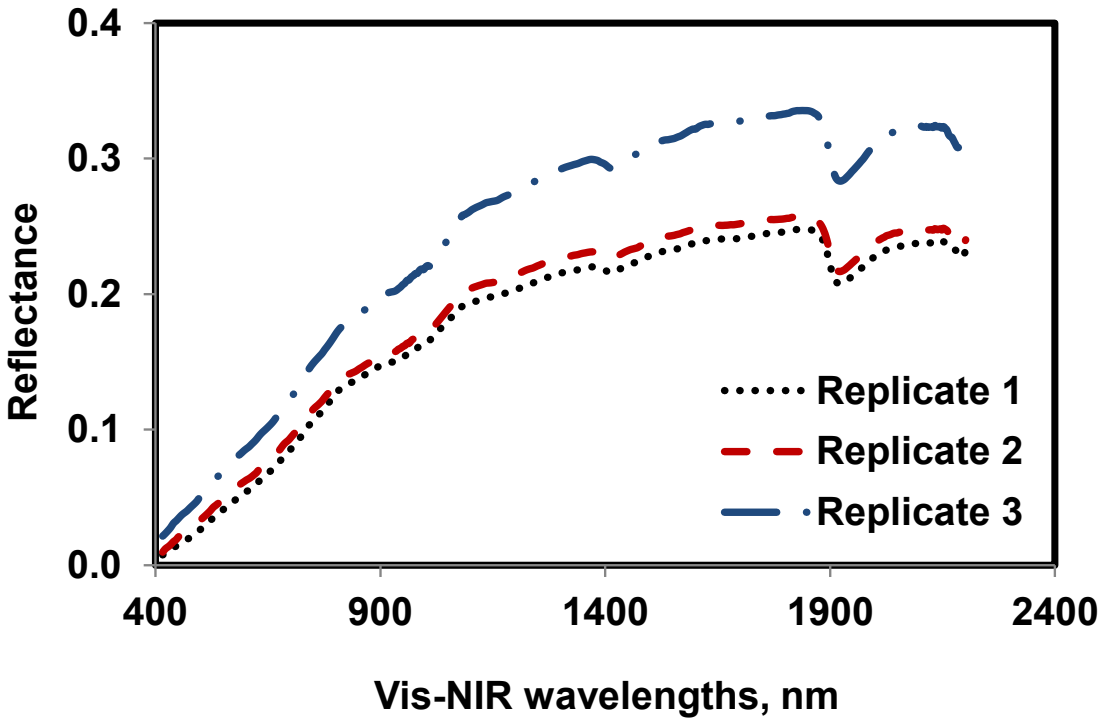


Fig. 6.1 Differences in *ex situ* soil spectra when measured on the same soil sample thrice

In practice, multiple measurements and averaging of several consecutive soil spectra had been accomplished to diminish the effect of random noise. However, it can be noted that certain parts of soil spectra as well as particular parameters constructed using a combination of spectral data values can be relatively repeatable from spectrum to spectrum. Some of these soil spectra parameters that do not vary substantially when testing the same soil, but alter a lot when considering soil samples with different attributes, might be the best candidates to be used for the site-specific calibration process. The objective of this research was to analyze both precision and

accuracy of hyperspectral soil reflectance measurements obtained using a commercial soil profiling tool for predicting percentages of sand, clay and SOM content *in situ* as well as in laboratory conditions.

6.2. Materials and Methods

6.2.1 Data Collection

Three fields at McGill University research farms (Ste-Anne-de-Bellevue, Quebec, Canada) were selected for this project. In total, nineteen locations were selected in these fields (Fig. 6.2) to represent diverse soil conditions ranging from sand to clay loam.

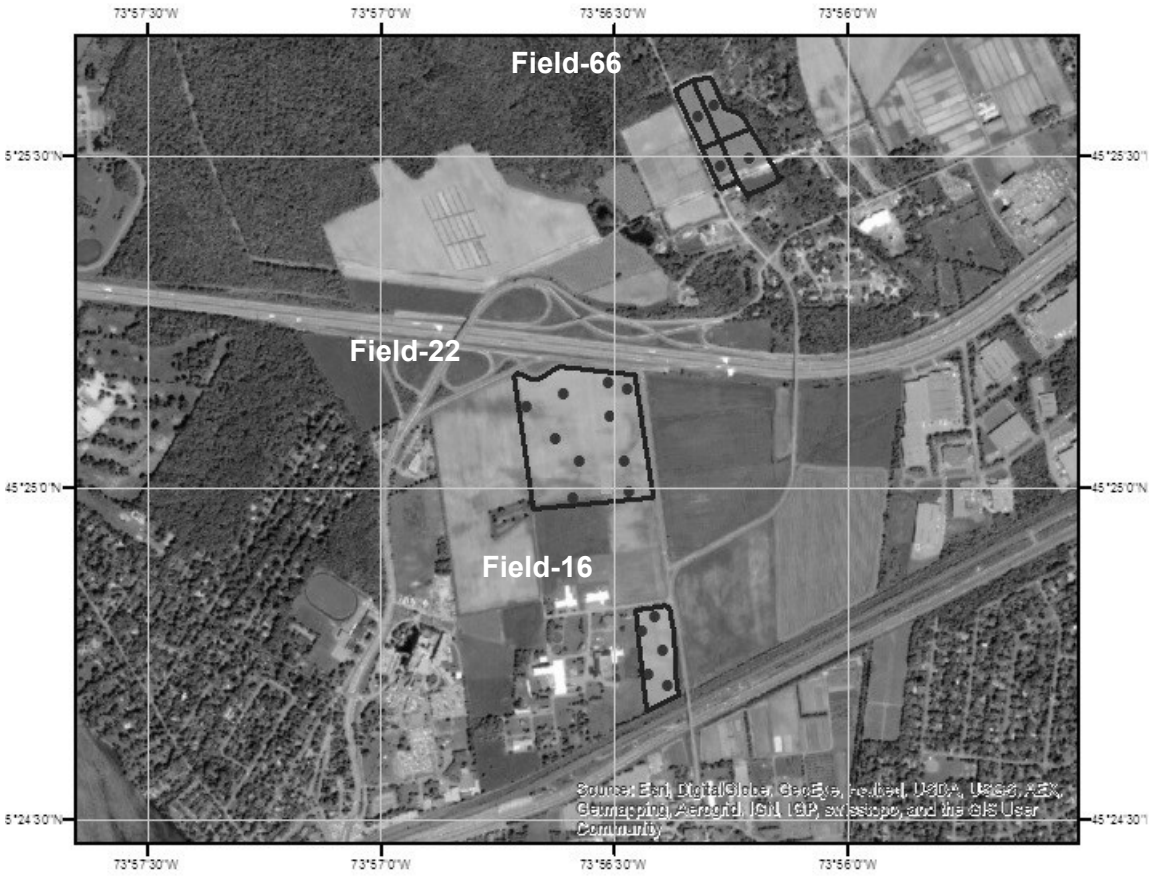


Fig. 6.2 Google earth view of the Macdonald research farm site, illustrating-sampling points

The fields had 2.5, 4.5 and 12 ha areas and were used to grow crop according to a rotation involving soybean, corn (grain and silage) and alfalfa. All the data were collected in 2012 when alfalfa/grass mix was grown in each of the three fields. In each location, composite soil samples were obtained from the top (0-20 cm) layer of soil. A 4-cm diameter, stainless steel auger was used to take five samples from within a 0.5-m radius. These soil cores were mixed, air dried and sieved through a 2-mm mesh. Then, each sample was divided into two subsamples: one to be used for conventional laboratory analysis and the other for *ex situ* spectral measurements.

In each location, composite soil samples were obtained from the top (0-20 cm) of soil. A 4-cm diameter, stainless steel auger was used to take five samples from within a 0.5-m radius Soil cores were mixed, air dried and sieved through a 2-mm mesh. Each sample was divided into two subsamples: one to be used for conventional laboratory analysis and the other for *ex situ* spectral measurements.

Particle size analysis (fractions of sand, silt and clay) as well as SOM content were evaluated for each laboratory sample, as summarized in Table 6.1. The particle size analysis was conducted using hydrometers (Gee and Bauder, 1986; Soil Survey Staff, 2004) and SOM was determined using loss-on-ignition technique (Schulte and Hopkins, 1996).

Table 6.1 Statistical results of percent sand and clay, and SOM content in 19 soils

Statistics	% sand	% clay	% SOM
Minimum	26	2	3.9
Median	56	9	5.8
Maximum	93	33	25.6
Mean	56	13	7.2
Standard Deviation	18	9	4.8

The core of Veris P4000 instrument is a combined dual type spectrometer operating in the visible and near-infrared parts of the electromagnetic spectra (Fig. 6.3). One of the two spectrometers was USB2000 (Ocean Optics, Dunedin, Florida, USA) operating between 342 and 1023 nm with a spectral resolution of 6 nm. The other spectrometer was C9914GB (Hammatsu Photonics. K.K., Tokyo, Japan) collecting data between 1070 and 2220 nm with a spectral resolution of 4 nm. The instrument included its own light source and was capable of maintaining a constant

distance between measured soil surfaces and detectors by the means of a sapphire contact probe with fibre-optic cables.

All *ex situ* measurements were conducted as triplicates in a different randomised order each replicate. A specially designed sample holder (Fig. 6.3 right) was filled with <1 g of soil sample and placed near the optical window of the soil profiling tool. At the beginning of each spectral measurement session, the instrument was optimised and calibrated by measuring dark current followed by white reference measurements using the specially provided reference blocks. The instrument was re-calibrated every 20 samples. Soil spectra were interpolated to about 5 nm of spectral resolution, yielding a total of 380 data points (wavelengths) per spectrum. To minimize the instrument noise, each spectrum was the average of 25-30 scans (~6 scans s^{-1}).

Fig. 6.3 Illustration of the vis-NIR soil profiling tool during laboratory measurements

The *in situ* measurements were collected using recommended equipment setup (Fig. 6.4) for topsoil profiles down to 20 cm, while penetrating soil at a speed approximately 2 cm s^{-1}. Three measurements were conducted consecutively along a straight line less than 0.5 m long. Soil spectra data collection procedure for in situ and ex situ measurements was the same. However, in this case the average of 50-60 scans represented soil spectra collected at different depths from 0 to 20 cm.

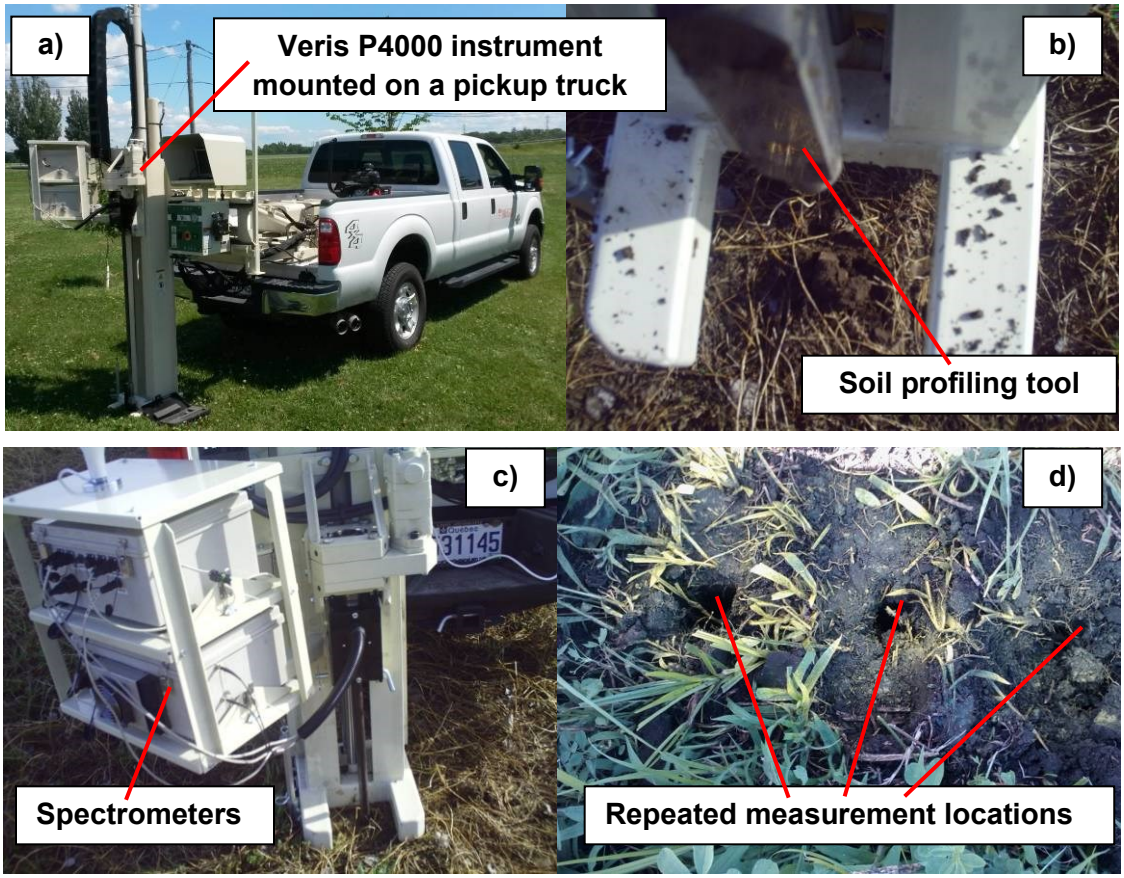


Fig. 6.4 Illustration of the vis-NIR soil profiling tool during *in situ* mneasurements

All raw spectral data was processed using MATLAB 2012a (The MathWorks, Inc. Natick, Massachusetts, USA) and ParLeS 3.1 software (University of Sydney, Sydney, Australia) described by Viscarra Rossel (2008). All spectra exhibited a step discontinuity from 1023 to 1070 nm, caused by the transition from one detector to another. After removal relatively noise parts of the spectra at the edges of detection ranges for each spectrometer (342-409 nm, 1014-1075 nm, and 2206-2220 nm) all resultant spectra consisted of the total 363 measurements at different wavelengths. The spectral data were then corrected for offset and processed using multiplicative scattering correction (MSC) algorithm (Geladi *et al.,* 1985) and mean centering (MC). In addition to these "original" spectra, the following spectra treatments were pursued: 1) 3-point Savitzky-Golay smoothing (Savitzky and Golay, 1964), 2) 11-point first derivative, and 3) 11-point second derivative.

6.2.2 Data Processing

Among the different statistics considered for assessing precision of spectral data, the ratio of spread over error (RSE) was used:

$$RSE = \frac{SD_{SA}}{RMSE} \tag{6.1}$$

where SD_{SA} is the standard deviation of nineteen sample mean values; RMSE is the root mean squared error calculated based on three replicated measurements for each of the nineteen samples.

$$SD_{SA} = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}} \tag{6.2}$$

where n=19 is the number of soil samples; X_i is an average measured or calculated spectra-related value for the i^{th} sample; X is the average of all X_i values.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\left(x_{ij} - X_i\right)^2}{n \cdot (m-1)}} \tag{6.3}$$

where m=3 is the number of replicated measurements; x_{ij} is a j^{th} replicate of the measured or calculated spectra-related value for the i^{th} sample.

$$X_i = \frac{\sum_{j=1}^{m} x_{ij}}{m} \tag{6.4}$$

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \tag{6.5}$$

Similar to the frequently used ratio of prediction over deviation (RPD), high RSE means a relatively strong ability of a given measurement to distinguish different soil samples. Use of RSE has been reported earlier in Adamchuk *et al.* (2006) and is directly related to ANOVA F-statistics used to compare the means of repeated measurements. Based on the degrees of freedom involved, the difference among the soil samples (means of three measurements) can be detected

at α=0.05 when RSE is greater than 0.79 ($\sqrt{\frac{1}{m}F_{stat}}$).This analysis evaluates measurement precision with the underlying hypothesis that a particular parameter that does not change when measuring the same soil sample and the change is at its maximum when measuring different samples should be considered reliable.

Percent sand, clay and SOM were predicted by fitting SLR models on each individual measured or treated spectral value versus these properties. A coefficient of determination (R^2) was the main indicator of the ability of a single spectral value to explain the variability in a particular soil property. However, the standard error of prediction (SEP) was used as a measure of the accuracy of soil property estimates obtained using each SLR model:

$$\hat{y}_{ij} = \beta_0 + \beta_1 x_{ij} \tag{6.6}$$

where y is the measured value of a given soil property for i^{th} sample; y is the predicted value of a given soil property for i^{th} sample and j^{th} replicate; β_0 (itercept) and β_1 (slope) are coefficients of SLR.

$$SEP = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\left(y_i - \hat{y}_{ij}\right)^2}{n \cdot m}} \tag{6.7}$$

where y is the measured value of a given soil property for i^{th} sample.

While SEP indicated the total error associated with each individual measurement and is primarily linked to the accuracy of prediction, RMSE is linked to the measurement precision. However, since RMSE is expressed in the units of spectra measurements and related calculated parameters, measurement precision (MP) can be expressed in physical units following:

$$MP = \beta_1 \cdot RMSE \tag{6.8}$$

When comparing different spectral wavelengths and transformation techniques, it is important to identify measurements that have maximum RSE and minimum RMSE, MP, and SEP. The RMSE is an indicator of measurement reproducibility. However, without considering the spread of values across different samples, it is impossible to conclude if given values is a strong value to

distinguish different soil samples from each other. Therefore, RSE is involved to elect candidates to differentiate between samples disregard of prediction property. Neither RMSE, nor RSE depend on the model used to predict a given soil property.

Because of SLR approach used to test one-input soil property prediction functions, RMSE can be expressed in terms of percent sand, clay or SOM as MP. The MP estimate is then evaluated together with SEP, which is the ultimate indicator of soil properties predictability. Unlike RMSE, MP as well as SEP can be compared across different spectra transformation procedure as both of them are expressed in physical units. From sensor development point of view, small RMSE and MP indicates stable soil-detector interface. High RSE means that the sensor can be applied for a particular set of soils. Finally, small SEP (high R^2 for a given set of samples) means sensor's ability to predict soil property of interest. SEP is always greater than MP and the greater this difference, the less uncertain linear relationship between the measured value and the property. Small difference between SEP and MP indicates applicability of the prediction model when reliable measurement estimates are obtained. In other words, small difference between SEP and MP indicated the potential for improved predictability by averaging multiple unbiased measurements, but larger difference means the limitation of the model and alternative prediction methods, such as PLSR, should be involved

In soil spectroscopy, PLSR is one of the most widely used techniques to aggregate measurements obtained at multiple wavelengths in a single prediction model. The PLSR is a bilinear regression technique that extracts a small number of latent factors, which are a combination of the independent variables, and uses these factors as a regression producer for the dependent variables (Geladi and Kowalski, 1986; de Jong and Kiers, 1992). The PLSR analysis is normally evaluated using the leave-one-out cross validation technique and RMSE, R^2 and Akaike Information Criterion (AIC) (Akaike, 1974) are the most common model performance indicators.

In this study, the orthogonalised PLSR-1 algorithm (Martens and Naes, 1989) was applied to 1) original spectra, 2) smoothed spectra, 3) first order derivative spectra, 4) second order derivative spectra, and 5) all the values combined to develop calibration models using ParLeS software (Viscarra Rossel, 2008). The number of factors to use in each model was selected using

leave-one-out cross validation. However, due to the limited number of soil samples used (n = 19), the selected models were not re-validated on a different set of soil samples, which is generally done in other reported studies. The developed models were used to estimate the performance indicators comparable to MP and SEP to define the superiority of PLSR over the SLR models.

6.3 Results and Discussion

Statistics on laboratory results for percentage sand, clay and SOM contents as listed in Table 6.1. These results point out the variability's observed over relatively small fields in terms of their texture and SOM content. It is interested to note that standard deviation of % sand in this dataset is two times higher than % clay and four times higher than % SOM.

Average reflectance spectra of nineteen soil samples, collected both *ex situ* and *in situ* are illustrated in Fig. 6.5. In general, the laboratory spectra exhibited higher reflectance as those from the field, which might be explained by dry versus wet soil. All spectra showed water absorption inversions near 1400 and 1900 nm. Less noticeable inversions were also observed in many other parts of the spectra (1100, 1500, 1600, 1700 and 2100 nm) which could be associated with primary or secondary effects of various minerals, carbonates and SOM. Several noticeable peaks were also observed at certain parts of the spectra, especially in the visible region that could be associated with soil color.

Fig. 6.6 illustrates the RSE values produced for original and all transformed spectra and Table 6.2, summarizes relevant statistical results. It is noticeable, that transforming the original soil spectra into 3-point Savitzky-Golay smooth and first and second derivative spectra, increased the reproducibility in many parts of the spectra. It is clear that the second derivatives had the highest relative reproducibility across the entire *ex situ* spectrum, while RSE for the first derivatives had the highest relative reproducibility across the entire *in situ* spectrum.

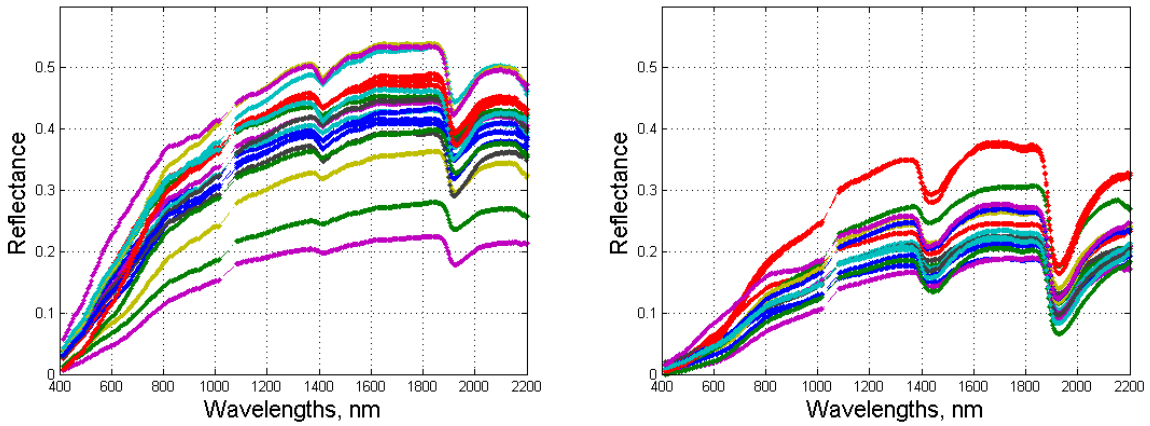


Fig. 6.5 Examples of *ex situ* (left) and *in situ* (right) vis-NIR soil spectra

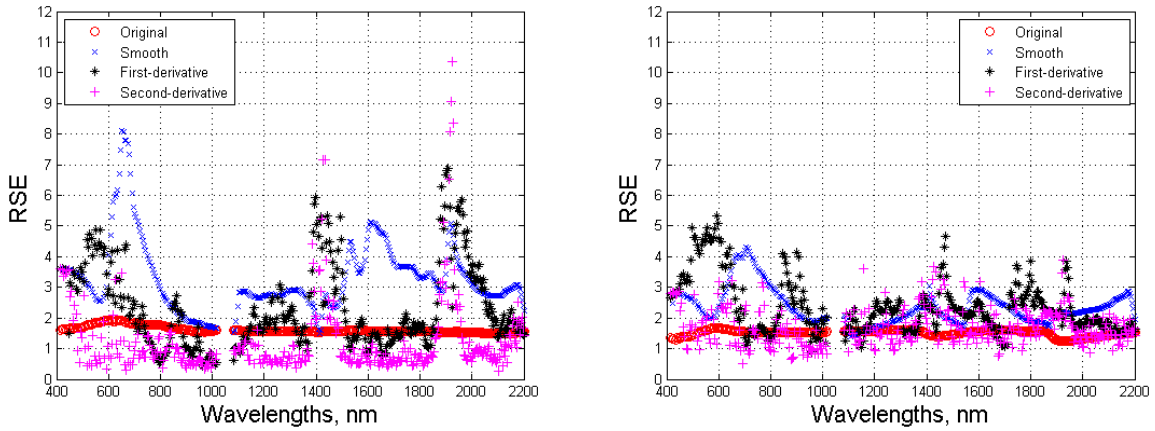


Fig. 6.6 RSE estimates for *ex situ* (left) *in situ* (right) soil spectra and their derivatives

One might expect that *ex situ* measurements were more reproducible than those conducted in natural soil conditions within a 0.5 m distance from each other. However, in both cases, first derivatives yielded data with comparable relative reproducibility in certain parts of the spectrum. In large, each transformed spectra had some portions with the ability to separate the means of the three replicated measurements between different soil samples (RSE > 0.79).

Figures 6.7-6.9 illustrate the coefficients of determination versus the RSE's, while predicting percentage sand, clay and SOM contents using SLR. Table 6.3, summarizes wavelengths and transformation methods indicating the highest level of predictability for the soil properties of

interest. Selection of these wavelengths was predominantly guided by the coefficient of determination as the weighting between R^2 and RSE is arbitrary.

Table 6.2 Ratio of spread over error (RSE) comparison

Test	Transformed Factors	Minimum	Median	Mean	Maximum
Ex situ	Original spectra	1.69	1.76	1.81	2.18
	Smooth spectra	1.62	3.04	3.37	8.71
	First derivative	0.47	1.91	2.49	6.93
	Second derivative	0.32	0.86	1.32	10.56
In situ	Original spectra	1.24	1.53	1.51	1.67
	Smooth spectra	1.38	2.23	2.35	4.30
	First derivative	0.89	2.14	2.33	5.34
	Second derivative	0.52	1.59	1.70	3.88

Alternatively, Figures 6.10-6.12 illustrate the SEP versus MP while predicting the soil properties. All values were calculated using both SLR and PLSR calibration models. To facilitate the comparison, the standard deviation of soil properties (laboratory measurements) was shown as a benchmark for SEP and the isolines represented positions in the chart with the constant sum of MP and SEP. To assure that both error terms are small, most promising wavelengths and spectra transformation methods were selected as those placed between two isolines indicating the lowest sum of MP and SEP.

Table 6.4 summarized these wavelengths and spectra transformation methods. Tables 6.5 and 6.6 lists key model performance indicators for each predicting variable listed in Table 6.4 along with all PLSR models. Fig. 6.13 points out selected wavelength using the average of all original spectra. It was observed that wavelengths near 1900 nm (short water absorbance band) transformed using either first or second derivative were found better correlated with percentage sand and clay contents as compared to the other parts of spectra. Smooth and first derivative spectra corresponding to different visible wavelengths indicated were superior when predicting SOM content.

From Tables 6.5 and 6.6, it is obvious that PLSR results had the lowest combination of MP and SEP with no overwhelming difference among the types of spectra transformation involved. In

most cases, PLSR that is based on second-order derivative spectra indicated the best performance and soil properties prediction did not improve when using all four methods of spectra transformation. This was an expected observation as PLSR uses several spectral factors and SLR just one, which is not capable to relate the difference in sensor response with a given soil property.

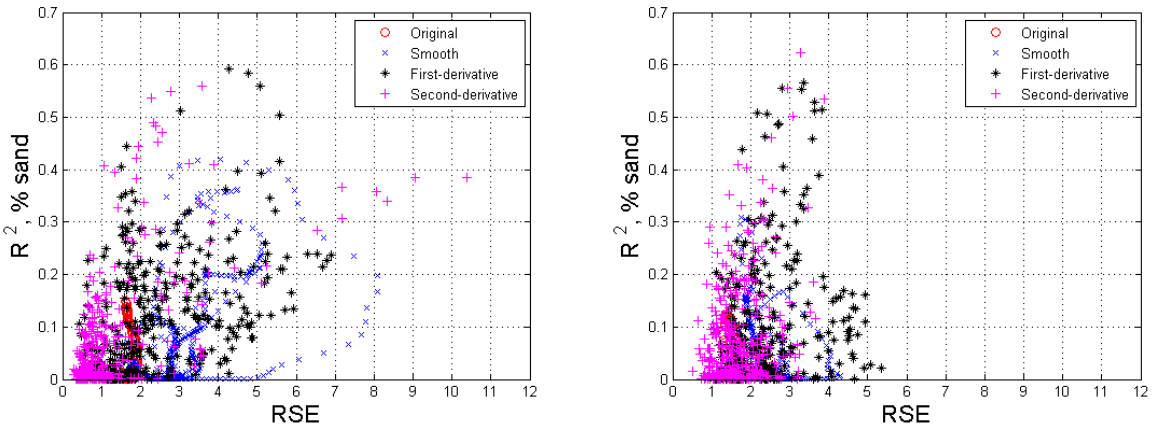


Fig. 6.7 % sand correlations verus the calculated RSEs for *ex situ* (left) and *in situ* (right) spectra and its derivatives

In general, results indicate that when using a single predicting factor, SEP for sand was about 10-12% with no significant difference between in situ and ex situ measurement. The MPs around 3% indicate relatively strong spectra reproducibility, which indicates the need for a more expanded model. This is apparent since SEP of PLRS is twice smaller indicating the ability to predict sand to within 4%. Percent clay, on another hand had 3-4% SEP and 1-2% MP indicating both reproducibility of the spectra and the ability of a single spectra parameter predict clay fairly accurate. In this case PLRS also reduced the errors two times, but only using first or second derivative spectra.

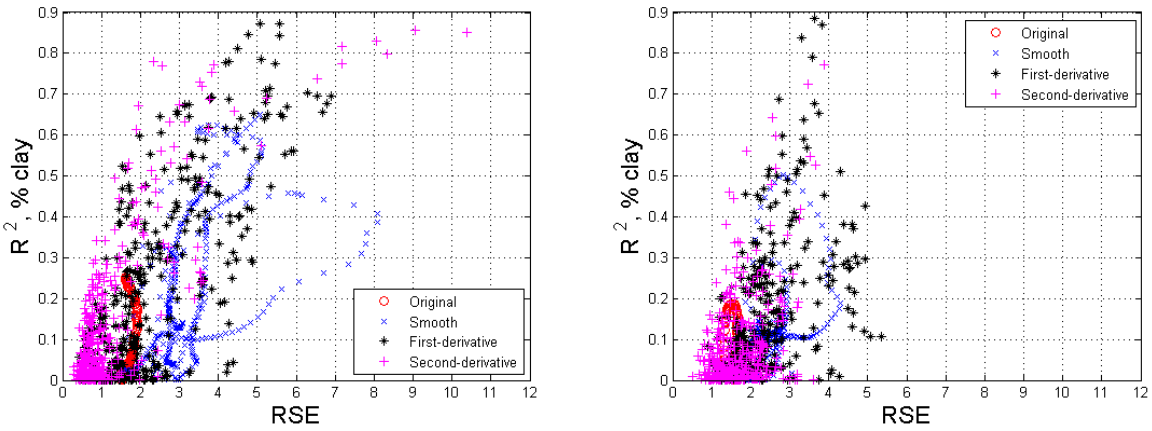


Fig. 6.8 % clay correlations verus the calculated RSEs for *ex situ* (left) and *in situ* (right) spectra and its derivatives

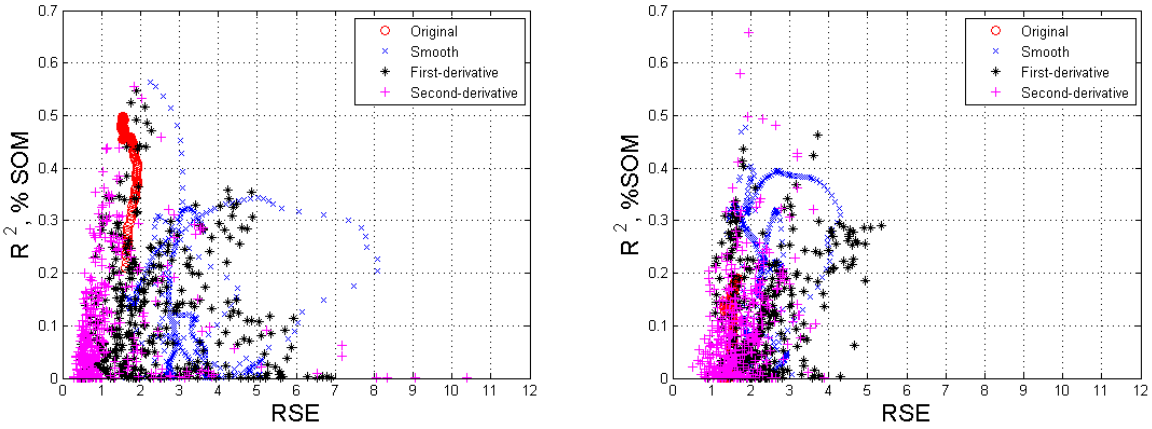


Fig. 6.9 % SOM content correlations verus the calculated RSEs for *ex situ* (left) and *in situ* (right) spectra and its derivatives

Finally, SEP for SOM was found only a quarter lower than the standard deviation of laboratory measurements indicating that SLR is not an appropriate model for this soil property for the given set of soils. PLSR has reduced SEP to about 1% SOM, which is reasonable considering the extensive spread of soil texture in this soil set. Similarly to the sand and clay, minor increase of error estimates corresponding to in situ versus ex situ measurements was much small than it was expected considering that different physical locations within 0.5 m distance was used each time.

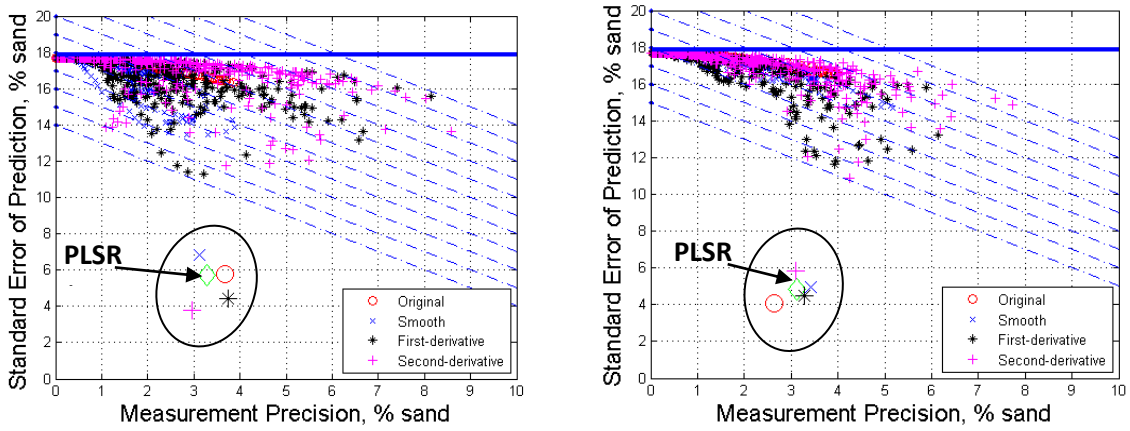


Fig. 6.10 Precision versus accurracy comparison for the vis-NIR soil spectra, their derivatives and PLSR when predicting % sand using *ex situ* (left) and *in situ* (right) measurements, (solid line - represents the standard deviation of sand content (laboratory measurements); isolines - represent the positions in the chart with the constant sum of MP and SEP; diamond symbol - represents PLSR results on all spectra combined)

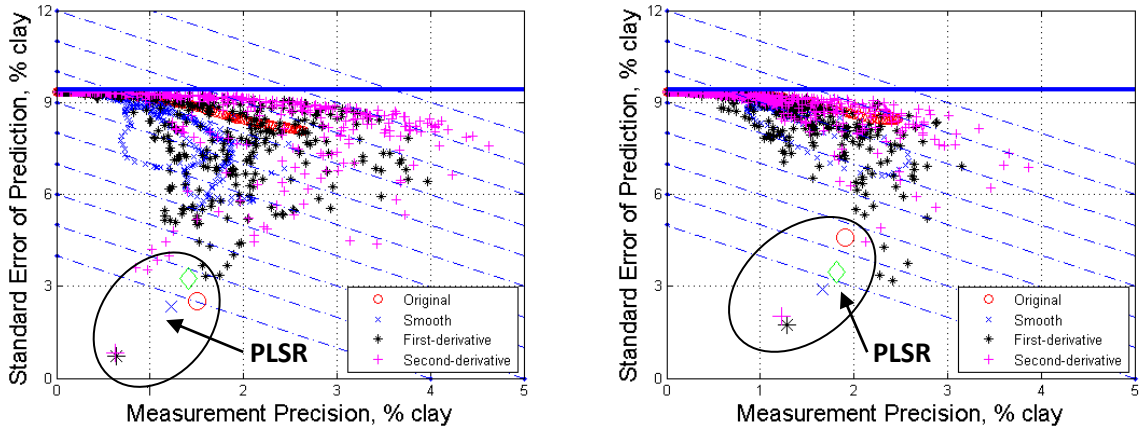


Fig. 6.11 Precision versus accurracy comparison for the vis-NIR soil spectral, their derivatives and PLSR when predicting % clay using *ex situ* (left) and *in situ* (right) measurements, (solid line - represents the standard deviation of clay content (laboratory measurements); isolines - represents the positions in the chart with the constant sum of MP and SEP; diamond symbol - represents PLSR results using all spectra combined)

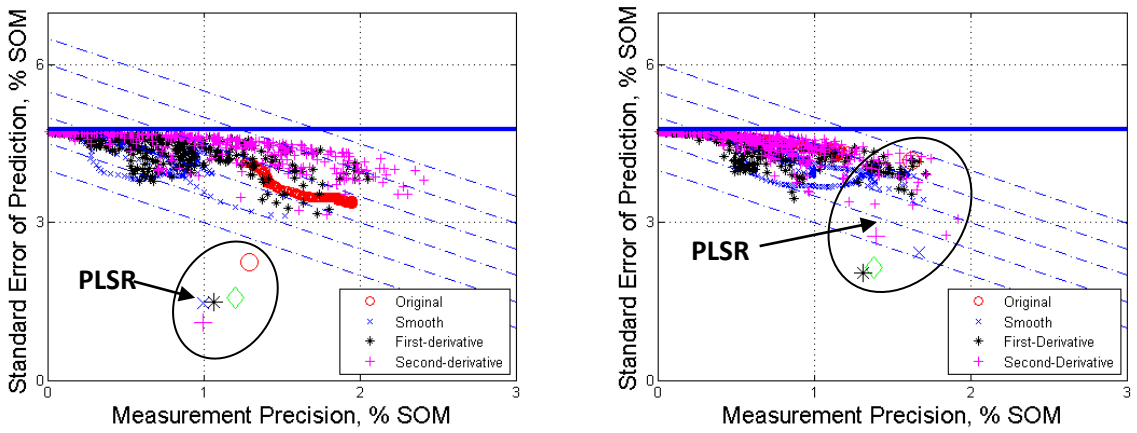


Fig. 6.12 Precision versus accurracy comparison for the vis-NIR soil spectral, their derivatives and PLSR when predicting SOM content using *ex situ* (left) and *in situ* (right) measurements[5].

Table 6.3 Vis-NIR wavelengths (nm) indicating the highest correlation (R^2) and ratio of spread over error (RSE) and with % sand, % clay and % SOM content

Soil property	Test	Original spectra	Smoothed spectra	First derivative	Second derivative
% sand	*Ex situ*	-	-	1932, 1936, 1940	1413
	In situ	-	-	1447, 1452, 1940, 1944	1457, 1462, 1466, 1923
% clay	*Ex situ*	-	-	1936, 1940, 1944, 1948	1433,1915, 1919, 1923
	In situ	-	-	1931, 1936, 1940, 1944	1919, 1923
% SOM	*Ex situ*	733, 786	712, 744	2114, 2117, 2121, 2148	-
	In situ	-	749, 770	-	602, 1738, 1742, 2159, 2182

[5] solid line - represents the standard deviation of % SOM content (laboratory measurements); isolines - represents the positions in the chart with the constant sum of MP and SEP; diamond symbol - represents PLSR results using all spectra combined

Table 6.4 Vis-NIR wavelengths (nm) indicating the lowest critical error's when predicting % sand, % clay and % SOM content

Soil property	Test	Smoothed spectra	First derivative	Second derivative
% sand	*Ex situ*	-	***1931***, **1936***, 1940, 1944	***1413***
	In situ	-	1452, 1457, 1936,1940, 1944	***1462***, 1923
% clay	*Ex situ*	-	1940, 1944, 1948	1915,1919, 1923
	In situ	-	1931, 1936, ***1940***, ***1944***	1923
% SOM	*Ex situ*	***674*,** 679, 685	551, ***556*,** 562, 568, 585	-
	In situ	421, 722, 728	421, 585, 590, ***1747***, 1751	1605, 543

*- Bold italics represents the most optimal spectral wavelengths

Table 6.5 The vis-NIR soil spectral bands found capable to exhibit the highest possible degree in terms of the precision and accuracy component, when predicting % sand, % clay and % SOM content during *in situ* measurement environments

Soil property	Spectra	Model	Wavelengths, nm	R^2	RSE	MP	SEP
% sand	Second derivative	PLSR	-	-	-	2.96	3.78
	First derivative	SLR	1931	0.59	4.26	3.22	11.31
	First derivative	SLR	1936	0.58	4.77	2.87	11.40
	Second derivative	SLR	1413	0.56	3.58	3.70	11.74
	First derivative	SLR	1940	0.56	5.06	2.65	11.75
	First derivative	SLR	1944	0.5	5.57	2.29	12.46
% clay	First derivative	PLSR	-	-	-	0.63	0.73
	First derivative	SLR	1944	0.87	5.57	1.59	3.33
	First derivative	SLR	1940	0.87	5.06	1.74	3.35
	Second derivative	SLR	1920	0.86	9.06	0.98	3.55
	Second derivative	SLR	1924	0.85	10.38	0.85	3.61
	Second derivative	SLR	1915	0.83	8.07	1.08	3.86
% SOM	Second derivative	PLSR	-	-	-	0.99	1.09
	First derivative	SLR	557	0.36	4.24	0.67	3.78
	First derivative	SLR	551	0.35	4.86	0.59	3.79
	First derivative	SLR	585	0.34	4.26	0.65	3.84
	First derivative	SLR	563	0.33	4.46	0.62	3.86
	Smooth	SLR	685	0.31	6.70	0.40	3.92
	First derivative	SLR	568	0.31	4.89	0.54	3.93
	Smooth	SLR	680	0.3	7.35	0.36	3.95
	Smooth	SLR	674	0.29	7.67	0.34	3.99

Table 6.6 Summary of vis-NIR factors demonstrating reproducibility, precision and accuracy during *in situ* measurements

Soil Property	Spectra	Model	Wavelengths, nm	R^2	RSE	MP	SEP
% sand	Original	PLSR	-	-	-	2.64	4.09
	Second derivative	SLR	1462	0.49	1.90	4.25	10.86
	First derivative	SLR	1452	0.40	3.75	3.93	11.67
	First derivative	SLR	1944	0.50	2.34	3.96	11.83
	Second derivative	SLR	1924	0.39	2.17	3.35	12.06
	First derivative	SLR	1940	0.56	2.74	3.54	12.13
	First derivative	SLR	1936	0.58	3.30	3.31	12.33
	First derivative	SLR	1457	0.36	4.31	3.50	12.35
% clay	First derivative	PLS	-	-	-	1.28	1.73
	First derivative	SLR	1940	0.87	2.74	2.41	3.17
	First derivative	SLR	1936	0.85	3.30	2.28	3.38
	First derivative	SLR	1944	0.87	2.34	2.58	3.75
	First derivative	SLR	1932	0.78	3.64	2.32	4.28
	Second derivative	SLR	1924	0.85	2.17	2.12	4.46
% SOM	First derivative	PLSR	-	-	-	1.31	2.05
	First derivative	SLR	1747	0.00	3.07	0.86	3.46
	Second derivative	SLR	1605	0.01	1.67	0.96	3.57
	First derivative	SLR	1752	0.00	2.67	0.85	3.59
	Second derivative	SLR	1544	0.00	1.39	0.96	3.59
	Smooth	SLR	422	0.28	2.77	0.74	3.76
	Smooth	SLR	728	0.33	3.87	0.72	3.78
	Smooth	SLR	723	0.34	3.97	0.69	3.82
	First derivative	SLR	591	0.00	4.60	0.48	3.98
	First derivative	SLR	422	0.00	2.80	0.51	3.99
	First derivative	SLR	585	0.00	5.07	0.52	3.99

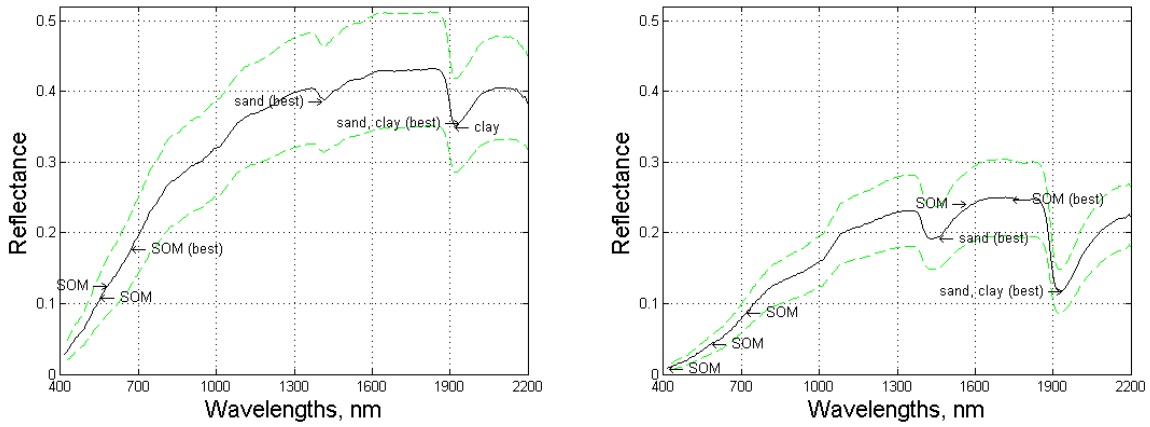


Fig. 6.13 The identified regions from the vis-NIR soil spectra capable to exhibit the highest possible degree in terms of the precision and accuracy component, when predicting % sand, % clay, and % SOM in *ex situ* (left) and *in situ* (right) measurment environments

6. 4. Conclusions

Based on the results of this study, it can be concluded that vis/NIR spectral measurements obtained using the commercially available equipment are compatible in terms of precision and accuracy when used under laboratory conditions or in the field and when soil water content varied substantially. Smoothing, first and second derivative transformation of the spectra has increased measurement reproducibility, which enhanced the system's ability to separate different soil samples. Naturally, this indicated the ability to predict soil properties that define the difference between soil samples. From tested soil attributes, clay content was shown to be predictable using spectral data derivatives around the short water absorption band. The prediction error was higher for sand and unacceptably high for SOM. Use of the popular PSLR technique reduced prediction errors, but a relatively small number of soil samples require caution when inferring these results in practice. The methodology of evaluating both the precision and accuracy components of spectral measurement errors was found to be useful when assessing performance of a given proximal soil sensing instrument.

Connecting Text to Chapter 7

Chapter 7 is also related to the second objective of this study as listed in Chapter 1. As discussed in Chapter 1, phosphorus in the top few centimetres of the soil surface may cause non-point source pollution when rain-driven soil erosion carries surface runoff to streams. As shown in Chapters 4 and 6, spectral data obtained in *ex situ* and *in situ* environments using commercially available portable vis-NIR equipment, showed promising results while predicting soil texture. This chapter discusses a study performed to evaluate the capability of vis-NIR spectral measurements in terms of predicting plant available phosphorus quantities. The soil data set was 86 archived soil samples. The results yielded a linear relationship between the predictions from the model and the reference measurements, with a coefficient of determination of 0.85 and a standard error of prediction of 28 mg kg^{-1}. The results of this study were presented at conferences shown below and at the time this thesis was submitted, a manuscript had been submitted to the "Canadian Biosystems Engineering Journal" as follows:

Dhawale, N.M., V.I. Adamchuk, R.A. Viscarra Rossel, S.O. Prasher, A.A. Ismail, & J.K. Whalen. 2013. Predicting plant available phosphorus using a vis-NIR spectral hyperspectral measurements obtained using a soil profiling tool. *Canadian Biosystems Engineering Journal* (under review).

Dhawale, N.M., V.I. Adamchuk, R.A. Viscarra Rossel, S.O. Prasher, A.A. Ismail, & J.K. Whalen. 2013. Predicting extractable soil phosphorus using Visible/Near-infrared hyperspectral soil reflectance measurements. *Paper No. 13-047*. Orleans, Ontario: CSBE.

Chapter 7

PREDICTNG PLANT AVAILABLE PHOSPHORUS USING VISIBLE/NEAR-INFRARED HYPERSPECTRAL MEASURMENTS OBTAINED USING A SOIL PROFILING TOOL

Dhawale, N., V.I. Adamchuk, R.A. Viscarra Rossel, S.O. Prasher, J.K. Whalen, A.A. Ismail

Abstract

Mehlich-3 extracting solution is widely used to assess plant available phosphorus ($P_{Mehlich-3}$) in Québec through colorimetric or inductively coupled plasma spectroscopic methods. While conventional analyses of $P_{Mehlich-3}$ is laborious and time-consuming, hyperspectral soil reflectance measurements provide a portable and low-cost alternative that can be performed *in situ*. The objective of this study was to evaluate the capability of a commercial soil profiling tool, equipped with a combined visible/near-infrared (vis-NIR) spectrometry system covering two spectral ranges from 342-1023 nm and 1070-2220 nm to predict $P_{Mehlich-3}$ using samples obtained from an agricultural field undergoing phosphorus fertility trials. Three repeated measurements of soil reflectance were collected using 86 homogenized air-dried soil samples. Using a spectrometer calibration procedure, each 376-band spectrum was transformed into a digital array of soil absorbance measurements. A partial-least-squares regression (PLSR) method was used to relate the averages of the three repeated spectra of each of the 86 soil samples to $P_{Mehlich-3}$ measurements. The results of the leave-one-out cross-validation of the spectral calibration model yielded a linear relationship between the predictions from the model and the reference measurements, with a coefficient of determination (R^2) equal 0.85 and a root mean squared error of $P_{Mehlich-3}$ prediction equal 28 mg kg^{-1}. The next step in this research will involve applying a similar data processing procedure to *in situ* measurements.

Keywords: Soil reflectance spectroscopy, soil available phosphorus, proximal soil sensing

7.1 Introduction

Agriculture-based non-point source pollution is an undesired effect of excessive fertilization traditionally practised to achieve higher yields. In result, it is of rising concern, due to the many cases of cyanobacteria and algal blooms affecting rivers, lakes and ponds across Canada (Canadian Council of Ministers for the Environment, 2007; Carpenter, 2008).

Phosphorus (P) is one among the three most essential crop nutrients added to agricultural soils by applying mineral fertilizers and manure. When not consumed by crop, excessive P may propagate to ground and surface waters due to leaching and runoff (Chambers *et al.,* 1997; Carefoot and Whalen 2003; Condron *et al.,* 2005; Jiao *et al.,* 2006; Enright *et al.,* 2009). Besides, phosphorous is a finite resource on earth and is available primarily through mining (Jasinski, 2011). According to Cordell *et al.* (2009), P fertilizer is going to reach peak production by 2030 and reserves are expected to be completely depleted in 50-100 years.

Conventionally, plant available P content is quantified through soil sampling and laboratory analysis. Mehlich-3 is a widely used extracting solution (Mehlich, 1984) for evaluating plant available phosphorus ($P_{Mehlich\text{-}3}$) in Québec soils. $P_{Mehlich\text{-}3}$ may be quantified using colorimetric or inductively coupled plasma (ICP) spectroscopic methods (Murphy and Riley 1962; Ziadi *et al.*, 2009).

Unfortunately, quantification of $P_{Mehlich\text{-}3}$ using conventional techniques is laborious and time-consuming, and, therefore, typical density of soil sampling is relatively low. Potentially, this limitation can be overcome using proximal soil sensing (PSS) (Viscarra Rossel, *et al.,* 2011), when an instrumented system is used to obtain a large number of measurements in field conditions. To date, there is no commercial PSS technology used to routinely produce high-density maps of $P_{Mehlich\text{-}3}$ prior to site-specific P management. Thus, a spectrometric method can be used to employs the interactions of visible and near infrared radiation with the sample under investigation.

The vis-NIR system is based on the sample's absorption of electromagnetic radiation at within a range from 400 to 2500 nm (Viscarra Rossel *et al.,* 2006b). A number of studies determining soil attributes by this method have resulted in the successful use of NIR spectroscopy to quantify soil organic matter, moisture, total C, total N, and cation exchange capacity. For example, Dalal and Henry (1986) simultaneously determined soil organic C, moisture, and total N by NIR spectroscopy. Similarly, there have been a number of attempts to predict P in soils by ultra violet (UV), vis, and/or NIR spectroscopy.

Krischenko *et al*. (1991) reported weak correlations between predicted and actual P ($R^2 = 0.42$). Similarly, Williams (2003) reported poor prediction of P using NIR spectroscopy. Chang *et al*. (2001) found an $R^2 = 0.40$. Thomsson *et al*. (2001), reported a similar value ($R^2 = 0.49$), noting that there were several spectral overtones for P in the wavelength range from 400 to 2475 nm in sandy loam, clay and silt clay soils. He *et al.* (2005) reporting an $R^2 = 0.46$, offered the explanation that this low correlation could be the result of a significant overlap in the signals for both P and C-H-O-N bonds.

Further, Bogrekci and Lee (2005) investigated the effects of common soil P compounds on reflectance spectra of sandy soils using UV, vis, and NIR reflectance spectroscopy. They added P to the sandy soils in four different forms ($FePO_4 \cdot 2H_2O$, $Mg_3(PO_4)_2 \cdot 2H_2O$, $CaPO_4$, and $AlPO_4$) and in seven different proportions. The authors reported successful prediction of P in sandy soils for all four compounds ($0.48 \leq R^2 \leq 0.73$) and found the strongest absorption peaks for $FePO_4 \cdot 2H_2O$, $Mg_3(PO_4)_2 \cdot 2H_2O$, $CaPO_4$, and $AlPO_4$ occurred at 286, 2548, 2516, and 228 nm, respectively. Maleki *et al*. (2006) reported soil P predictions for 2 different data sets with an $R^2 = 0.75$ and 0.73, with soil texture classes ranging from silt loam, to sandy loam. Wetterlind *et al.* (2010) reported soil P predictions for 2 different set of soils with an $R^2 = 0.48$ and 0.78. Kodaira and Shibusawa (2013) reported soil P predictions with an $R^2 = 0.72$.

The objective of this study was to evaluate the capability of a commercial, soil profiling tool equipped with a combined visible/near-infrared spectrometry system covering a spectral range from 342 to 2220 nm to predict $P_{Mehlich\text{-}3}$ in the laboratory, using samples obtained from an agricultural field in Québec undergoing P fertility trials.

7.2 Materials and Methods

7.2.1 Soil Samples

Soil samples were selected from a set of 1440 soil samples originally collected for a different study that involved P fertility trials (Carefoot and Whalen, 2003). Soil samples were collected in the fall after harvest, but before fall tillage, from 0-15, 15-30, and 30-60 cm depth intervals using a tractor-mounted soil auger. The samples at each depth were composites of two cores (7.5-cm diameter) obtained from each split plot area (Carefoot *et al*., 2003). Soils were then dried in a forced-air oven (60 °C for 48 h), ground, and passed through a 2-mm mesh sieve.

Soils nutrients were extracted using Mehlich-3 solution (1:10 soil: solution) after shaking for 5 min at 130 rpm. Phosphate concentrations in Mehlich-3 extracts were evaluated by the molybdenum blue reaction (Holman, 1943). To cover a wide range of $P_{Mehlich-3}$ content (Fig. 7.1) a subset of 86 soils samples was selected. A summary of $P_{Mehlich-3}$ concentrations for the selected 86 soil samples is provided in Table 7.1.

7.2.2. Instrumentation

A commercial soil profiling tool (Veris® P4000, Veris Technologies, Inc., Salina, Kansas, USA[6]) equipped with a combined, dual type spectrometer instrument (Fig. 7.2), operating in the visible and near-infrared regions of the electromagnetic spectrum, was used in this study. This instrument can be operated both *in situ* and *ex situ*. One of the two spectrometers was used to collect soil absorbance data between 342 and 1023 nm and 7 nm of spectral resolution, and the other spectrometer measured between 1070 and 2220 nm and 4 nm of spectral resolution. The instrument included its own halogen light source (2700°K) and was capable of maintaining a constant distance between the measured soil and the fibre optic probes using a designated sapphire window.

[6] Mention of a trade name, proprietary product, or company name is for presentation clarity and does not imply endorsement by the authors or McGill University and CSIRO, nor does it imply exclusion of other products that may also be suitable.

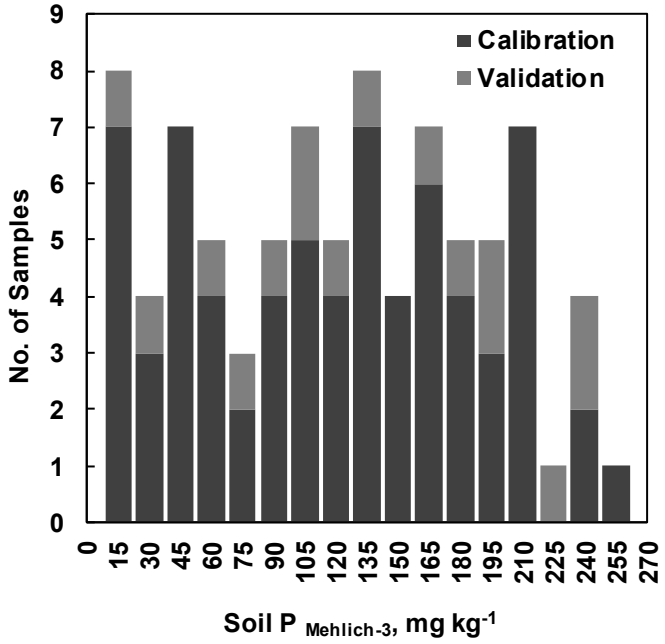


Fig. 7.1. Distribution of soil samples according to their $P_{Mehlich-3}$ content

Table 7.1 Statistical results of $P_{Mehlich-3}$ (mg kg^{-1}) observed in the 86 soil samples used in the study

Number of samples	Texture	Depth	Minimum	Mean	Maximum	Standard deviation (SD)
31	silty- loam	0-15 cm	96	177	244	38
31	sandy-silty- loam	15-30 cm	44	109	196	40
21	sandy-clay	30-60 cm	4	32	154	32

7.2.3. Methods

Three sets of *ex situ* measurements of soil absorbance were collected from each of 86 soil samples in random order. The data collection process involved filling about ~1 g of soil into a customized sample holder (radius of 0.5 mm and thickness of 5 mm) and placing it in contact with an optical window of area ~707 mm^2. To minimize the instrument noise, each spectrum was recorded as an average of 30-32 scans. At the beginning of each set of measurements and after every 20 samples, the instrument was recalibrated by measuring dark current followed by white

reference measurements using the specially provided reference blocks. This yielded a total of 376 data points (wavelengths) with about 4-7 nm of spectral resolution.

Fig. 7.2 Vis-NIR setup in bench top configuration

7.2.4 Processing of the Spectra

All raw absorbance soil spectral measurements were transformed into reflectance measurements ($10^{-absorbance}$) and the quality of the vis-NIR soil reflectance spectra were improved by removing the noisy parts (tails) observed in the spectra corresponding to wavelengths 342-373, 1018-1023, 1070-1075, 2216-2220 nm. Three replicated spectra were averaged for each soil sample to represent raw averaged (RA) soil spectra. A 3-point median filter was applied to the RA soil spectra to obtain average denoised spectra. Then the first (I) derivatives (DER) using 13 consecutive wavelengths were taken according to Savitzky and Golay (1964). This has brought the total number of spectra-based quantise equal 722 (RA, FD spectra).

Table 7.2. Statistical results of $P_{Mehlich-3}$ (mg kg^{-1}) observed in the calibration and validation examples

Number of samples	Set	Minimum	Mean	Maximum	Standard deviation (SD)
70	Calibration	4	112	244	67
16	Validation	4	127	228	71

7.2.5 Modeling

After a simple regression analysis relating each individual measure and $P_{Mehlich-3}$ was performed, a partial least squares regression (PLSR) method was used to determine their combination that could better predict $P_{Mehlich-3}$. The 86 soils samples were randomly divided into two data sets consisting of 70 calibration samples and 16 validation samples and the ParLeS software (version 3.1, 2007, University of Sydney, Sydney, Australia) (Viscarra Rossel, 2008) was used for calibration model development. Table 7.2 shows the statistical results with fairly even distribution of $P_{Mehlich-3}$ in both sets of soil samples.

PLSR is a bilinear regression technique that extracts a small number of latent factors, which are a combination of the independent variables *i.e.* reflectance/absorbance (at spectral wavelengths/wavenumbers), and uses these factors as dependent variables or chemical laboratory measured reference values in the regression. The PLSR analysis was tied to the leave-one-out cross-validation technique, which is essential to the method to construct a steadfast prediction model of $P_{Mehlich-3}$ in soils.

During $P_{Mehlich-3}$ spectral model calibration, several data pre-treatment options (available in the ParLeS application) were applied to the calibration examples. The accepted selection of a pre-treatment was based on comparing the statistical results of the different models. This process resulted in two models. In the first few efforts to develop a $P_{Mehlich-3}$ model, no pre-treatment after denoising the RA spectra (Model-1) and the first derivative (FD) of denoised RA spectra using Savitzky and Golay (1964) followed by mean centering (MC) technique (Model-2) yielded into the best results.

7.2.6. Statistical Analysis

Model performances were evaluated using root mean square error (RMSE), coefficient of determination (R^2) and Akaike information criterion (AIC) (Akaike, 1974). An additional statistical quality parameter; the relative percent deviation (RPD) was used to evaluate the calibration models (Williams, 2001). The RPD is the ratio of standard deviation (SD) of the measured value of soil properties to the value for RMSE of prediction or the RMSE of validation (Stenberg *et al.,* 2004). For samples of heterogeneous material such as soil, the following levels of performance are defined: calibrations with values of $R^2 > 0.95$ and RPD > 4 are outstanding, values for R^2 of 0.90-0.95 and RPD of 3.00-4.00 are high, values for R^2 of 0.80-0.90 and RPD of 2.25-3.00 are moderately high and values for R^2 of 0.70-0.80 and RPD of 1.75-2.25 are moderate (Chang *et al.,* 2001; Williams, 2001; Nduwamungu *et al*., 2009).

7.3 Results and Discussion

Fig. 7.3 illustrates the RA vis-NIR spectrum between 379-2212 nm collected on three air-dried laboratory soil samples containing three different $P_{Mehlich-3}$ concentrations.

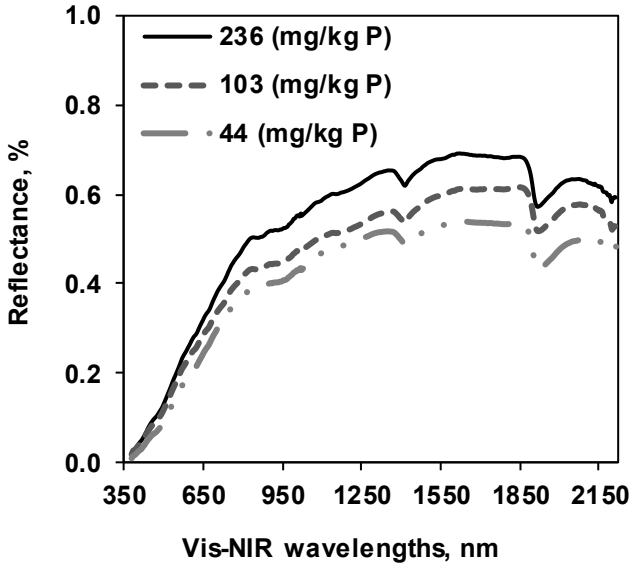


Fig. 7.3 Vis-NIR spectra illustrating differences when collected from soil samples having three different $P_{Mehlich-3}$ concentrations

As illustrated in Fig. 7.3 significant differences were observed between the vis-NIR spectra and the three different $P_{Mehlich-3}$ concentrations. Fig. 7.4 illustrates the R^2 values of RA and FD vis-NIR spectra. A value of $R^2 = 0.59$ was found around 600 nm with the RA vis-NIR spectra. In addition, significantly stronger values of $R^2 > 0.75$ were observed at several parts of the FD vis-NIR spectra (827, 897, 2037-2044 nm).

Fig. 7.5 shows the AIC values corresponding to the number of PLSR factors reported by ParLeS, for both the models. The PLSR factor numbers, corresponding to the lowest AIC values were chosen for the models. The statistical results illustrating the performances of both $P_{Mehlich-3}$ prediction models are summarised in Table 7.3. The linear relationships between laboratory measurements and predicted $P_{Mehlich-3}$ by both vis-NIR models in the calibration stage are shown in figures 7.6a-b. The predictions of $P_{Mehlich-3}$ resulting into negative values are forced to zero.

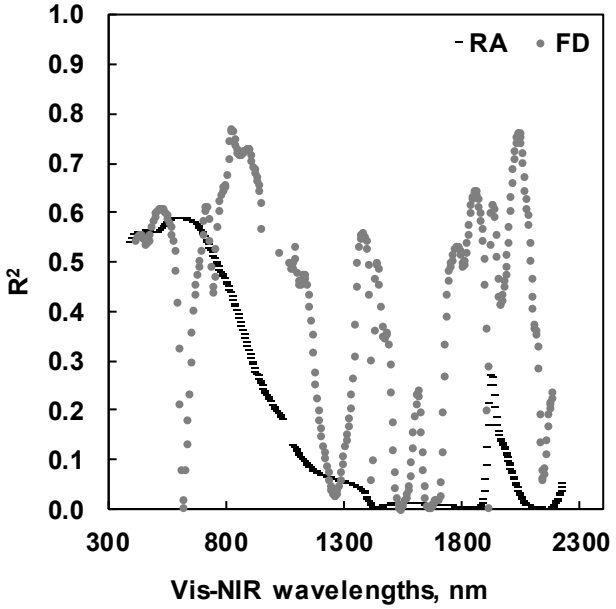


Fig. 7.4 Coefficient of determination (R^2) between vis-NIR spectra and $P_{Mehlich-3}$

The R^2 and RPD were either 0.83 and 2.42 or 0.85 and 2.37, when no pre-treatment was applied (Model-1) or when the denoised RA spectral data was pre-treated by the first derivative followed by mean centering (Model-2). The later pre-treatment also resulted in the lowest

RMSE = 28 mg kg^{-1} $P_{Mehlich-3}$. Based on these results both of the vis-NIR spectral models can be categorised as being moderately useful for predicting $P_{Mehlich-3}$ (Chang *et.al.* 2001). It is also important to note that these models covered a greater range of soil types (silt loam, sandy silt loam and sandy clay) as compared to earlier relevant studies (Bogrekci and Lee, 2005; Maleki *et al.,* 2006; Kodaira and Shibusawa, 2013; Wetterlind *et al.,* 2010).

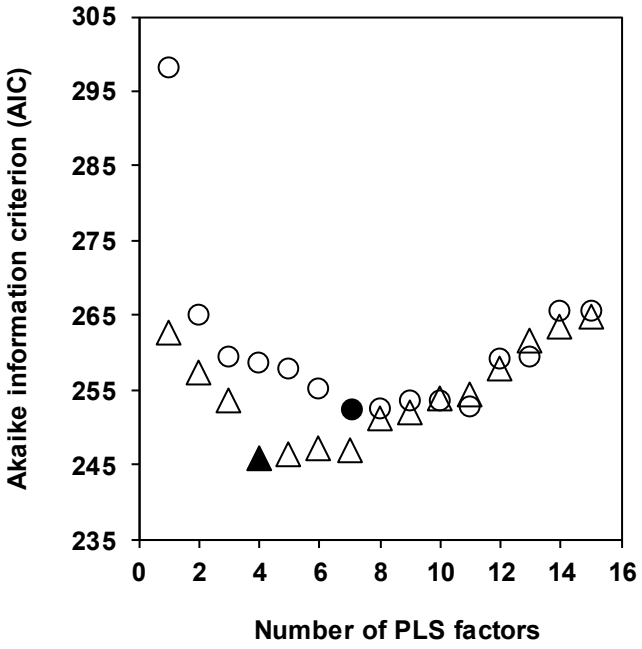


Fig. 7.5 Selection of the PLSR model factors, (Circles: averaged raw calibration data; Triangles: first derivatives and mean centered on averaged raw data).

Table 7.3 Statistical results of PLSR model calibration and validation, developed and tested using vis-NIR soil spectral data

Statistics*	Model-1		Model-2	
Data set	Calibration	Validation	Calibration	Validation
SD, mg kg^{-1}	67	71	67	71
Number of PLSR factors	7	7	4	4
RMSE, mg kg^{-1}	30	29	30	28
L CI RMSE, mg kg^{-1}	-	22	-	22
H CI RMSE, mg kg^{-1}	-	45	-	46
RPD	2.22	2.42	2.23	2.37
R^2	0.80	0.83	0.80	0.85

*SD - standard deviation; RMSE - root mean square error; L CI - lower confidence interval; H CI - higher confidence interval; RPD - relative percentdeviation; R^2 - coeefficient of determination.

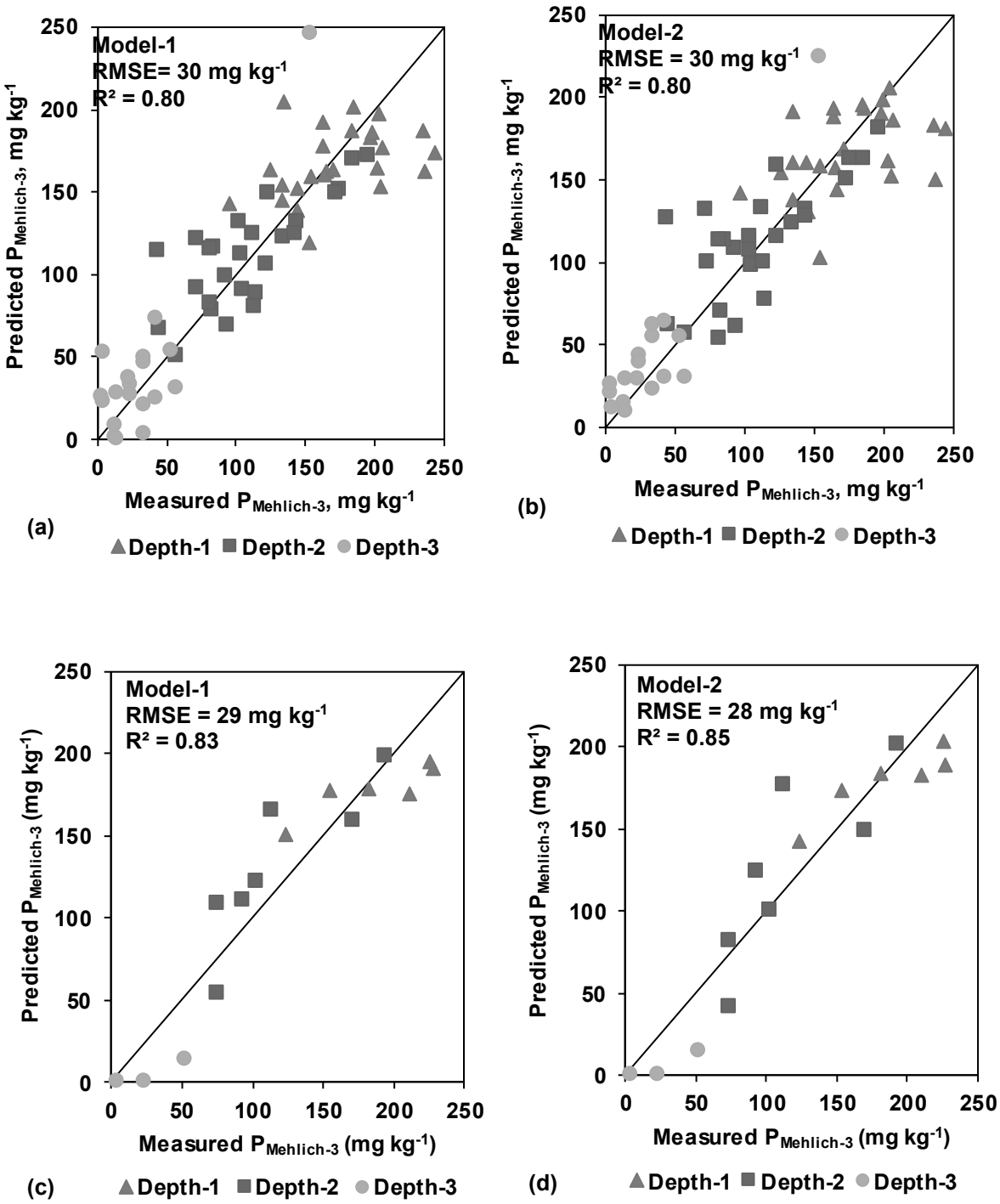


Fig. 7.6 P Predicted versus laboratory chemical measured soil available phosphorus ($P_{Mehlich-3}$) for 86 samples showing, (a-b) PLSR Model 1 and Model-2 calibration with leave one out cross validation and (b-c) Model-1 and Model-2 validation, R^2 - coefficient of determination, RMSE - root mean square error

The PLS regression coefficients (beta) for the best model (Model-2) are shown in Fig. 7.7. The size of the regression coefficients (negative or positive) represents the importance of the portable vis-NIR wavelengths being used for the prediction of $P_{Mehlich-3}$. In the vis range, the higher peaks for $P_{Mehlich-3}$ for soils used in the calibration set and in Model-2 were observed at wavelengths of 439, 457, 463, 545, 596, 635 and 738 nm whereas, in the NIR range, most high peaks were observed at wavelengths of 806, 1000, 1009, 1374, 1418, 1873, 1911, 2032 and 2174 nm. The highest peak in the vis region was found at 635 nm and in the NIR region at 1418, 1911 and 2174 nm. Among these wavelengths, 439-457 and 1418 could be compared with 421, 441, 448, 454 and 1464 found by Maleki *et al.* (2006). Wavelength, like 806, 1418, 1911 and 2174 can be compared with 867, 1419-1464, 1938-1948 and 2211-2152 found by Bogrecki and Lee (2005). In general bands near 600 nm are associated with soil color and bands near 1400 and 1900 are associated with soil water, which can be supported for these results.

Unfortunately, there is no proof available in the literature concerning direct absorption by P in the vis-NIR region (Maleki *at al.,* 2006) as the absorption of NIR radiation by organic molecules is due to overtone and combination bands primarily of C-H, N-H and O-H groups whose fundamental band is related to molecular stretching occurring in the mid-infrared (mid-IR) spectral region (Viscarra Rossel *et al.,* 2006b). However, many successful studies reported earlier and including this are perhaps due to the existing and hidden relations between soil P and other soil constituents, as discussed by Chang *et al.* (2001).

According to Maleki *et al*. (2006), all the correlations for $P_{Mehlich-3}$ in the vis and NIR region are indirect correlations with soil components that bind with P and show spectral activity. P binds with oxides present in sandy soils, and in sandy loam soils, like calcium carbonate, will bind P as calcium phosphate. The presence of moisture could also darken the color of these complexes which in turn would be responsible for a better correlation between the P complex and the spectral signal in the vis and NIR ranges. On the other hand, since the soils were treated with inorganic fertilizers the presence of calcium phosphate might have been detected by the NIR spectrometer for its corresponding wavelength and since the soils were treated with manure the presence of organic complexes with P could also be responsible for a better correlation between organic P and the spectral signal in the vis range. Through future research, a detailed analysis of

each indicative wavelength and $P_{Mehlich-3}$ concentration along with the influence of other soil attributes needs to be carried out.

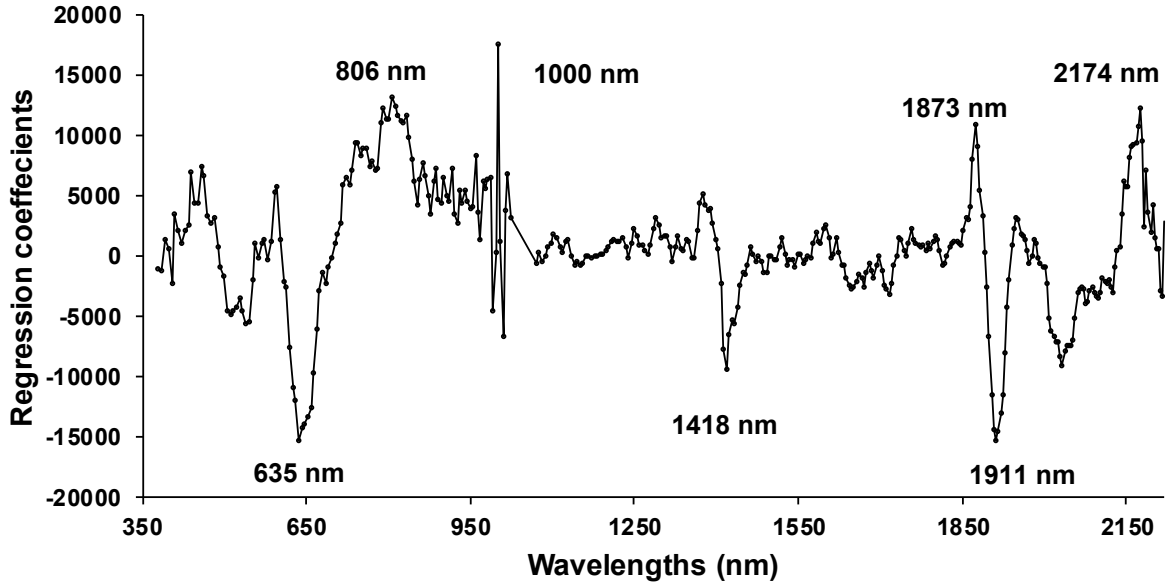


Fig. 7.7 Regression coefficients (betas) of the PLSR Model-2

7.4 Conclusions

In this study soil absorbance was measured by using a commercial soil profiling tool equipped with a combined dual spectrometer in the range of 342-2220 nm of the visible (vis) and near infrared (NIR) spectrum was combined with partial least squares regression (PLSR) leave-one-cross-validation techniques to correlate soil reflectance with plant available phosphorus ($P_{Mehlich-3}$) in a calibration data set of air-dried soil samples. Two methods of pre-processing the averaged spectral data resulted in two steadfast models using vis-NIR spectral data in the range 379-2220 nm. The results of the leave-one-out cross-validation of the spectral calibration model yielded a linear relationship between the predictions from the model and the reference measurements, with $R^2 = 0.85$ and RMSE = 28 mg kg^{-1} for the best model. The next step in this research will involve applying a similar data processing procedure to *in situ* measurements on the same and different fields.

Connecting Text to Chapter 8

The importance of site specific soil and crop management and the role of proximal soil sensing have been addressed in the first two chapters. As shown in Chapter 3, an on-the-spot soil analysing platform has been developed to make point-based measurements of multi-soil chemical properties in field. Two independent sets of active hyperspectral sensor systems were evaluated and compared in terms of their potential use for OSA. The results have been discussed in Chapters 4 through 6. Chapter 8 is related to the third objective listed in Chapter 1. In this chapter, a novel algorithm has been developed to combine multilayer high-density spatial data, obtained using proximal soil sensing systems. Apparent soil electrical conductivity and field topography obtained using on-the-go sensing platforms are used to illustrate the algorithm. As a result, the algorithm was found capable to partition several tested fields into spatially contiguous and relatively homogenous areas (groups), representing the measured variability between input data layers to be nearly constant within the formed groups. Due to this unique feature, the algorithm is anticipated to maximize the value of prescribed point-based measurements by OSA or other similar platforms, however it requires more research. Different parts of this study were presented at the conferences listed below and a manuscript has been prepared for the "Geoderma".

Dhawale, N.M., V.I. Adamchuk, S.O. Prasher, P.R.L. Dutilleul, & R.B. Ferguson. 2014. Spatially constrained data clustering for multilayer high-density proximal soil sensing data. *Geoderma* (to be submitted).

Dhawale, N.M., V.I. Adamchuk, S.O. Prasher, P.R.L. Dutilleul, & R.B. Ferguson. 2014. Spatially constrained geospatial data clustering for multilayer sensor based measurements. *Paper No. MTSTC2-127.* ISPRS/IGU Joint International Conference on Geospatial Theory, Processing, Modelling and Applications, 6-8 October 2014, Toronto, Canada.

Dhawale, N.M., V.I. Adamchuk, S.O. Prasher, & P.R.L. Dutilleul, 2012. Spatial data clustering using neighbourhood analysis. *Paper No. 121337939.* St. Joseph, Michigan: ASABE.

Chapter 8

SPATIALLY CONSTRAINED DATA CLUSTERING OF MULTILAYER HIGH-DENSITY PROXIMAL SOIL SENSING DATA

Dhawale, N., V. I. Adamchuk, S. O. Prasher, P. R. L. Dutilleul, R. B. Ferguson

Abstract

One of the most popular approaches to process high-density proximal soil sensing data is to aggregate similar measurements representing unique field conditions. An innovative, constraint-based spatial clustering algorithm has been developed. The algorithm seeks to minimize the mean squared error during the interactive grouping of spatially adjacent measurements similar to each other and different from other parts of the field. After successful implementation of a one soil property scenario, the goal of this research was to accommodate multiple layers of soil properties representing the same area under investigation. Six agricultural fields across Nebraska, USA, were chosen to illustrate the algorithm's performance. Field elevation and apparent soil electrical conductivity at deep and shallow depths were investigated. The algorithm was implemented in MATLAB, R2013b. Prior to the process of interactive grouping, geographic coordinates were projected and erroneous data were filtered out. Additional data pre-processing included bringing each data layer to a 20 x 20 m raster to facilitate multilayer computations. An interactive grouping starts with a new "nest" search to initiate the first group of measurements which includes those that are most different from the rest of the field. This group is expanded using a neighbourhood search approach and once the group fails to reduce the overall mean squared error, the algorithm seeks to locate a new "nest", which will grow into another group. This process continues until there is no benefit from separating out additional parts of the field. Results of the six-field trial showed that each case generated a reasonable number of groups which corresponded to agronomic knowledge of the fields. The unique feature of this approach is spatial continuity of each group and the capability to process

multiple data layers. Further development will involve comparison with a more traditional k-means clustering approach and agronomic model calibration using targeted soil sampling.

Keywords: precision agriculture, proximal soil sensing, geospatial data clustering, management zones

8.1 Introduction

While conventional soil sampling techniques are laborious and time consuming, proximal soil sensing (PSS) allows rapid and inexpensive collection of high-density data (Viscarra Rossel and McBratney, 1998; Viscarra Rossel *et al.,* 2010). To pursue various site-specific management practices, spatial data is frequently split into groups (clusters or zones) to represent significantly different growing conditions (Fraisse *et al.,* 2001; Ping and Dobermann, 2003). Geo-spatial data clustering is an important process (Li and Wang, 2010), which is widely used in remote sensing (Deng *et. al.* 2003), neuroanatomy analysis (Prodanov *et. al,* 2007), and other areas. Several different spatial clustering algorithms have been developed to group geospatially dense PSS-based measurements of soil attributes into management zones.

For example, Management Zone Analyst (Fridgen *et al.,* 2004) represents a publicly available tool accepted by a number of practitioners. The algorithm is based on computing a distance matrix and performing clustering over this new distance matrix. It is closely related to the popular k-means clustering algorithm, where quality of the resulting clusters heavily depends on the selection of initial centroids and the results are not repeatable. However, this method requires cross-validation to select the best, among several runs. (Abdul-Nazeer and Sebastian, 2009). Although the method provides multidimensional data analysis, complexity and frequently occurring discontinuities of management zones make this technology non-robust for potential users (Kerby *et al.,* 2007; Shatar and McBratney, 2001).

Spatial continuity of formed clusters can be achieved by restricting grouping measurements that are not adjacent to each other (Dhawale *et al.,* 2012) through so called Neighbourhood Search Analysis (NSA). This is a form of clustering built on the principle of growing new groups of data points or grid cells with a fixed size through minimization of the mean squared error (MSE). Since previous trials with one measured soil attribute revealed positive outcomes, the objective of this study was to advance an algorithm to allow multiple data layers to be used for delineating spatially constrained groups of high-density soil sensor-based measurements. Field elevation and apparent soil electrical conductivity (ECa) at two depths obtained from six agricultural fields with different levels of spatial structure were used to illustrate the performance of this algorithm.

8.2 Materials and Methods

8.2.1 Data Collection

Six production fields in Nebraska were mapped using a Veris®3100[7] (Veris Technologies, Salina, Kansas, USA) galvanic contact soil EC_a mapping unit equipped with an RTK-level, AgGPS 442[8] (Trimble Navigation Ltd., Sunnyvale, California, USA) and a global navigation satellite system (GNSS) receiver. The three data layers were: 1) field elevation 2) deep soil EC_a (~0-90 cm) obtained with a wide pair of Wenner array electrodes and 3) shallow soil EC_a (~0-30 cm) obtained using a narrow pair of electrodes. Table 8.1 summarises data from the six fields. It has been noted that the two layers of EC_a represent similar but not identical spatial patterns, while field elevation does not always correspond to the overall pattern of changing EC_a. Therefore, the ideal map of field partitioning would delineate areas with different combinations of the three values significantly different from the average field conditions.

8.2.2 Data Pre-processing

All data processing was accomplished using MATLAB R2013b (The MathWorks, Inc., Natick, Massachusetts, USA). To obtain three 2D matrices representing each field, sensor-based data pre-processing involved four steps: 1) removing erroneous data using predefined threshold values of physically feasible measurements, 2) 1D data smoothing using a 5-point moving average technique, 3) projection of local coordinates according to Adamchuk (2001), and 4) 20 x 20 m averaging of all measurements inside each grid cell. Field elevation data were relative to the lowest grid cell found in every field. The resulting rectangular matrix representing each field covered the entire spatial domain. Grid cells outside field boundaries were assigned zero values. Therefore, no grid cells inside the fields were without corresponding sensor measurements. Smaller grid cell size would also be possible, but would require more

[7] Mention of a trade name, proprietary product, or company name is for presentation clarity and does not imply endorsement by the authors or McGill University, nor does it imply exclusion of other products that may also be suitable.

[8] Mention of a trade name, proprietary product, or company name is for presentation clarity and does not imply endorsement by the authors or McGill University, nor does it imply exclusion of other products that may also be suitable.

computation power. The selected resolution using a total of 600-1500 grid cells per field was considered reasonable to reveal field macro-variability.

Table 8.1. Statistical summary of the proximal soil sensing (PSS) data from agricultural fields

Field ID	Area, ha	Mean	Range	SD
		Field elevation, m		
1	25	1.50	3.20	0.53
2	46	4.95	17.82	3.99
3	50	7.10	11.54	2.09
4	55	8.07	27.44	5.68
5	67	4.22	8.09	1.60
6	44	6.15	10.59	2.15
		Shallow ECa, mSm^{-1}		
1	25	0.73	1.58	0.28
2	46	3.99	13.14	1.67
3	50	6.21	11.64	1.84
4	55	2.44	9.04	1.72
5	67	7.25	9.32	1.88
6	44	2.29	7.42	0.82
		Deep ECa, mSm^{-1}		
1	25	7.62	27.66	3.76
2	46	30.24	86.90	14.39
3	50	4.10	8.68	1.71
4	55	16.31	61.97	12.06
5	67	51.01	80.77	14.07
6	44	25.72	81.74	14.36

8.2.3 Data Clustering Algorithm

The data clustering algorithm was constructed using the assumption that treating a group of adjacent grid cells separately from the rest of the field would reduce the MSE between individual cell values and the average for corresponding groups:

$$MSE = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(X_{ij} - \dot{X}_j\right)^2}{N} \tag{8.1}$$

where X_{ij} = sensor-value for ith grid cell within jth group

$\dot{X}_j$ = sensor-value average for j^{th} group

k = the number of grid cell groups

n_j = the number of grid cells within j^{th} group

N = the total number of non-zero grid cells

The interactive process of grid cell grouping starts with the assumption that all grid cells belong to the group labelled "1" designated as "the rest of the field". Grid cells can be grouped together only when they have at least one common side. This assumption is typically referred to as "rook's rule".

Only nine neighbouring grid cells in a 3x3 configuration can form a new group. The beginning of a new group as well as a merger of a new grid cell to an existing group is accepted when the result produces the lowest MSE. Group enlargement as well as the search for a new group stops when neither action results in a further decrease in MSE. Fig. 8.1 illustrates the flowchart of the algorithm developed.

One minus the ratio of the MSE calculated using equation (8.1) and the initial MSE (considering that k = 1) indicates the fraction of variability accounted for by the grouping and is equivalent to the coefficient of determination (R^2) typically used to quantify the quality of a linear regression model:

$$R^2 = 1 - \frac{MSE}{MSE_{k=1}} \tag{8.2}$$

As shown by Dhawale *et al.* (2012), this algorithm can be successfully used for a single data layer. To achieve multilayer analysis, MSE for each data layer should be minimized and R^2 maximized. This can be realized by multiplying R^2 values. Thus, perfect recognition of spatial variability would mean $R^2 = 1$ (measured values within each group are exactly the same), and $R^2 < 1$ once a fraction of the variability is not accounted for. Therefore, the product of three R^2 (elevation and two depths of EC) will be small if at least one of the three multipliers is relatively low. Since MSE_k=1 is a constant value, the same grid cell grouping result will occur when minimizing the product of three MSE estimates as when maximizing the product of three R^2 values. Quality partitioning of an agricultural field would occur when R^2 for all data layers

would be relatively high with the smallest possible number of identified groups of relatively homogeneous grid cells different from their surroundings.

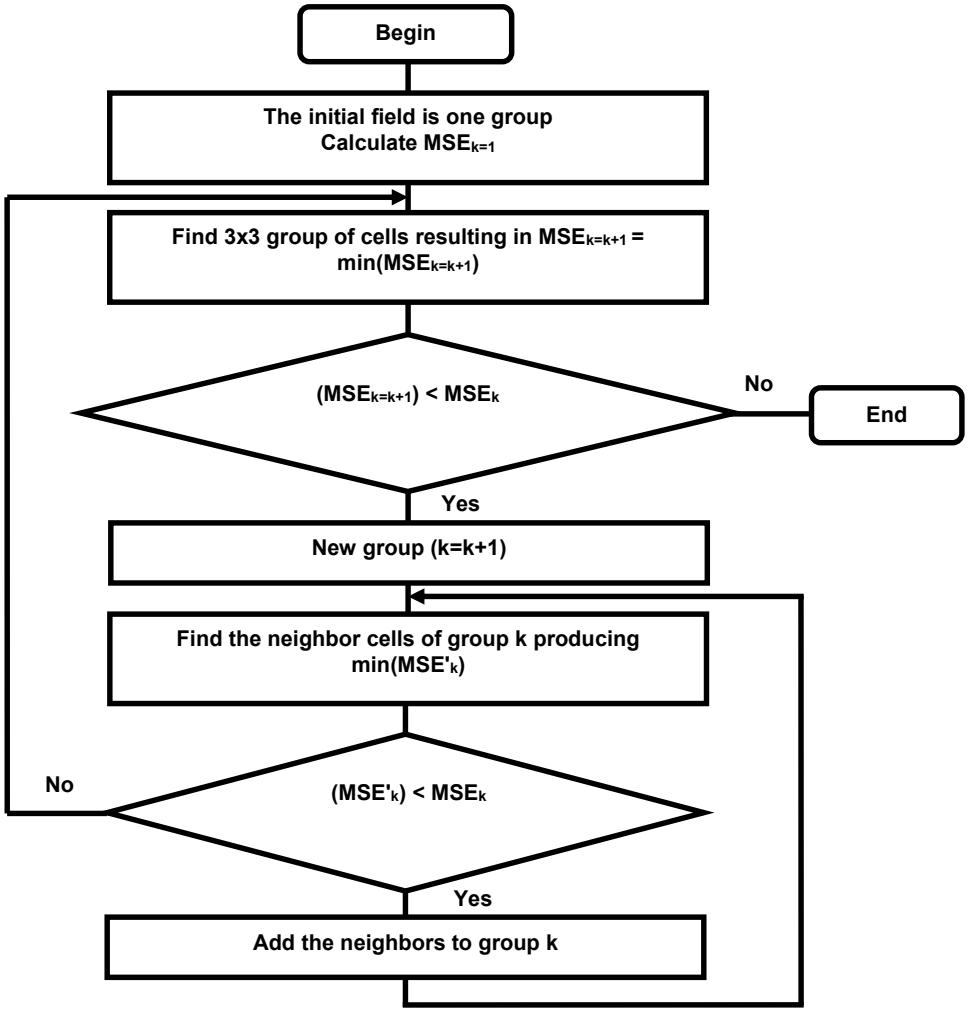


Fig. 8.1 Algorithm flow chart

Since the two ECa measurements frequently correlate, the influence of field elevation in this study was made similar to the influence of ECa by raising the elevation R^2 estimate to the second power:

$$R^2_{Product} = R^2_{ShallowEC_a} \cdot R^2_{DeepEC_a} \cdot \left(R^2_{Elevation}\right)^2 \tag{8.3}$$

Therefore, the algorithm shown in Fig. 1 was implemented to maximize product R^2 instead of the MSE for a single data layer. No formal statistical analysis and comparison with more traditional spatial clustering techniques were performed at this preliminary stage.

8.3 Results and Discussion

Fig. 8.2 illustrates one of the six fields from the initial trial of the developed algorithm. Areas of the field representing low elevation in the east and high elevation in the west were delineated first with two additional groups emerging later. Fig. 8.3 illustrates that R^2 values increased as new groups were formed. Apparently, delineation of groups 2 and 4 were primarily caused by the spatial variability of soil ECa, while groups 3 and 5 emerged predominantly due to differences in field elevation. The algorithm did not locate any new groups of 3 x 3 grid cells that could further increase the R^2 product.

Fig. 8.4 illustrates grid cell grouping for all the fields resulting in a total of 2-8 groups per field. Fig. 8.5 summarises the resulting R^2 values. The products of these values are shown in Fig. 8.6. Fields 2 and 4 revealed only one group of grid cells that could be separated from the rest of the field while Fields 3 and 4 had 6 and 7 groups, respectively. At the same time, the algorithm produced groups with relatively strong three data layer partitioning for Fields 1, 3, 4, and 6. However, sub-division of Fields 2 and 5 were mainly dominated by field elevation, which resulted in relatively low R^2 products. In both cases, soil ECa measurements differed significantly among neighbouring cells, indicating relatively poor spatial structure.

Although the strength of this algorithm is spatial continuity of each group of grid cells, group edges may need smoothing for improved field manageability. Since grid cells poorly associated with their neighbours occur mostly due to field anomalies or erroneous measurements, edge smoothing will always reduce the R^2 product objective function. This suggests that more research to include a comparison of resulting field partitioning with equivalent processing could be conducted using more traditional k-means-type clustering algorithms (Fraisse *et al.*, 2001; Ping and Dobermann, 2003) with suitable pre- and post-processing techniques.

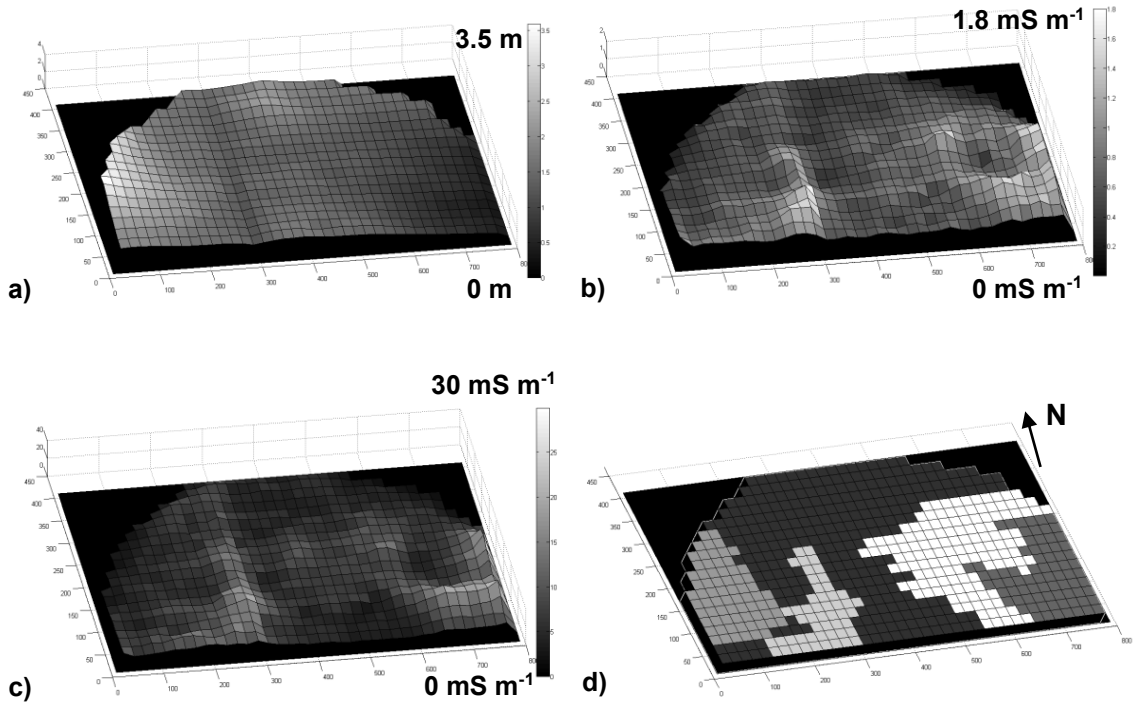


Fig. 8.2 Maps of field elevation (a), shallow ECa (b), deep ECa (c), and delineated groups of grid cells (d) for Field 1

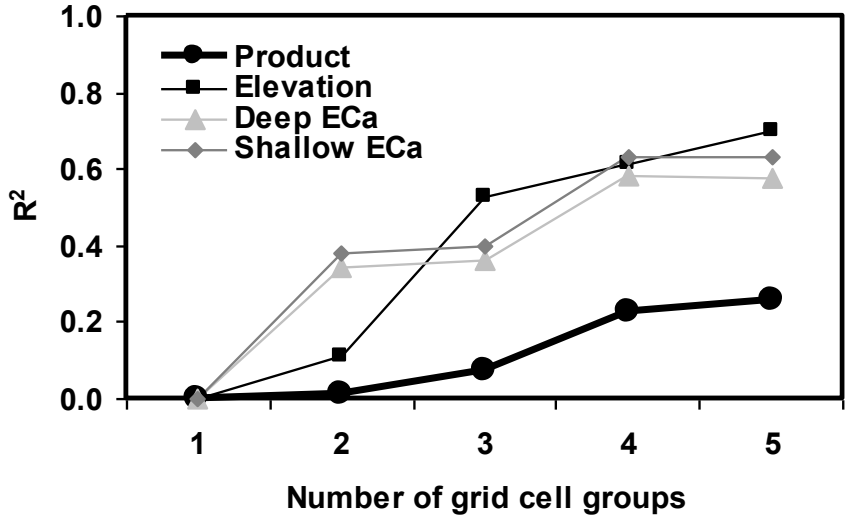


Fig. 8.3 Change in R^2 product with the number of delineated grid cell groupings

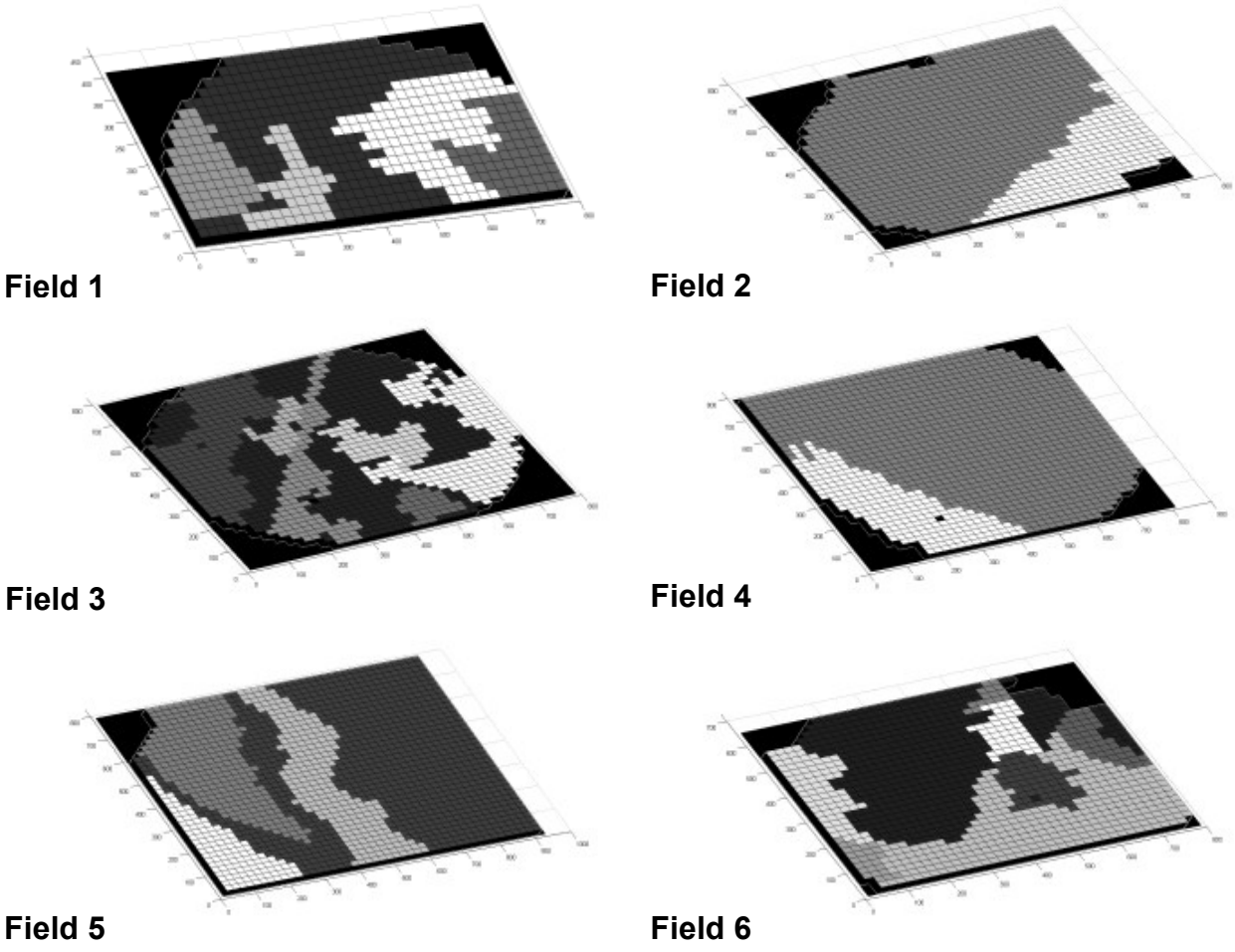


Fig. 8.4. Maps of the partitioned fields

In comparison with previously reported algorithms (Fraisse *et al.*, 2001; Ping and Dobermann, 2003; Fridgen *et al.,* 2004) where the quality of the resulting field partitioned areas heavily depends on the selection of initial centroids and the results are not repeatable, this algorithm produced repeatable field partition areas that were relatively homogenous and spatially contiguous. Spatial contiguity of formed partitioned areas were achieved by restricting the grouping measurements that are not adjacent to each other through Neighbourhood Search Analysis (NSA). In addition, one does not need to provide the number (k) of desired field partitioned areas in advance as the algorithm is capable of finding the optimal number of k on its own. In this way, the algorithm avoids the extra step needed to find the number of *k's* as done in other algorithms by re-running the algorithms for several runs for several numbers of *k's*.

In presented algorithm, the initialization of new groups of fixed size data points and the later of merging its neighbours is all achieved through minimization of the mean squared error (MSE). Therefore, the algorithm is capable of producing the same outcome for any number of repeated runs and does not require cross-validation to select the best among all runs, which is the conventional practice, using traditional clustering.

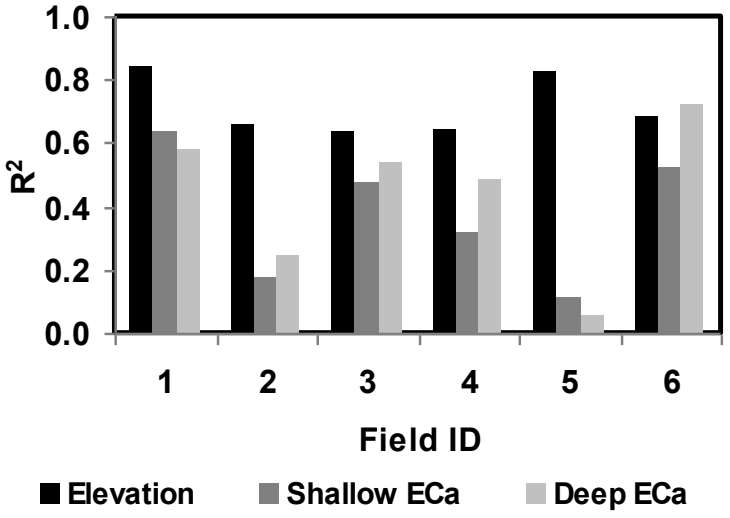


Fig 8.5 R^2 values for the three data layers used to partition the six experimental fields

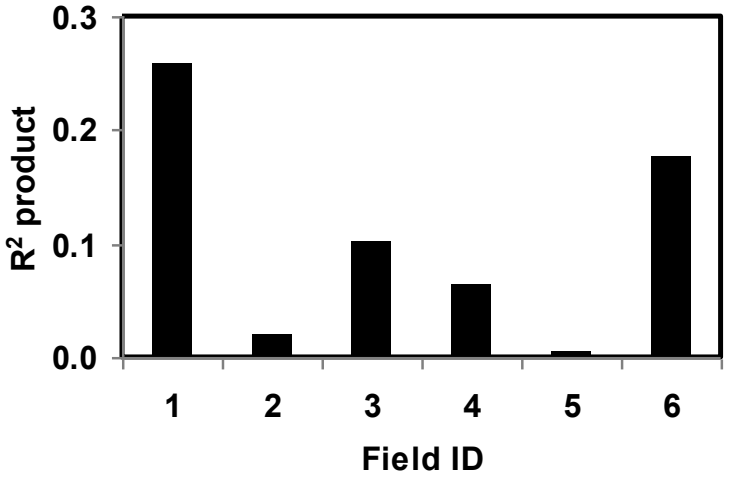


Fig. 8.6 R^2 products for the six fields

In general, the results from the new multi-dimensional spatial data clustering algorithm when evaluated on traditional high-density proximal soil sensing data was capable of delineating five

out of the six agricultural fields into relatively homogeneous areas. The successful outcome of the algorithm indicates its potential for use in optimizing the number of point based measurements within partitioned field areas with OSA, other *in situ* sensing systems, or conventional soil sampling.

8.4 Conclusions

The spatial clustering algorithm developed in this study is based on a neighbourhood search method and seeks to minimize variance inside each group of interpolated grid pixels corresponding to an unlimited number of sensor-based data layers. Preliminary tests of the algorithm using six production fields illustrated algorithm robustness (consistent outcomes) when delineating field areas with different field elevations and soil ECa measurements. Each spatially constrained group of grid cells emerged in response to each unique combination of data values which were relatively constant within each group. Concurrently, more sets of traditional high-density proximal soil sensing data composed of contrasting agricultural fields could be used to test the overall robustness of the algorithm in future.

Chapter 9

SUMMARY AND GENERAL CONCLUSIONS

9.1 Summary

This research provided several groundbreaking advancements for making Proximal Soil Sensing (PSS) technologies more efficient and effective. Firstly, an advanced proximal soil sensing equipment was prototyped to make point based measurements of multi-soil chemical properties in a simultaneous and automated manner. Further, spectroscopic techniques were evaluated in laboratory and in field conditions by comparing the ability of two distinct and advanced hyperspectral instruments for indirect and simultaneous prediction of soil physical and chemical properties. Finally, a smart mutli-dimensional spatial data clustering algorithm was developed to process high-density proximal soil sensing data and then it was evaluated on classical, on-the-go proximal soil sensing data.

A prototype “On-the-spot Soil Analyser" (OSA) was designed, fabricated and evaluated. In its current modification, it used an array of two potentiometer sensors and could simultaneously determine ions such as H^+ and NO_3^- on-the-spot. As OSA is relatively inexpensive and easy to use, it will be attractive to agribusinesses involved in soil fertility management. The technology is expected to provide an opportunity to extend the suite of deployable sensors and thus, allow for sensor fusion algorithms and integrated data acquisition practices.

Four studies were performed to evaluate the usefulness of two distinct field deployment-ready spectrometers to make point based measurements. The first, was a prototype, operating in the mid-infrared (mid-IR) region and the second, was a dual commercial unit, operating in visible and near-infrared (vis-NIR) regions. Several soil properties were determined indirectly by applying, partial least squared regression (PLSR) models, on diffuse reflectance measurements obtained in *in situ* and *ex situ* environments. To assess the precision and accuracy component of

mutli-spectral measurement error, a new technique based on simple linear regression was developed and evaluated on vis-NIR spectral data.

The ability of a commercial vis-NIR unit to accurately predict total soil phosphorus (P) was highly promising (coefficient of determination of 0.85) and both spectrometers resulted in no significantly different results in terms of soil textural predictions, though % sand tended to be better predicted with the prototyped mid-IR, while % clay was better determined using the vis-NIR unit; however, the ability to accurately predict soil organic carbon was relatively poor (prediction error < 0.4 %) for the soils used in the studies. In addition, % clay and % sand were predicted accurately (90% accuracy), using spectral data obtained by the mid-IR instrument in naturally moist soil environments. However, the predictability of soil organic matter was relatively poor and suggests the need for further research. The vis-NIR spectral measurements and model derived predictions of % sand and % clay were also found to be compatible in terms of precision and accuracy under laboratory or field conditions, when soil water content varies substantially.

Finally, a new algorithm based on neighbourhood search analysis which seeks to minimize variance inside each group of interpolated grid pixels corresponding to an unlimited number of sensor-based data layers was developed. Tests of the algorithm using six production fields illustrated the algorithms robustness when delineating field areas with different field elevations and apparent soil electrical conductivity (EC_a) measurements. Each spatially constrained group of grid cells, with the exception of the initial group designated as “the rest of the field”, emerged in response to every unique combination of data values as relatively constant within each group.

9.2 General Conclusions

Proximal soil sensing is an emerging technology resulting in on-the-go mapping of soil characteristics. It provides an alternative to on-the-go mapping techniques where on-the-spot measurements can be made in situations where spatially sporadic test locations are needed or soil coverage does not allow for the continuous engagement of the soil with parts of the sensor system.

The on-the-spot soil analyser presented in this thesis illustrates a new mechanism that enables rapid and reliable soil manipulation to facilitate the necessary interface between the sensors and soil in an automated mode. Furthermore, when compared to existing solutions, this method is applicable in difficult soil conditions and enables sufficient opening of a subsurface soil area to engage an array of proximal soil sensors. This technology was designed to extend the suite of other deployable sensors, and thus, allow for sensor fusion algorithms and integrated data acquisition practices.

Promising results obtained from the mid-infrared spectrometer while predicting % sand and % clay indicates its potential use for the OSA platform; however, further testing is needed to investigate potential optimizations of the soil/sensor interface needed to facilitate the deployment of mid-IR spectroscopy for proximal soil sensing.

The results from the new multi-dimensional spatial data clustering algorithm when evaluated on traditional high-density proximal soil sensing data was capable of delineating several agricultural fields into homogeneous and contiguous groups. The successful outcome suggests further research on its use in optimizing the number of point-based measurements with OSA, other *in situ* sensing systems, or conventional soil sampling.

Overall, it can be concluded that, the six independent studies, showing three novel approaches, conveyed several promising results and together, they illustrate the desired advancements brought to PSS for improving the quality of data required by decision-making algorithms for the Precision Agriculture practitioners. Thus, they demonstrate the potential to improve the economic, environmental and social aspects of the modern farmer.

Chapter 10

CONTRIBUTIONS TO KNOWLEDGE AND SUGGESSIONS FOR FUTURE RESEARCH

10. 1 Contributions to Knowledge

This research describes the following contributions to knowledge that will make Proximal Soil Sensing (PSS) more effective and efficient:

1. An automated On-the-spot Soil Analyser (OSA) has been developed which has a distinct mechanism for the deployment of multiple sensors at a given measurement depth, and in a completely automated mode of operation. This OSA can be used in agricultural and natural resources management.

2. For the first time, a portable, mid-infrared spectrometer has been used in naturally moist soil environments for on-the-spot determination of soil textural components (% sand and % clay). This instrument can be used for proximal soil sensing in both agricultural and industrial applications. Presently, no research was found that reported on similar developments elsewhere.

3. This thesis presents three different studies for evaluating a field deployment-ready vis-NIR instrument (Veris® P4000) in terms of predicting on-the-spot key soil properties, including soil particle size distribution, organic carbon content, soil organic matter and total soil phosphorus content.

4. Based on neighbourhood search analysis, a new algorithm to process traditional multi-dimensional, high density, proximal soil sensing data, has been developed for delineating agricultural fields into spatially homogenous and contiguous groups that

can be used for optimizing the number of point-based measurements, with OSA, other soil sensing systems or conventional soil sampling.

10. 2 Suggestions for Further Research

1. Further development of the OSA platform by adding more sensors in the form of either ion-selective electrodes and soil optical sensors will help in determining several other soil chemical properties, making it a very useful tool between farmers and environmental consultants.

2. Further studies are required on investigating the potential optimizations of the soil/sensor interface needed to facilitate the deployment of portable mid-infrared spectroscopy for proximal soil sensing.

3. More sets of traditional high-density proximal soil sensing data collected between several number of contrasting agricultural sites is required to explore the robustness of the newly developed spatial data clustering algorithm.

REFERENCES

Abdul-Nazeer, K.A., & Sebastian. M.P. 2009. Improving the Accuracy and Efficiency of the k-means Clustering Algorithm. In: Proceedings of the World Congress on Engineering, 23, 1-3.

Abidine, A.Z., Heidman, B.C., Upadhyaya, S.K., & Hills D.J. 2004. Autoguidance system operated at high speed causes almost no tomato damage. California Agriculture, 58, 44-47.

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

Adamchuk, V.I. 2001. Untangling the GPS String. University of Nebraska Co-operative Extension, EC-01-157, University of Nebraska, Lincoln, Nebraska, USA.

Adamchuk, V.I., Perk, R.L., & Schepers, J.S. 2003. Application of Remote Sensing in Site-specific Management. University of Nebraska, Cooperative Extension. Institute of Agriculture and Natural Resources.

Adamchuk, V.I., Hummel, J.W., Morgan, M.T., & Upadhyaya, S.K. 2004. On-the-go soil sensors for precision agriculture. Computer and Electronics in Agriculture, 44, 71-91.

Adamchuk, V.I., & Mulliken, J. 2005. Site-Specific Management of Soil pH (FAQ). University of Nebraska-Lincoln, Extension EC05-705.

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

Adamchuk, V.I., Morgan, M.T., & Brouder, S.M. 2006. Development of an On-the-go Soil pH Mapping Method: Analysis of Measurement Variability. Applied Engineering in Agriculture, 22, 223-344. Adamchuk, V.I., Morgan, M.T., & Lowenberg-DeBoer, J.M. 2004. A model for agro-economic analysis of soil pH mapping. Precision Agriculture, 5, 109-127.

Adamchuk, V.I., & Lund, E.D. 2008. On-the-go mapping of soil pH using antimony electrodes. Paper No. 08-3995. St. Joseph, Michigan: ASABE.

Adamchuk, V.I., Viscarra Rossel, R.A., Marx, D.B., & Samal, A.K. 2008. Enhancement of On-the-go Soil Sensor Data Using Guided Sampling. In: Proceedings of the Ninth

International Conference on Precision Agriculture (ed., R. Kholsa), Denver, Colorado, CD Publication, 13pp. Colorado State University, Fort Collins.

Adamchuk, V.I., Viscarra Rossel, R.A., Marx, D.B., & Samal, A.K. 2011. Using targeted sampling to process multivariate soil sensing data. Geoderma, 163, 63-73.

Adamchuk, V.I., Viscarra Rossel, R.A., Sudduth, K.A., & Schulze Lammers, P. 2011. Sensor Fusion for Precision Agriculture. In: Sensor Fusion - Foundation and Applications, Chapter 2, pp. 27-40, ed. C. Thomas, InTech, Rijeka, Croatia.

Adamchuk, V., Dhawale, N., & Rene-Laforest, F. 2014. Development of On-the-spot Soil Analyser for Measuring Soil Chemical Properties. In: proceedings of 12th International Conference on Precision Agriculture, July 20-23, 2014, Sacramento, California, USA.

Adsett, J.F., & Zoerb, G.C. 1991. Automated Field Monitoring of Soil Nitrate Levels. In: Automated Agriculture for the 21st Century, Proceedings of the 1991 Symposium, pp. 326-335. American Society of Agricultural Engineers, St. Joseph.

Akaike, H. 1974. A New Look at the Statistical Model Identification. IEEE Transactions on Automatic Control, 19, 716-23.

Anderson-Cook, C.M., Alley, M.M., Roygard, J.K., Khosla, R., Noble, R.B., & Doolittle, J.A. 2002. Differentiating soil types using electromagnetic conductivity and crop yield maps. Soil Science Society of America Journal, 66, 1562-1570.

Artigas, J., Beltran, A., Jimenez, C., Baldi, A., Mas, R., Dominguez, C., & Alonso, J. 2001. Application of ion sensitive field-effect transistor based sensors to soil analysis. Computer and Electronics in Agriculture, 31, 281-293.

Ben-Dor, E., & Banin, A. 1994. Visible and near-infrared (0.4-1.1 um) analysis of arid and semiarid soils. Remote Sens. Environ. 48, 261-74.

Ben-Dor, E., & Banin, A. 1995. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. Soil Sci. Soc. Am. J. 59, 364-72.

Birrell, S.J., & Hummel, J.W. 2001. Real-time multi ISFET/FIA soil analysis system with automatic sample extraction. Computer and Electronics in Agriculture, 32, 45-67.

Bishop, J. L., Pieters, C.M., & Edwards, J.O. 1994. Infrared spectroscopic analyses on the nature of water in montmorillonite. Clays and Clay Minerals, 42, 702-16.

Blackmore, S. 2000. The interpretation of trends from multiple yield maps. Computers and electronics in agriculture, 26, 37-51.

Blackmore, S. 2002. Developing the Principles of Precision Farming. In Proceedings of the International Conference on Agropoles and Agro-Industrial Technological Parks (Agrotech 99).

Blackmore, S., Have, H., & Fountas, S. 2002. A specification of behavioural requirements for an autonomous tractor. Automation technology for off-road equipment, ed., Q. Zhang, ASAE Publication, 33-42.

Blackmore, S. 2003. THE ROLE OF YIELD MAPS IN PRECISION FARMING. A Doctoral dissertation, Cranfield University, Cranfield, Bedfordshire, UK.

Blackmore, S., Stout, B., Wang, M., & Runov, B. 2005. Robotic Agriculture-the Future of Agricultural Mechanisation. In: Procedings of 5th European Conference on Precision Agriculture (ECPA), Upsala, Sweden.

Bogrekci, I., & Lee, W.S. 2005. Spectral measurement of common soil phosphates. Transactions of the ASAE. 48, 2371-2378.

Carpenter, S.R. 2008. Phosphorus control is critical to mitigating eutrophication. In Proceedings of the National Academy of Sciences of the United States of America, 11039-11040. Washington, DC. August.

Canadian Council of Ministers for the Environment. 2007. Canadian Guidance Framework for the Management of Nutrients in Near-shore Marine Systems. Available at http://www.ccme.ca/assets/pdf/ntrnts_gdnc_frmwrk_ssd_1387.pdf (seen May. 10, 2012)

Carefoot, J.P., & Whalen, J.K. 2003. Phosphorus concentration in subsurface water as influenced by cropping systems and fertilizer sources. Canadian Journal of Soil Science. 83, 203-212.

Chambers, P.A., Allard, M., Walker, S.L., Marsalek, J., Lawrence, J., Servos, M., Busnarda, J., Munger, K.S., Adare, K., Jefferson, C., Kent, R.A., & Wong, M.P. 1997. Impacts of Municipal wastewater effluents on Canadian waters: a review. Water Qual. Res. J. 32, 659-713.

Chang, C.W., Laird, D.A., Mausbach, M.J., & Hurburgh, C.R. 2001. Near infrared reflectance spectroscopy-principal components regression analysis of soil properties. Soil Science Society of American Journal, 65, 480-490.

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

Chodak, M., Ludwig, B., Khanna, P., & Beese, F. 2002. Use of near infrared spectroscopy to determine biological and chemical characteristics of organic layers under spruce and beech stands. J. Plant Nutr. Soil Sci. 165, 27-3.

Christy, C., Collings, K., Drummond, P., & Lund, E. 2004. A mobile sensor platform for measurement of soil pH and buffering. ASAE Paper No. 041042. St. Joseph, Mich.: ASAE.

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

Clark, R.N., King, T.V.V., Klejwa, M., Swayze, G.A., & Vergo, N. 1990. High spectral resolution reflectance spectroscopy of minerals. J. Geophys. Res. 95, 12653-2680.

Clark, R.N. 1999. Spectroscopy of Rocks and Minerals and Principles of Spectroscopy. In: Remote Sensing for the Earth Sciences. (ed., A.N. Rencz), pp. 3-8. John Wiley & Sons, Chichester, UK.

Coates, J.P. 2000. New microspectrometers: building on the principle that is beautiful. Spectroscopy. 15, 21-27.

Condron, L.M., Turner, B.L., & Cade-Menun, B.J. 2005. Chemistry and Dynamics of Soil Organic Phosphorus. In Phosphorus: Agriculture and the Environment, eds. J.T. Sims and A.N. Sharpley, 87-121, Agronomy Monograph No. 46. Agronomy Society of America, Madison, WI. USA.

Cordell, D., Drangert, J-O., White, S. 2009. The story of phosphorus: Global food security and food for thought. Global Environmental Change 19, 292-305.

Corwin, D.L., & Lesch, S.M. 2005. Apparent soil electrical conductivity measurements in COMPAG. 46, 11-43.

Couteaux, M.M., Berg, B., & Rovira, P. 2003. Near infrared reflectance spectroscopy for determination of organic matter fractions including microbial biomass in coniferous forest soils. J. Soil Biol. Biochem. 35, 1587-600.

Cozzolino, D., & Moron, A. 2003. The potential of near-infrared reflectance spectroscopy to analyse soil chemical and physical characteristics. Journal of Agricultural Sciences, 140, 65-71.

Dalal, R.C., & Henry, R.J. 1986. Simultaneous determination of moisture, organic carbon, and total nitrogen by near infrared reflectance spectrophotometry. Soil Sci. Soc. Am. J. 50, 120-23.

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

de Gruijter, J.J., McBratney, A.B., & Taylor, J. 2010. Sampling for High Resolution Soil Mapping. In: Proximal Soil Sensing (eds., R.A. Viscarra Rossel, A.B. McBratney, and B. Minasny), pp. 3-14. Springer-Verlag, New York.

de Jong, S., & Kiers, H. 1992. Principal covariates regression. J. Chemometrics and Intelligent Laboratory Systems, 14, 155-64.

Deng, X., Wang, Y., & Peng, H. 2003. The Clustering of High Resolution Remote Sensing Imagery. Journal of Electronics & Information Technology, 25, 1073-1080.

Dhawale, N., Adamchuk, V.I., Prasher, S.O., & Dutilleul, P.R.L. 2012. Spatial data clustering using neighbourhood analysis. Paper No. 121337939. St. Joseph, Michigan: ASABE.

Dhawale, N., Adamchuk, V., Prasher, S., Ismail, A., & Viscarra Rossel, R.A. 2013a. Analysis of the Repeatability of Soil Spectral Data Obtained Using Different Measurement Techniques. In: Proceedings of the 3rd Global Workshop on Proximal Soil Sensing, Potsdam, Germany, 26-29 May 2013, (eds., R. Gebbers, E. Luck, & J. Ruhlmann), pp. 161-165. Potsdam, Germany: ATB Leibniz-Institut fur Agratechnik Potsdam-Bornim.

Dhawale, N.M., Adamchuk, V.I., Viscarra Rossel, R.A., Prasher, S.O., Whalen, J.K., & Ismail, A.A. 2013b. Predicting Extractable Soil Phosphorus using vis/NIR Hyperspectral Soil Reflectance Measurements. Paper No. 13-047. Orleans, Ontario, Canada: CSBE.

Dhawale, N.M., Adamchuk, V.I., Prasher, S.O., Dutilleul, P.R.L, & Ferguson, R.B. 2014a. Spatially constrained geospatial data clustering for multilayer sensor-based measurements. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 40, 187-190.

Dhawale, N.M., Adamchuk, V.I., Prasher, S.O., Viscarra Rossel, R.A., Ismail, A.A., Whalen, J.K., & Louargant, M. 2014b. Comparing vis/NIR/mid-IR hyperspectrometry for measuring soil physical properties. Paper No. 141909453. St. Joseph, Michigan: ASABE.

Dobermann, A., Blackmore, S., Cook, S.E., & Adamchuk, V.I. 2004. Precision Farming: Challenges and Future Directions. In: Proceedings of the 4th International Crop Science Congress, 26 Sep - 1 Oct 2004, Brisbane, Queensland, Australia.

Doerge, T., Kitchen, N.R., & Lund, E.D. 2002. Soil Electrical Conductivity Mapping, Site-specific Management Guidelines. Available at - www.ppi-far.org (Seen Sep.10. 2012).

Easterly D.R., Adamchuk, V.I., Kocher, M.F., & Hoy, R.M. 2010. Using a vision sensor system for performance testing of satellite-based tractor auto-guidance. Computers and Electronics in Agriculture 72, 107-118.

Enright, P., Begg, C., & Dodds, G. 2009. Targeted P Applications Based upon Grid Soil Sampling: final report. In: Livestock Environmental Intiative. Brace Centre for Water Resources Management, McGill University. Canada.

Etzion, Y., Linker, R., Cogan, U., & Shmulevich, I. 2004. Determination of protein concentration in raw milk by mid-infrared Fourier transform infrared/attenuated total refectance spectroscopy. Journal of Dairy Science, 87, 2779-2788.

Fraisse, C.W., Sudduth, K.A., & Kitchen, N.R. 2001. Delineation of site-specific management zones by unsupervised classification. Trans. ASAE. 44, 155-166.

Fridgen, J.J., Kitchen, N.R., Sudduth, K.A., Drummond, S.T., Wiebold, W.J., & Fraisse, C.W. 2004. Management Zone Analyst (MZA): software for subfield management zone delineation. Agron. J. 96, 100-108.

Gandonou, J-M. 2005. ESSAYS ON PRECISION AGRICULTURE TECHNOLOGY ADOPTION AND RISK MANAGEMENT. A Phd Thesis, University of Kentucky. Lexington, KY. 40506. USA.

Gebbers, R. Herbst, R., & Wenkel, K-O. 2009. Sensitivity Analysis of Soil Nutrient Mapping. In: EFITA Vonference ´09 (eds A. Bregt, S. Wolfert, J. E. Wien & C. Lokhorst), pp. 513-519. Wageningen Academic Publishers, Wageningen, The Netherlands.

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

Gee, G.W., & Bauder, J.W. 1986. Particle-size Analysis. In: Methods of soil analysis, Part1-Physical and mineralogical methods. pp. 383-411. American Society of Agronomy, Madison, Wisconsin, USA.

Geladi, P., & Kowalski, B. 1986. Partial least squares regression: A tutorial. Journal of Analytica Chimica Acta, 185, 1-17.

Geladi, P., Macdougall, D., & Martens, H. 1985. Linearization and scatter-correction for near-infrared reflectance spectra of meat. J. Appl. Spectrosc. 39, 491-00.

Griffin, T. 2009. Whole-farm benefits of GPS-enabled navigation technologies. In: Proceedings of. 2009 ASABE Annual International Meeting. Paper No. 095983.St. Joseph, Michigan: ASABE.

Grisso, R., Alley, M., Wysor, W.G., Holshouser, D., & Thomason, W. 2009. Precision Farming Tools: Soil Electrical Conductivity. Virginia Cooperative Extension, Virginia Tech, and Virginia State University. Petersburg, Virginia, USA.

Grisso, R., Alley, M., Thomason, W., Holshouser, D., & Roberson, G.T. 2011. Precision Farming Tools: Variable-rate Application. Virginia Cooperative Extension, College of Agriculture and Life Sciences, Virginia Polytechnic Institute and State University. Blacksburg, Virginia. USA.

Haapala, H.E.S., Pesonen, L., & Nurkka, P. 2006. Usability as a Challenge in Precision Agriculture-case Study: an ISOBUS VT. Agricultural Engineering International: the CIGR Ejournal. vol. 8, Manuscript MES 05 001.

He, Y., Song, H., Pereira, A.G., & Gómez, A.H. 2005. A New Approach to Predict N, P, K and OM Content in a Loamy Mixed Soil by Using Near Infrared Reflectance Spectroscopy. In: Advances in Intelligent Computing. Lecture Notes in Computer Science, vol. 3644. (eds., D.S. Huang, X.-P. Zhang, G.-B. Huang) pp. 859-867. Springer-Berlin, Heidelberg, Germany.

Henderson, T.L., Baumgardner, M.F., Franzmeier, D.P., Stott, D. E., & Coster, D.C. 1992. High dimensional reflectance analysis of soil organic-matter. Soil Sci. Soc. Am. J. 56, 865-72.

Hodge, A.M., & Sudduth, K.A. 2012. Comparison of two spectrometers for profile soil carbon sensing. Paper No. 121338240. St. Joseph, Michigan, USA: ASABE.

Hoeft, R.G., Peck, T.R., & Boone, L.V. 1996, Soil Testing and Fertility. p. 70-101. In: Illinois Agronomy Handbook 1995-1996, Circular 1333, Cooperative Extension Service, College of Agriculture, University of Illinois at Urbana-Champaign, USA.

Holland, J.K., Erickson, B., & Widmar, D.A. 2013. In: Precision Agricultural Services Dealership Survey Results. Dept. of Agricultural Economics, Purdue University, West Lafayette, Indiana. USA.

Holman, J. P. 2001. Experimental Methods for Engineers, 7th edition. McGraw Hill, New York, New York, USA.

Holman, W.I.M. 1943. A new technique for the determination of phosphorus by the molybdenum blue method. Biochemical Journal, 37,256-259.

Hooda, P.S., Edwards, A.C., Anderson, H.A., & Miller, A. 2000. A review of water quality concerns in livestock farming areas. Science of the Total Environment, 250, 143-167.

Hummel, J.W., Gaultney, L.D., Sudduth, K.A. 1996. Soil property sensing for site-specific crop management. Computers and Electronics in Agriculture, 14, 121-136.

Ima, C.S., & Mann, D.D. 2004. Ergonomic concerns with light bar guidance displays. Agricultural Safety and Health, 10, 91-102.

Ji, W.J., Viscarra Rossel, R.A. & Shi, Z. 2015. Accounting for the effects of water and the environment on proximally sensed vis-NIR spectra and their calibrations. European Journal of Soil Science (in press).

Jiao, Y., Whalen, J.K., & Hendershot, W.H. 2006. No-tillage and manure applications increase aggregation and improve nutrient retention in a sandy-loam soil. Geoderma, 134, 24-33

Janik, L.J., & Skjemstad, J.O. 1995. Characterisation and analysis of soils using mid-infrared partial least squares: II. Correlation with some laboratory data. Australian Journal of Soil Research, 33, 637-650.

Janik, L.J., Merry, R.H., & Skjemstad, J.O. 1998. Can mid infra-red diffuse reflectance analysis replace soil extractions? Australian Journal of Experimental Agriculture, 38, 681-696.

Janik, L.J., Skjemstad, J.O., & Raven, M.D. 1995. Characterization and analysis of soils using mid infrared partial least-squares. I. Correlations with XRF-determined major-element composition. Australian Journal of Soil Research, 33, 621-636.

Jasinski, S. M. 2011. Phosphate Rock, Statistics and Information. In US Geological Survey. Mineral Commodity Summaries.

Kerby, A., Marx, D., Samal, A., & Adamchuk, V. 2007. Spatial Slustering Using the Likelihood Function. pp. 637-642. In: ICDM Workshops, Seventh IEEE International Conference.

Khosla, R. 2008. The 9th International Conference on Precision Agriculture opening ceremony presentation. July 20-23rd.

Kim, H.J., Sudduth, K.A., Hummel, J.W., & Birrell, S.J. 2007. Evaluation of phosphate ion-selective membranes and cobalt-based electrodes for soil nutrient sensing. Transactions of the ASABE. 50, 415-425.

Kim, H.J., Sudduth, K.A., & Hummel, J.W. 2009. Soil macronutrient sensing for precision agriculture. J. Environ. Monit. 11, 1810-1824.

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

Kodaira, M. & Shibusawa, K. 2013. Using a mobile real-time soil visible-near infrared sensor for high resolution soil property mapping. Geoderma, 199, 64-79.

Krischenko, V.P., Samokhvalov, S.G., Fomina, L.G., & Novikova, G.A. 1991. Use of Infrared Spectroscopy for the Determination of Some Properties of Soil. p. 239-249. In: Interagrotech, Moscow, USSR.

Krishnan, P., Alexander, D.J., Butler, B., & Hummel, J.W. 1980. Reflectance technique for predicting soil organic matter. Soil Sci. Soc. Am. J. 44, 1282-285.

Li, Z., & Wang, X. 2010. Spatial Clustering Algorithm Based on Hierarchical-Partition Tree. International Journal of Digital Content Technology and its Applications, 4, 26-35.

Linker, R. 2008. Soil classification via mid-infrared spectroscopy. Computer and Computing Technologies in Agriculture, 2, 1137-1146.

Linker R. 2011. Application of FTIR Spectroscopy to Agricultural Soils Analysis, Chapter 18. Fourier Transforms - New Analytical Approaches and FTIR Strategies, (eds., Prof. Goran Nikolic), pp. 387-404. InTech, Rijeka, Croatia.

Lobsey, C., Viscarra Rossel, R., & McBratney, A. 2010. An Automated System for Rapid In-Field Soil Nutrient Testing. In: Proceedings of 19th World Congress of Soil Science, Soil Solutions for a Changing World, 1-6 August 2010, Brisbane, Queensland, Australia.

Madari, B.E., Reeves III, J.B., Machado, P.L.O.A., Guimaraes, C.M., Torres, E., & McCarty, G.W. 2006. Mid and near-infrared spectroscopic assessment of soil compositional parameters and structural indices in two Ferralsols. Geoderma, 136, 245-9.

Maleki, M.R., Van Holm, L., Ramon, H., Merckx, R., De Baerdemaeker, J., & Mouazen, A.M. 2006. Phosphorus sensing for fresh soils using visible and near infrared spectroscopy. Biosystems Engineering, 95, 425-436.

Malley, D.F., Martin, P.D., & Ben-Dor, E. 2004. Application in Analysis of Soils. In: Near-infrared spectroscopy in agriculture, Agronomy 44. (eds., C.A., Roberts, J. Workman (Jr.), J.B. Reeves III), pp. 729-84. ASA, CSSA, SSSA, Madison, WI, USA.

Martens, H., & Næs, T. 1989. Multivariate Calibration. John Wiley & Sons, Chichester, New York, New York, USA.

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

Masserschmidt, I., Cuelbas, C.J., Poppi, R.J., De Andrade, J.C., De Abreu, C.A., & Davanzo, C.U. 1999. Determination of organic matter in soils by FTIR/diffuse reflectance and multivariate calibration. Journal of Chemometrics, 13, 265-273.

McBratney, A.B., Mendonca Santos, M.L., & Minasny, B., 2003. On digital soil mapping. Geoderma, 117, 3-52.

McBratney, A., Whelan, B., Ancev, T., & Bouma, J. 2005. Future Directions of Prescision Agricuulture. Precision Agriculture, 6, 7-23.

McCarty, G.W., Reeves III, J.B., Reeves, V.B., Follett, R.F., & Kimble, J.M. 2002. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurements. Soil Science Society of America Journal, 66, 640-646.

Miller, R.O. 2006. A western evaluation of soil testing laboratory performance. Better Crops, 90, 26-29.

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

Murphy, J., & Riley, J.P. 1962. A modified simple solution method for the determination of phosphate in natural waters. Analytica Chimica Acta. 27: 31-36.

Nduwamungu, C., Ziadi, N., Parent, L.-É., Tremblay, G.F., & Thuriès, L. 2009. Opportunities for, and limitations of, near infrared reflectance spectroscopy applications in soil analysis: A review. Can. J. Soil Sci. 89, 531-41.

Nemenyi, M., Mesterhazi, P.A., Pecze, Zs., & Stepan, Zs. 2003. The role of GIS and GPS in precision farming. Comput. Electron. Agric. 40, 45-55.

Nguyen T.T., Janik L.J., & Raupach, M. 1991. Diffuse reflectance infrared Fourier transform (DRIFT) spectroscopy in soils studies, Australian Journal of Soil Research, 29, 49-67.

Nyiraneza, J., N'Dayegamiye, A., Gasser, M.O., Giroux, M., Grenier, M., Landry, C., & Guertin, S. 2010. Soil and crop parameters related to corn nitrogen response in Eastern Canada. Agron. J. 102, 1478 1490.

Pecze, Zs. 2001. CASE MAPS OF THE PRECISION (SITE-SPECIFIC) FARMING. Thesis of Ph.D. Dissertation. University of West-Hungary, Sopron, Mosonmagyarovar. Hungary.

Pierce, F.J., & Nowak, P. 1999. Aspects of Precision Agriculture, p. 1-85. In: Advances in Agronomy, vol. 67, (ed., Donald L. Sparks), New York: Academic Press.

Piikki, K., Wetterlind, J., Söderström, M., & Stenberg, B. 2014. Three-dimensional digital soil mapping of agricultural fields by integration of multiple proximal sensor data obtained from different sensing methods. Precision Agriculture, 16, 29-45

Ping, J.L., & Dobermann, A. 2003. Creating spatially contiguous yield classes for site-specific management. Agronomy Journal, 95, 121-1131.

Prodanov, D.P., Nagelkerke, N., & Marani, E. 2007. Spatial clustering analysis in neuroanatomy: Applications of different approaches to motor nerve fiber distribution. Journal of Neuroscience Methods, 160, 93 108.

Reeves III, J.B. 2010. Near-versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: Where are we and what needs to be done? Geoderma, 158, 3-14.

Reeves III, J.B., McCarty, G.W., & Hively, W.D. 2010. Mid-Versus Near-Infrared Spectroscopy for On-Site Analysis of Soil. Chapter 11. Proximal Soil Sensing: Progress in Soil Sci. 1 (eds., R.A. ViscarraRossel, A. B. McBratney & B. Minasny), pp. 133-142, Springer-Verlag, New York, USA.

Robert, P., Rust, R., & Larson, W. 1994. Site-Specific Management for Agricultural Systems. In: Proceedings of the 2nd International Conference on Precision Agriculture, Madison, WI. ASA/CSSA/SSSA.

Roberts, R.K., English, B.C., & Mahajanashetti, S.B. 2001. Environmental and Economic Effects of Spatial Variability and Weather. In: Proceedings of the 3rd European Conference on Precision Agriculture. (eds., S. Blackmore & G. Grenier), AGRO, Montpellier, France, 545-550.

Savitzky, A., & Golay, M. J. E. 1964. Smoothing and differentiation of data by simplified least squares procedures. J.Anal. Chem. 36, 1627-1639.

Saxton, K.E., Rawls, W.J., Romberger, J.S., & Papendick, R.I. 1986. Estimating generalized soil water characteristics from texture. Trans. ASAE. 50, 1031-1035.

Schmidt, O., & Hughes-Games, G. 2010. Factsheet - No. 7. Potassium Considerations for Nutrient Management Nutrient Management. British Columbia Ministry of Agriculture and Lands. Available at http://www.agf.gov.bc.ca/resmgmt/publist/600Series/631500-5_Potassium_Considerations.pdf (seen Sept. 10, 2012).

Schulte E.E., & Hopkins, B.G. 1996. Estimation of Organic Matter by Weight Loss-On-Ignition. In: Soil Organic Matter: Analysis and Interpretation. (eds., F.R. Magdoff, et al.), pp. 21-1. Soil Science Society of America, Special Publication No. 46, SSSA, Madison, Wisconsin, USA.

Sethuramasamyraja, B. 2006. DEVELOPMENT OF ON-THE-GO SOIL SENING TECHNOLOGY FOR MAPPING SOIL pH, POTASSIUM AND NITRATE CONTENTS. A Phd Thesis, University of Nebraska. Lincoln, NE., USA.

Shatar, T.M., & McBratney, A. 2001. Subdividing a Field into Contiguous Management Zones Using a K-Zones Algorithm. pp. 115-120. In: Proceedings of the 3rd European Conference on Precision Agriculture, (eds., G. Grenier and S. Blackmore), Agro-Montpellier ENSAM. France.

Shepherd, K., & Walsh, M. 2002. Development of reflectance spectral libraries for characterization of soil properties. Soil Science Society of America Journal, 66, 988-998.

Shibusawa, S. 2006. Soil Sensors for Precision Agriculture. In: Handbook of Precision Agriculture. Principles and Applications, (ed., A. Srinivasan), pp. 57-90. Food Products Press, New York, New York, USA.

Shockley, J.M. 2010. WHOLE FARM MODELLING OF PRECISION AGRICULTURE TECHNOLOGIES. A Phd Thesis, Paper 105. University of Kentucky. Lexington, KY. 40506. USA.

Sibley, K.J., Astatkie, T., Brewster, G., Struik, P.C., Adsett, J.F., & Pruski, K. 2009. Field-scale validation of an automated soil nitrate extraction and measurement system. Precision Agriculture, 10, 162-174.

Soil Survey Staff. 2004. Soil Survey Laboratory Methods Manual (version 4.0). USDA-NRCS, US Government Print Off, Washington, DC, USA.

Sparks, D.L. 2003. Environmental Soil Chemistry 2nd ed. New York: Elsevier Science.

St. Luce, M., Ziadi, N., Nyiranezab, J., Tremblayb, G., Zebarthc, B., Whalen, J., & Laterrière, M. 2012. Near-Infrared reflectance spectroscopy prediction of soil nitrogen supply in humid temperate regions of Canada. Soil Science Society of American Journal,76(4): 1454-1461.

Stenberg B., Jenson, L.S., Nordkvist, E., Breland, T.A, Pederson, A., GuOmundsson, J., Bruun, S., Salo, T., Palmason, F., Henriksen, T.M., & Korsaeth, A. 2004. Near infrared reflectance spectroscopy for quantification of crop residue, green manure and catch crop C and N fractions governing decomposition dynamics in soil. Journal of Near Infrared Spectroscopy, 12, 331-346.

Stenberg, B. 2010. Effects of soil sample pre-treatment and standardized rewetting as interacted with sand classes on vis-NIR predictions of clay and soil organic carbon. Geoderma, 158, 15-22.

Sudduth, K.A., & Hummel, J.W. 1993a. Portable, near-infrared spectrophotometer for rapid soil analysis. Transaction -American Society of Agricultural Engineers, 36, 185-194.

Sudduth, K.A., & Hummel, J.W. 1993b. Soil organic matter, CEC, and moisture sensing with a portable NIR spectrophotometer. Transactions-American Society of Agricultural Engineers, 36, 1571-1582.

Sudduth, K.A., Drummond, S.T., & Kitchen, N.R. 2001. Accuracy issues in electromagnetic induction sensing of soil electrical conductivity for precision agriculture. Comput. Electron. Agric. 31, 239-264.

Sudduth, K.A., Hummel, J.W., & Birrell, S.J. 1997. Sensors for site-specific management. In: The State of Site-Specific Management for Agriculture, (eds., F.T. Pierce and E.J. Sadler), pp. 183-210, ASA-CSSA-SSSA, Madison, Wisconsin, USA

Thomasson, J.A., Sui, R., Cox, M.S., & Al-Rajehy, A. 2001. Soil reflectance sensing for determining soil properties in precision agriculture. Transactions of the ASAE. 44, 1445-1453.

Udelhoven, T., Emmerling, C., & Jarmer, T. 2003. Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least-square regression: A feasibility study. J. Plant Soil. 251, 319-29.

Viscarra Rossel, R.A., & McBratney, A.B. 1997. Preliminary Experiments Towards the Evaluation of a Suitable Soil Sensor for Continuous 'On-the-Go' Field pH Measurements.

In: Precision Agriculture '97, Vol. 2, Technology, IT and Management, (ed., J.V. Stafford), pp. 493-502. Bios Scientific Publishers Ltd, Oxford. UK.

Viscarra Rossel, R.A., & McBratney, A.B. 1998. Laboratory evaluation of a proximal sensing technique for simultaneous measurement of clay and water content. Geoderma, 85, 19 39.

Viscarra Rossel R.A. 2001. DEVELOPMENT OF A PROXIMAL SOIL SENING SYSTEM FOR THE CONTINOUS MANAGEMENT OF ACID. Unpublished PhD Thesis. The University of Sydney, New South Wales, Australia.

Viscarra Rossel, R.A., Gilbertsson, M., Thyle´n, L., Hansen, O., McVey, S., & McBratney, A.B. 2005. Field Measurements of Soil pH and Lime Requirement Using an On-the-Go Soil pH and Lime Requirement Measurement System. In: Precision Agriculture '05 (ed., J.V. Stafford), Wageninen Academic Publishers, Wageninen, The Netherlands.

Viscarra Rossel, R.A., McGlynn, R.N., & McBratney, A.B. 2006a. Determining the composition of mineral-organic mixes using UV-vis-NIR diffuse reflectance spectroscopy. Geoderma, 137, 70-82.

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

Viscarra Rossel, R.A. 2008. ParLeS: Software for chemometric analysis of spectroscopic data. Chemometrics and Intelligent Laboratory Systems, 90:72-83.

Viscarra Rossel, R.A., Jeon, Y.S., Odeh, I.O.A., & McBratney, A.B. 2008. Using a legacy soil sample to develop a mid-IR spectral library. Australian journal of Soil Research, 46, 1-16.

Viscarra Rossel, R.A. 2009. The Soil Spectroscopy Group and the development of a global soil spectral library. NIR News, 20, 17-8.

Viscarra Rossel, R.A., McBratney, A.B., & Minasny, B., (eds.). 2010. Proximal Soil Sensing, Progress in Soil Science Series, Springer-Verlag, New York. New York. USA.

Viscarra Rossel, R.A., Adamchuk, V.I., Sudduth, K.A., McKenzie, N.J. & Lobsey, C. 2011. Proximal Soil Sensing: An Effective Approach for Soil Measurements in Space and Time, In: Chapter 5. Advances in Agronomy, vol. 113 (ed., D.L Spark), pp. 237-283. Academic Press, Burlington, Massachusetts, USA.

Volpe, R., Balaram, J., Ohm, T., & Ivlev, R. 1996. Rocky 7: a next generation Mars rover prototype. Advanced Robotics, 11, 341-358.

Wäldchen, J., Schning, I., Mund, M., Schrupf, M., Bock, S., Herold, N. et al. 2012. Estimation of clay content from easily measurable water content of air-dried soil. Journal of Plant Nutrition, 175, 367-376.

Watson, M. E., & Brown, J. R. 1998. pH and Lime Requirement. In: Recommended Chemical Soil Test Procedures for the North Central Region, ed. J. R. Brown, 13-16. North Central Regional Publication No. 221. Columbia, Mo.: Missouri Agricultural Experiment Station.

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

Wetterlind, J., Piikki, K., Stenberg, B. & Soderstrom, M. 2013. Improved predictions of soil texture and soil organic matter content by combining simultaneous in situ measurements of visible and near infrared reflectance, electrical conductivity and insertion force. In: Proceedings of the 3rd Global Workshop on Proximal Soil Sensing, Potsdam, Germany, 26-29 May 2013, (eds., R. Gebbers, E. Luck, & J. Ruhlmann), pp. 161-165. Potsdam, Germany: ATB Leibniz-Institut fur Agratechnik Potsdam-Bornim.

Williams, P.C. 2001. Variables Affecting Near-Infrared Reflectance Spectroscopic Analysis. In Near-Infrared Technology in the Agricultural and Food Industries (eds P. Williams and K. Norris, 143-166. Saint Paul, MN: American Association of Cereal Chemists Inc. Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19, 716-723.

Williams, P. C. 2003. Near-InfraRed Technology-Getting the Best Out of Light. PDK Project, Inc., Nanaimo, British Columbia, Canada, pp. 109.

Winstead, A. T., Norwood, S. H., Griffin, T., Runge, M. W., Adrian, A. M., Fulton, J. P., & Kelton, J. 2010. Adoption and Use of Precision Agriculture Technologies by Practitioners. In: Proceedings of the 10th International Conference of Precision Agriculture. Denver, Colorado, USA.

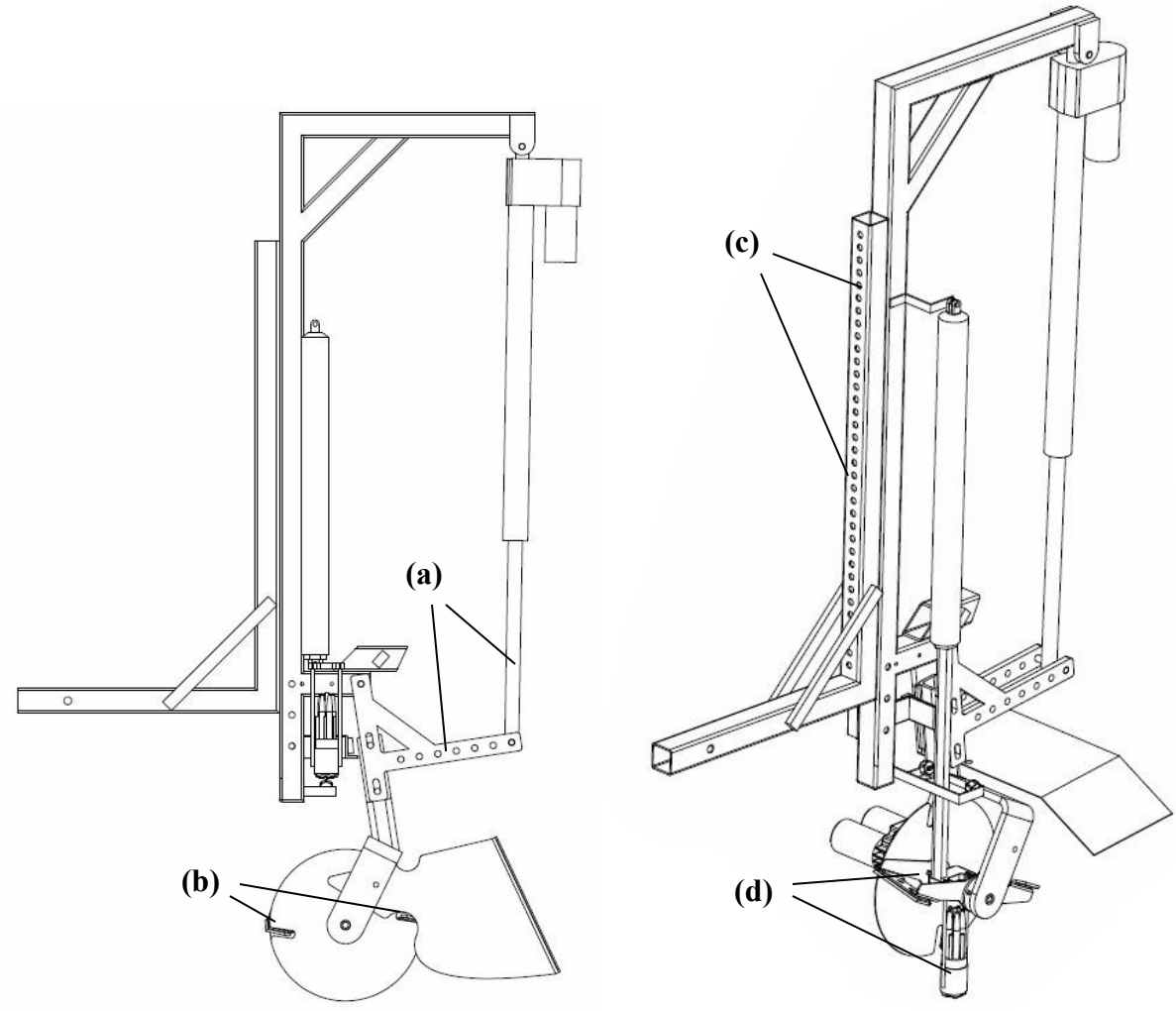
Wold, S., Sjöström, M., & Eriksson, L. 2001. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems, 58, 109-130.

Ziadi, N., Bélanger, G., Gagnon, B., & Mongrain, D. 2009. Mehlich 3 soil phosphorus as determined by colorimetry and inductively coupled plasma. Communications in Soil Science and Plant Analysis, 40, 1-6.

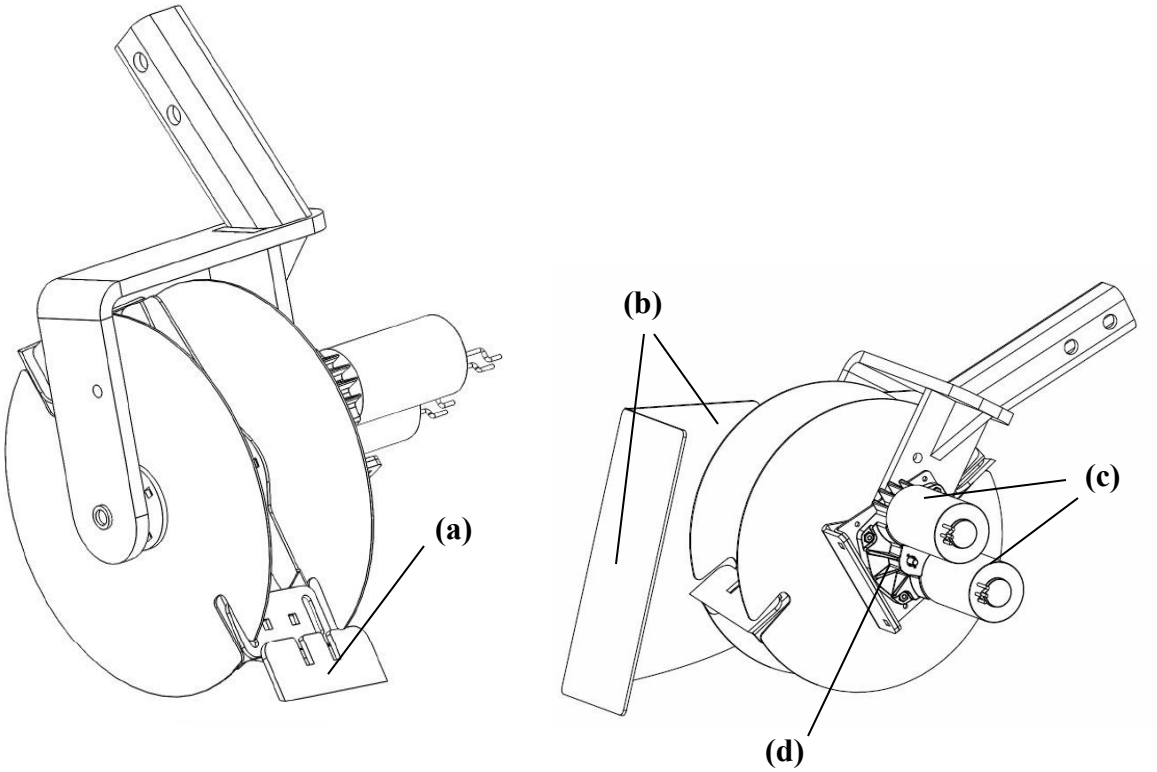
APPENDICES

APPENDIX A

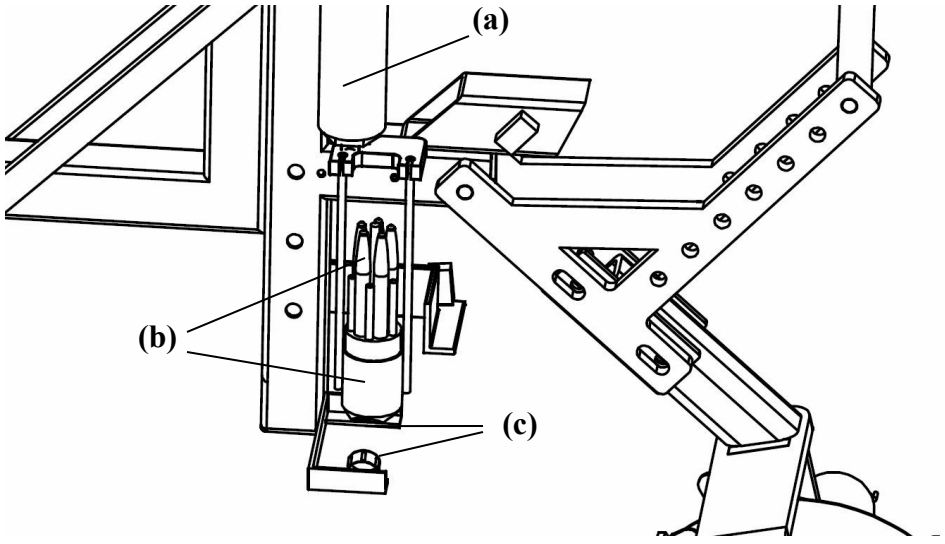
OSA DRAWINGS



Assembly drawings of On-the-spot Soil Analyser (OSA) illustrating: (a) tool cutter in ground position, (b) stopped horizontal postion of the cutting blades, (c) holes with fixed spacing to adjust hitch heights, and (d) deployment of sensing array through provided disc space



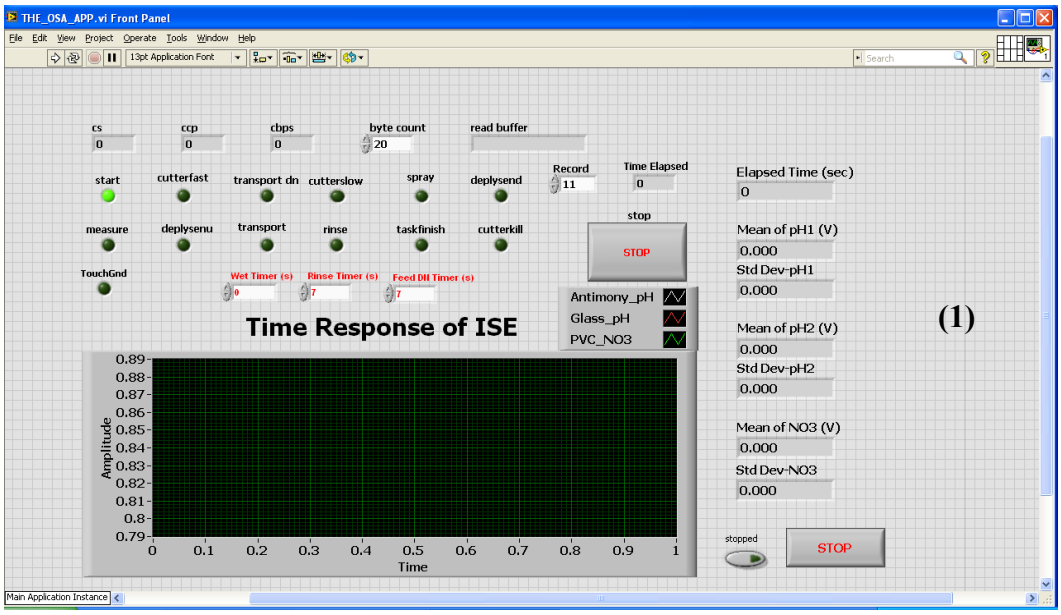
The OSA cutter tool assembly illustrating: (a) removable cutting blades, (b) removable guard, (c) pair of DC motors, and (d) gearbox

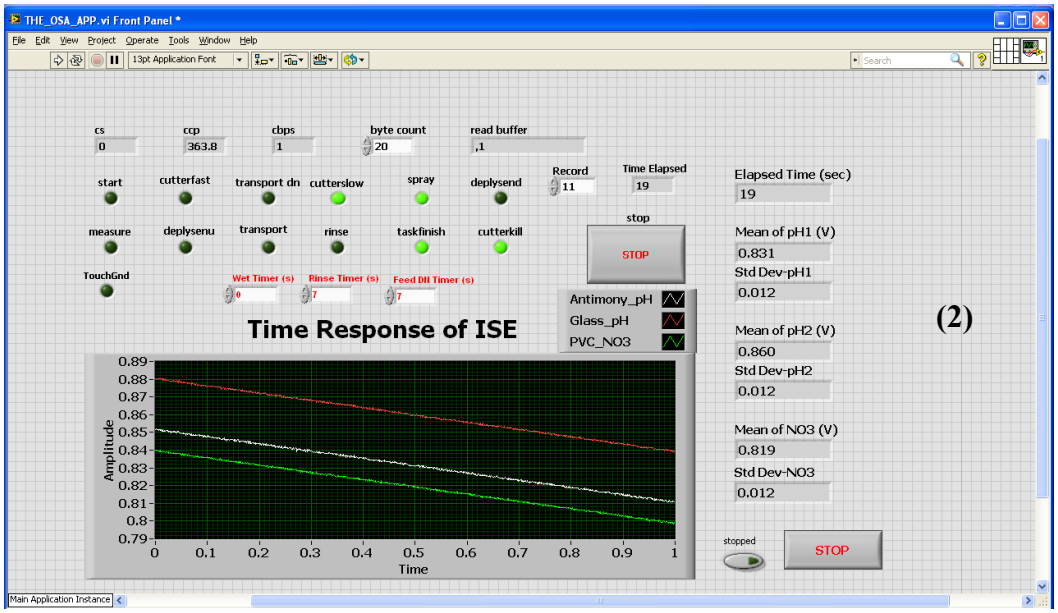


The OSA sensor deployment and sensor cleansing assembly illustrating: (a) the end portion of the linear motor, (b) sensor mount and sensing array, and (c) pair of nozzle sprayers

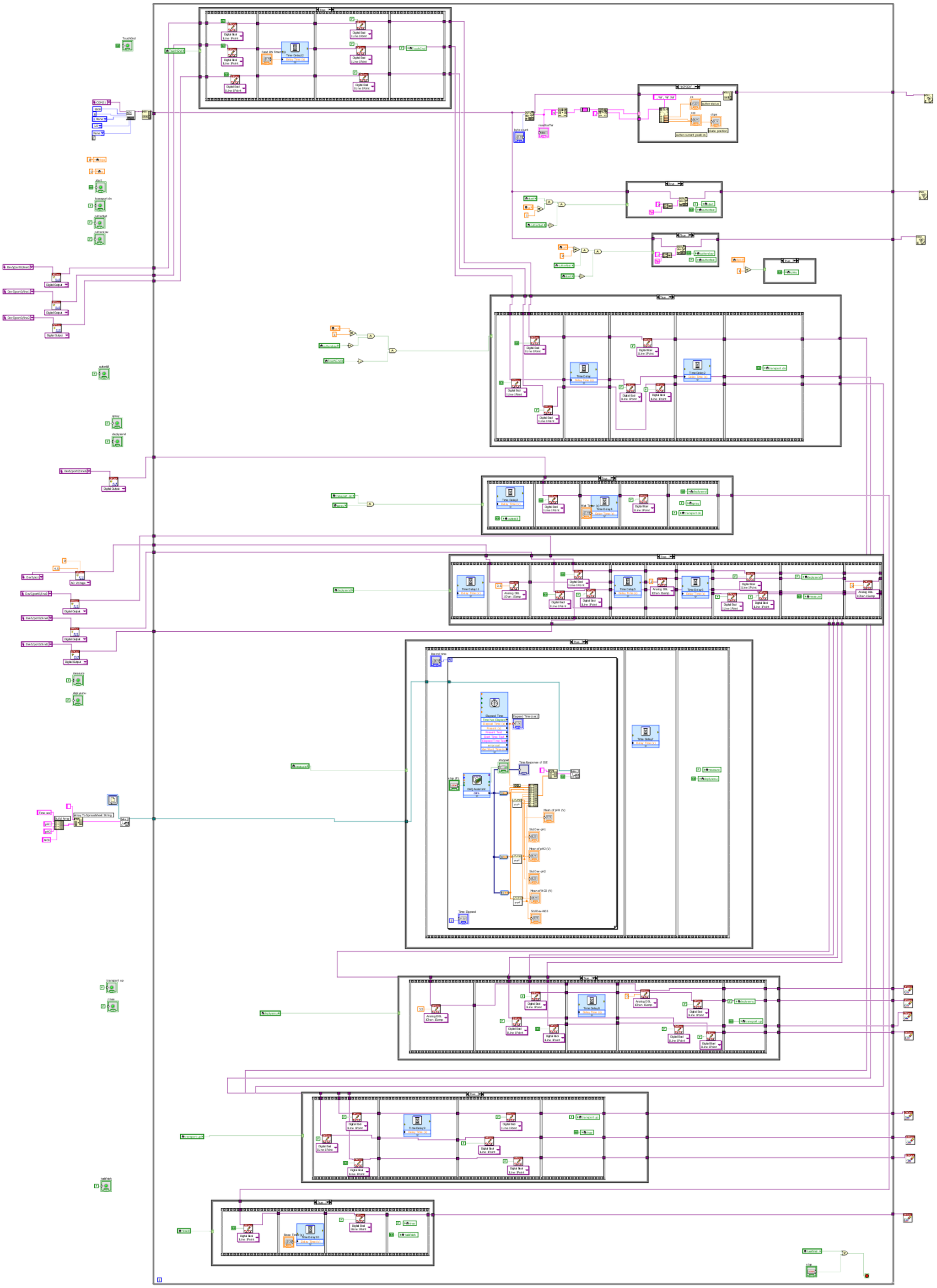
APPENDIX B

LABVIEW ROUTINES





Front panel of the OSA's DAQ and Control Application illustrating task status and process variables (input and output) before (1) and after completion (2) of a typical data collection cycle



LabVIEW code (block diagram) illustrating the OSA’s DAQ and Control Application.

APPENDIX C

ARDUINO ROUTINE

C-1. Arduino script to control speed and blade postion for the cutter on OSA (Chapter 3)

```
/* This example code is in the public domain.Speed and postion control of OSA Cutter motor to
forward, reverse and break OSA Cutter motor

The circuit:
 * CMD signal is attached on pin 3
 * POS Signal is attached on A0

 Created 2014, by Nandkishor Dhawale
  */
String inputString = "";     // a string to hold incoming data
//String stringOne, stringTwo, stringThree;
boolean stringComplete = false;  // whether the string is complete

// Constants won't change and used here to set pin numbers:
const int motCmdPin =  3;      // the number of the Motor pin
int motPosPin = A0;    // select the input pin for the potentiometer
int motSloPin = 4; // to read in for slowing motor if 1
int motSloPinEn = 8; // to enable the pin 4 high

// Variables will change:
int motSloPinVal =0;  //to store the value of motSloPin
int CutBlStat =0; // to inform cutter blade postion
int motCmdState = LOW;             // ledState used to set the LED
unsigned long initial_time = millis();       //initial time
unsigned long time =0 ;                    //  time
int PulseCmd=1500;  // hh
int FWDCmd = 1550;
int REVCmd =1450;
int STPCmd =1500;

double SetPosValue = 490;
double ErrPosValue = 0.00;
int ChgSpd = 1;

// Variables for filtering position data
const int numReadings = 10;
```

```
double readings[numReadings]; // read analog 0
double PosValue = 0.00;  // variable to store the value coming from the position sensor
int index = 0;
double total = 0.00;

//Initialise
void setup() {
//Set the digital pin as output:
  pinMode(motCmdPin, OUTPUT);
  pinMode(motSloPinEn, OUTPUT);
  pinMode(motSloPin, INPUT_PULLUP);
  digitalWrite(motSloPinEn, LOW);
  Serial.begin(9600);
  // reserve 200 bytes for the inputString:
 inputString.reserve(50);
 for (int thisReading = 0; thisReading < numReadings; thisReading++)
  readings[thisReading] = 0;
}

// The main program
void loop()
{
    motSloPinVal = digitalRead(motSloPin); //read cutter slow command from magnetic switch
  if (stringComplete) {

    ChgSpd=inputString.toInt();

 // Clear the string:
    inputString = "";
    stringComplete = false;
  }
    if (motSloPinVal == 1 && ChgSpd == 1){

// Cutter motor stop
      PulseCmd=1500;
      motCmdState = HIGH;
      digitalWrite(motCmdPin, motCmdState);
      delayMicroseconds(PulseCmd);
      motCmdState = LOW;
      digitalWrite(motCmdPin, motCmdState);
      delay(1);
   }
     if (motSloPinVal == 0 && ChgSpd == 0){

 // Cutter motor stop
      PulseCmd=1500;
```

```
   motCmdState = HIGH;
   digitalWrite(motCmdPin, motCmdState);
   delayMicroseconds(PulseCmd);
   motCmdState = LOW;
   digitalWrite(motCmdPin, motCmdState);
   delay(1);
  }
      if (motSloPinVal == 1 && ChgSpd == 0) {

   // Motor fast forward command
   PulseCmd=1950;
   motCmdState = HIGH;
   digitalWrite(motCmdPin, motCmdState);
   delayMicroseconds(PulseCmd);
   motCmdState = LOW;
   digitalWrite(motCmdPin, motCmdState);
   delay(1);
}
 if (motSloPinVal == 0 && ChgSpd == 1)  {
     STPCmd= 1500;
     FWDCmd= 1535;
     REVCmd= 1465;
     ErrPosValue = PosValue-SetPosValue;

 if (ErrPosValue >= -35 && ErrPosValue <= 35) {
  PulseCmd= STPCmd;
  delay(5);
  if (ErrPosValue >= -45 && ErrPosValue <= 45) {
  //delay(2);
  PulseCmd= STPCmd;
  CutBlStat =1;
  //delay(2);
  }
  } //Stop

  if (ErrPosValue > 46) {
  PulseCmd= REVCmd;
  CutBlStat =0;}  // Reverse Speed

  if (ErrPosValue < -46) {
  PulseCmd= FWDCmd;
  CutBlStat =0; }  // Fowward Speed

 // Motor slow forward command
   motCmdState = HIGH;
   digitalWrite(motCmdPin, motCmdState);
```

```
    delayMicroseconds(PulseCmd);
    motCmdState = LOW;
    digitalWrite(motCmdPin, motCmdState);
    delay(1);
   }
  total= total - readings[index];    // read from position the sensor:
  //Readings[index] = ((analogRead(motPosPin)/10)*0.3381)-1.0177;
  readings[index] = analogRead(motPosPin);  // add the reading to the total:
  total= total + readings[index];  // advance to the next position in the array:
  index = index + 1;
   if (index >= numReadings)    {
   // ...Wrap around to the beginning:
   index = 0;

  // Calculate the average
  //PosValue = (total / numReadings)/2.84; //Position in cm
  PosValue = total / numReadings; //Position in cm
  time = millis() - initial_time;
  }

// Send it to the computer as ASCII digits
 Serial.print("$GPGGA"); Serial.print(',');
 //Serial.print(time/1000);  Serial.print(',');
 //Serial.print(PulseCmd); Serial.print(',');
 Serial.print(motSloPinVal); Serial.print(',');
 //Serial.print(SetPosValue); Serial.print(',');
 Serial.print(PosValue); Serial.print(',');
 //Serial.println(ErrPosValue); Serial.print(',');
 Serial.println(CutBlStat); //Serial.print(',');
 delay(1); // delay in between reads for stability
  }
    void serialEvent() {
    while (Serial.available()) {
   // Get the new byte:
   char inChar = (char)Serial.read();
   // Add it to the inputString:
   inputString += inChar;
   // If the incoming character is a newline, set a flag
   // So the main loop can do something about it
   if (inChar == '\n') {
   stringComplete = true;
   }
  }
}
```

APPENDIX D

MATLAB ROUTINES

D-1 Program to analyse *in situ* collected vis-NIR spectral data (Chapter 6)

```
%% Import raw and processed in situ spectral data
exsitu1 = xlsread('VIS_NIR_In_Situ.xlsx', 'ALL_RW');exsitu1 =exsitu1';
exsitu2 = xlsread('VIS_NIR_In_Situ.xlsx', 'ALL_F');exsitu2 =exsitu2';
exsitu3 = xlsread('VIS_NIR_In_Situ.xlsx', 'ALL_FD');exsitu3 =exsitu3';
exsitu4 = xlsread('VIS_NIR_In_Situ.xlsx', 'ALL_SD');exsitu4 =exsitu4';

%% Combine factors
vis_nir_spectra=[exsitu1(6:end,2:end); exsitu2(6:end,2:end); exsitu3(6:end,2:end); ...
exsitu4(6:end,2:end); exsitu1(2:5,2:end)];
vis_nir_factors=[exsitu1(6:end,1); exsitu2(6:end,1); ...
exsitu3(6:end,1); exsitu4(6:end,1);];
wavelengths=vis_nir_factors(1:363,:);

%% Calculate average spectrum
j=1;
for i=1:1:19,
 avg1(:,j)=nanmean([vis_nir_spectra(:,i) vis_nir_spectra(:,i+19) vis_nir_spectra(:,i+38)],2);
 var1(:,j)=nanvar([vis_nir_spectra(:,i) vis_nir_spectra(:,i+19) vis_nir_spectra(:,i+38)]')';
 j=j+1;
end

%% Calculate some statistics
den= sqrt(mean(var1,2));
num=std(avg1')';
RSE=num./den;
VAR_DATA2=var(vis_nir_spectra')';
MST=var(avg1')'*3;
MSE=mean(var1,2);
TOTAL=VAR_DATA2*56;
SST=MST*19;
SSE=MSE*38;
F_MEAN=MST./MSE;
F_IND=F_MEAN/3;
RMSE=power(MSE, 1/2);
```

```
%% Plot soil spectral figure
figure(1);
plot(vis_nir_factors(1:363,:), RSE(1:363,:), 'or'); hold on
plot(vis_nir_factors(364:726,:), RSE(364:726,:), 'xb'); hold on
plot(vis_nir_factors(727:1089,:), RSE(727:1089,:), '*k'); hold on
plot(vis_nir_factors(1090:1452,:), RSE(1090:1452,:), '+m'); grid on
xlabel('Wavelengths, nm', 'FontSize',16)
ylabel('RSE ', 'FontSize',16)
set(gca,'XTick',400:200:2200); set(gca,'YTick',0.0:1:12);
legend('Original', 'Smooth', 'First-derivative', 'Second-derivative', ...
 'Location','NorthEast')
xlim([400 2200])
ylim([0.0 12.0])

%% Calculate the correlations with raw spectrum
ll=1; hl=363;
Y=vis_nir_spectra(1453:1456,1:19)';
X=vis_nir_spectra(ll:hl,1:19)';

for i=1:1:363
  RHOt11(i,1) = corr(X(:,i),Y(:,1));
  RHOt12(i,1) = corr(X(:,i),Y(:,2));
  RHOt13(i,1) = corr(X(:,i),Y(:,3));
end

X=vis_nir_spectra(ll:hl,20:38)';

for i=1:1:363
  RHOt21(i,1) = corr(X(:,i),Y(:,1));
  RHOt22(i,1) = corr(X(:,i),Y(:,2));
  RHOt23(i,1) = corr(X(:,i),Y(:,3));
end

X=vis_nir_spectra(ll:hl,39:57)';

for i=1:1:363
  RHOt31(i,1) = corr(X(:,i),Y(:,1));
  RHOt32(i,1) = corr(X(:,i),Y(:,2));
  RHOt33(i,1) = corr(X(:,i),Y(:,3));
end

X=avg1(ll:hl,:)';
for i=1:1:363
  RHOM1(i,1) = corr(X(:,i),Y(:,1));
  RHOM2(i,1) = corr(X(:,i),Y(:,2));
  RHOM3(i,1) = corr(X(:,i),Y(:,3));
```

```
end

X=vis_nir_spectra(ll:hl,:)';
Y=vis_nir_spectra(1453:1456,:)';

for i=1:1:363
  RHOall1(i,1) = corr(X(:,i),Y(:,1));
  RHOall2(i,1) = corr(X(:,i),Y(:,2));
  RHOall3(i,1) = corr(X(:,i),Y(:,3));
end

corr_raw= [RHOall1, RHOall2 RHOall3, RHOM1, RHOM2, RHOM3, RHOt11, RHOt12,
RHOt13, RHOt21, RHOt22, RHOt23, RHOt31, RHOt32, RHOt33];

%% Calculate the correlations with filtered spectrum
ll=364; hl=726;
Y=vis_nir_spectra(1453:1456,1:19)';
X=vis_nir_spectra(ll:hl,1:19)';

for i=1:1:363
  RHOt11(i,1) = corr(X(:,i),Y(:,1));
  RHOt12(i,1) = corr(X(:,i),Y(:,2));
  RHOt13(i,1) = corr(X(:,i),Y(:,3));
end

X=vis_nir_spectra(ll:hl,20:38)';

for i=1:1:363
  RHOt21(i,1) = corr(X(:,i),Y(:,1));
  RHOt22(i,1) = corr(X(:,i),Y(:,2));
  RHOt23(i,1) = corr(X(:,i),Y(:,3));
end

X=vis_nir_spectra(ll:hl,39:57)';

for i=1:1:363
  RHOt31(i,1) = corr(X(:,i),Y(:,1));
  RHOt32(i,1) = corr(X(:,i),Y(:,2));
  RHOt33(i,1) = corr(X(:,i),Y(:,3));
end

X=avg1(ll:hl,:)';
for i=1:1:363
  RHOM1(i,1) = corr(X(:,i),Y(:,1));
  RHOM2(i,1) = corr(X(:,i),Y(:,2));
  RHOM3(i,1) = corr(X(:,i),Y(:,3));
```

```
end

X=vis_nir_spectra(ll:hl,:)';
Y=vis_nir_spectra(1453:1456,:)';

for i=1:1:363
  RHOall1(i,1) = corr(X(:,i),Y(:,1));
  RHOall2(i,1) = corr(X(:,i),Y(:,2));
  RHOall3(i,1) = corr(X(:,i),Y(:,3));
end

corr_filt= [RHOall1, RHOall2 RHOall3, RHOM1, RHOM2, RHOM3, RHOt11, RHOt12,
RHOt13, RHOt21, RHOt22, RHOt23, RHOt31, RHOt32, RHOt33];

%% Calculate correlations with first derivative of spectra
ll=727; hl=1089;
Y=vis_nir_spectra(1453:1456,1:19)';
X=vis_nir_spectra(ll:hl,1:19)';

for i=1:1:363
  RHOt11(i,1) = corr(X(:,i),Y(:,1));
  RHOt12(i,1) = corr(X(:,i),Y(:,2));
  RHOt13(i,1) = corr(X(:,i),Y(:,3));
end

X=vis_nir_spectra(ll:hl,20:38)';

for i=1:1:363
  RHOt21(i,1) = corr(X(:,i),Y(:,1));
  RHOt22(i,1) = corr(X(:,i),Y(:,2));
  RHOt23(i,1) = corr(X(:,i),Y(:,3));
end

X=vis_nir_spectra(ll:hl,39:57)';

for i=1:1:363
  RHOt31(i,1) = corr(X(:,i),Y(:,1));
  RHOt32(i,1) = corr(X(:,i),Y(:,2));
  RHOt33(i,1) = corr(X(:,i),Y(:,3));
end

X=avg1(ll:hl,:)';
for i=1:1:363
  RHOM1(i,1) = corr(X(:,i),Y(:,1));
  RHOM2(i,1) = corr(X(:,i),Y(:,2));
  RHOM3(i,1) = corr(X(:,i),Y(:,3));
```

```
end

X=vis_nir_spectra(ll:hl,:)';
Y=vis_nir_spectra(1453:1456,:)';

for i=1:1:363
  RHOall1(i,1) = corr(X(:,i),Y(:,1));
  RHOall2(i,1) = corr(X(:,i),Y(:,2));
  RHOall3(i,1) = corr(X(:,i),Y(:,3));
end

corr_fd= [RHOall1, RHOall2 RHOall3, RHOM1, RHOM2, RHOM3, RHOt11, RHOt12,
RHOt13, RHOt21, RHOt22, RHOt23, RHOt31, RHOt32, RHOt33];

%% Calculate correlations with second derivative spectra
ll=1090; hl=1452;
Y=vis_nir_spectra(1453:1456,1:19)';
X=vis_nir_spectra(ll:hl,1:19)';

for i=1:1:363
  RHOt11(i,1) = corr(X(:,i),Y(:,1));
  RHOt12(i,1) = corr(X(:,i),Y(:,2));
  RHOt13(i,1) = corr(X(:,i),Y(:,3));
end

X=vis_nir_spectra(ll:hl,20:38)';

for i=1:1:363
  RHOt21(i,1) = corr(X(:,i),Y(:,1));
  RHOt22(i,1) = corr(X(:,i),Y(:,2));
  RHOt23(i,1) = corr(X(:,i),Y(:,3));
end

X=vis_nir_spectra(ll:hl,39:57)';

for i=1:1:363
  RHOt31(i,1) = corr(X(:,i),Y(:,1));
  RHOt32(i,1) = corr(X(:,i),Y(:,2));
  RHOt33(i,1) = corr(X(:,i),Y(:,3));
end

X=avg1(ll:hl,:)';
for i=1:1:363
  RHOM1(i,1) = corr(X(:,i),Y(:,1));
  RHOM2(i,1) = corr(X(:,i),Y(:,2));
  RHOM3(i,1) = corr(X(:,i),Y(:,3));
```

```
end

X=vis_nir_spectra(ll:hl,:)';
Y=vis_nir_spectra(1453:1456,:)';

for i=1:1:363
  RHOall1(i,1) = corr(X(:,i),Y(:,1));
  RHOall2(i,1) = corr(X(:,i),Y(:,2));
  RHOall3(i,1) = corr(X(:,i),Y(:,3));
end

corr_sd= [RHOall1, RHOall2 RHOall3, RHOM1, RHOM2, RHOM3, RHOt11, RHOt12,
RHOt13, RHOt21, RHOt22, RHOt23, RHOt31, RHOt32, RHOt33];

figure(2)
plot(RSE(1:363,:), corr_raw(:,1).^2, 'or'); hold on
plot(RSE(364:726,:), corr_filt(:,1).^2, 'xb'); hold on
plot(RSE(727:1089,:), corr_fd(:,1).^2, '*k'); hold on
plot(RSE(1090:1452,:), corr_sd(:,1).^2, '+m'); grid on
xlabel('RSE', 'FontSize',16)
ylabel('R ^2, % sand', 'FontSize',16)
set(gca,'XTick',0:1:20); set(gca,'YTick',0.0:0.10:1.0);
legend('Original', 'Smooth', 'First-derivative', 'Second-derivative', ...
 'Location','NorthEast')
xlim([0 12])
ylim([0.0 0.7])

figure(3)
plot(RSE(1:363,:), corr_raw(:,2).^2, 'or'); hold on
plot(RSE(364:726,:), corr_filt(:,2).^2, 'xb'); hold on
plot(RSE(727:1089,:), corr_fd(:,2).^2, '*k'); hold on
plot(RSE(1090:1452,:), corr_sd(:,2).^2, '+m'); grid on
xlabel('RSE', 'FontSize',16)
ylabel('R ^2, % clay', 'FontSize',16)
set(gca,'XTick',0:1:20); set(gca,'YTick',0.0:0.10:1.0);
legend('Original', 'Smooth', 'First-derivative', 'Second-derivative', ...
 'Location','NorthEast')
xlim([0 12])
ylim([0.0 0.9])

figure(4)
plot(RSE(1:363,:), corr_raw(:,3).^2, 'or'); hold on
plot(RSE(364:726,:), corr_filt(:,3).^2, 'xb'); hold on
plot(RSE(727:1089,:), corr_fd(:,3).^2, '*k'); hold on
plot(RSE(1090:1452,:), corr_sd(:,3).^2, '+m'); grid on
xlabel('RSE', 'FontSize',16)
```

```
ylabel('R ^2, %SOM', 'FontSize',16)
set(gca,'XTick',0:1:20); set(gca,'YTick',0.0:0.10:1.0);
legend('Original', 'Smooth', 'First-derivative', 'Second-derivative', ...
 'Location','NorthEast')
xlim([0 12])
ylim([0.0 0.7])

figure(5)
plot(wavelengths, avg1(1:363,:), '.-.'); grid on
xlabel('Wavelengths, nm', 'FontSize',16)
ylabel('Reflectance ', 'FontSize',16)
xlim([400 2200])
ylim([0.0 0.6])

%% Calculate precsison and accuracy parameters on raw spectra
ll=1; hl=363;
Y=vis_nir_spectra(1453:1456,:)';
X=vis_nir_spectra(ll:hl,:)';
c=length(Y);

for i=1:1:363
p = polyfit(X(:,i),Y(:,1),1); yhat=p(2)+p(1).*X(:,i); Pall1(i,1)=p(1);
RMSEall1(i,1)=sqrt(mean((Y(:,1)-yhat).^2));P2all1(i,1)=p(2);
SDEall1(i,1)=sqrt(sum((Y(:,1)-yhat).^2)/c);% SDE(i,1)=b=yhat-mean(yhat); a= X(:,i)-
mean(X(:,i)); sqrt((sum(b.^2) -((sum(a.*b)).^2/(sum(a.^2))))/c);
p = polyfit(X(:,i),Y(:,2),1); yhat=p(2)+p(1).*X(:,i); Pall2(i,1)=p(1);
RMSEall2(i,1)=sqrt(mean((Y(:,2)-yhat).^2));P2all2(i,1)=p(2);
SDEall2(i,1)=sqrt(sum((Y(:,2)-yhat).^2)/c);
p = polyfit(X(:,i),Y(:,3),1); yhat=p(2)+p(1).*X(:,i); Pall3(i,1)=p(1);
RMSEall3(i,1)=sqrt(mean((Y(:,3)-yhat).^2));P2all3(i,1)=p(2);
SDEall3(i,1)=sqrt(sum((Y(:,3)-yhat).^2)/c);

end

rmse_raw= [RMSEall1, RMSEall2 RMSEall3];
sde_raw=[SDEall1, SDEall2, SDEall3];
slope_raw=[Pall1, Pall2, Pall3];
intcpt_raw=[P2all1, P2all2, P2all3];

%% Calculate precsison and accuracy parameters on filtered spectra
ll=364; hl=726;
X=vis_nir_spectra(ll:hl,:)';

for i=1:1:363
p = polyfit(X(:,i),Y(:,1),1); yhat=p(2)+p(1).*X(:,i); Pall1(i,1)=p(1);
RMSEall1(i,1)=sqrt(mean((Y(:,1)-yhat).^2));P2all1(i,1)=p(2);
```

```
SDEall1(i,1)=sqrt(sum((Y(:,1)-yhat).^2)/c);
p = polyfit(X(:,i),Y(:,2),1); yhat=p(2)+p(1).*X(:,i); Pall2(i,1)=p(1);
RMSEall2(i,1)=sqrt(mean((Y(:,2)-yhat).^2));P2all2(i,1)=p(2);
SDEall2(i,1)=sqrt(sum((Y(:,2)-yhat).^2)/c);
p = polyfit(X(:,i), Y(:,3),1); yhat=p(2)+p(1).*X(:,i); Pall3(i,1)=p(1);
RMSEall3(i,1)=sqrt(mean((Y(:,3)-yhat).^2));P2all3(i,1)=p(2);
SDEall3(i,1)=sqrt(sum((Y(:,3)-yhat).^2)/c);

end

rmse_filt= [RMSEall1, RMSEall2 RMSEall3];
sde_filt=[SDEall1, SDEall2, SDEall3];
slope_filt=[Pall1, Pall2, Pall3];
intcpt_filt=[P2all1, P2all2, P2all3];

%% Calculate precsison and accuracy parameters on first derivative spectra
ll=727; hl=1089;
X=vis_nir_spectra(ll:hl,:)';

for i=1:1:363
p = polyfit(X(:,i),Y(:,1),1); yhat=p(2)+p(1).*X(:,i); Pall1(i,1)=p(1);
RMSEall1(i,1)=sqrt(mean((Y(:,1)-yhat).^2));P2all1(i,1)=p(2);
SDEall1(i,1)=sqrt(sum((Y(:,1)-yhat).^2)/c);
p = polyfit(X(:,i),Y(:,2),1); yhat=p(2)+p(1).*X(:,i); Pall2(i,1)=p(1);
RMSEall2(i,1)=sqrt(mean((Y(:,2)-yhat).^2));P2all2(i,1)=p(2);
SDEall2(i,1)=sqrt(sum((Y(:,2)-yhat).^2)/c);
p = polyfit(X(:,i), Y(:,3),1); yhat=p(2)+p(1).*X(:,i); Pall3(i,1)=p(1);
RMSEall3(i,1)=sqrt(mean((Y(:,3)-yhat).^2));P2all3(i,1)=p(2);
SDEall3(i,1)=sqrt(sum((Y(:,3)-yhat).^2)/c);

end

rmse_fd= [RMSEall1, RMSEall2 RMSEall3];
sde_fd=[SDEall1, SDEall2, SDEall3];
slope_fd=[Pall1, Pall2, Pall3];
intcpt_fd=[P2all1, P2all2, P2all3];

%% Calculate precsison and accuracy parameters on second derivative spectra
ll=1090; hl=1452;
X=vis_nir_spectra(ll:hl,:)';

for i=1:1:363
p = polyfit(X(:,i),Y(:,1),1); yhat=p(2)+p(1).*X(:,i); Pall1(i,1)=p(1);
RMSEall1(i,1)=sqrt(mean((Y(:,1)-yhat).^2));P2all1(i,1)=p(2);
SDEall1(i,1)=sqrt(sum((Y(:,1)-yhat).^2)/c);
```

```
p = polyfit(X(:,i),Y(:,2),1); yhat=p(2)+p(1).*X(:,i); Pall2(i,1)=p(1);
RMSEall2(i,1)=sqrt(mean((Y(:,2)-yhat).^2));P2all2(i,1)=p(2);
SDEall2(i,1)=sqrt(sum((Y(:,2)-yhat).^2)/c);
p = polyfit(X(:,i), Y(:,3),1); yhat=p(2)+p(1).*X(:,i); Pall3(i,1)=p(1);
RMSEall3(i,1)=sqrt(mean((Y(:,3)-yhat).^2));P2all3(i,1)=p(2);
SDEall3(i,1)=sqrt(sum((Y(:,3)-yhat).^2)/c);
end

rmse_sd= [RMSEall1, RMSEall2 RMSEall3];
sde_sd=[SDEall1, SDEall2, SDEall3];
slope_sd=[Pall1, Pall2, Pall3];
intcpt_sd=[P2all1, P2all2, P2all3];

figure(6)
plot(abs(slope_raw(:,1)).*RMSE(1:363,:),rmse_raw(:,1), 'or'); hold on
plot(abs(slope_filt(:,1)).*RMSE(364:726,:),rmse_filt(:,1), 'xb'); hold on
plot(abs(slope_fd(:,1)).*RMSE(727:1089,:),rmse_fd(:,1), '*k'); hold on
plot(abs(slope_sd(:,1)).*RMSE(1090:1452,:),rmse_sd(:,1), '+m'); hold on
zz=std(Y(:,1)); plot((0:2:10), [zz,zz,zz,zz,zz,zz], 'linewidth',3); grid on
xlabel('Measurement Precision, % sand', 'FontSize',16)
ylabel('Standard Error of Prediction, % sand', 'FontSize',16)
set(gca,'XTick',0:1:20); set(gca,'YTick',0.0:2:20);
plot([15,0], [0,15], '.-.'); plot([16,0], [0,16], '.-.')
plot([17,0], [0,17], '.-.'); plot([18,0], [0,18], '.-.')
plot([19,0], [0,19], '.-.'); plot([20,0], [0,20], '.-.')
plot([21,0], [0,21], '.-.');plot([22,0], [0,22], '.-.')
plot([23,0], [0,23], '.-.')
plot(3.12, 4.82, 'dg', 'markersize',12)
plot(2.64, 4.09, 'or', 'markersize',12)
plot(3.43, 4.96, 'xb', 'markersize',12)
plot(3.29, 4.46, '*k', 'markersize',12)
plot(3.09, 5.85, '+m', 'markersize',12)
xlim([0 10]); ylim([0 20])
legend('Original', 'Smooth', 'First-derivative', 'Second-derivative', ...
 'Location','SouthEast')

figure(7)
plot(abs(slope_raw(:,2)).*RMSE(1:363,:),rmse_raw(:,2), 'or'); hold on
plot(abs(slope_filt(:,2)).*RMSE(364:726,:),rmse_filt(:,2), 'xb'); hold on
plot(abs(slope_fd(:,2)).*RMSE(727:1089,:),rmse_fd(:,2), '*k'); hold on
plot(abs(slope_sd(:,2)).*RMSE(1090:1452,:),rmse_sd(:,2), '+m'); hold on
zz=std(Y(:,2)); plot((0:5), [zz,zz,zz,zz,zz zz], 'linewidth',3); grid on
xlabel('Measurement Precision, % clay', 'FontSize',16)
ylabel('Standard Error of Prediction, % clay', 'FontSize',16)
set(gca,'XTick',0:1:10); set(gca,'YTick',0:3:20);
plot([6,0], [0,6], '.-.'); plot([7,0], [0,7], '.-.')
```

```
plot([8,0], [0,8], '.-.'); plot([9,0], [0,9], '.-.')
plot([10,0], [0,10], '.-.'); plot([11,0], [0,11], '.-.')
plot([12,0], [0,12], '.-.'); plot([5,0], [0,5], '.-.');
plot(1.81, 3.48, 'dg', 'markersize',12)
plot(1.91, 4.59, 'or', 'markersize',12)
plot(1.66, 2.89, 'xb', 'markersize',12)
plot(1.28, 1.73, '*k', 'markersize',12)
plot(1.23, 2.04, '+m', 'markersize',12)
xlim([0 5]); ylim([0 12])
legend('Original', 'Smooth', 'First-derivative', 'Second-derivative', ...
 'Location','SouthEast')

figure(8)
plot(abs(slope_raw(:,3)).*RMSE(1:363,:),rmse_raw(:,3), 'or'); hold on
plot(abs(slope_filt(:,3)).*RMSE(364:726,:),rmse_filt(:,3), 'xb'); hold on
plot(abs(slope_fd(:,3)).*RMSE(727:1089,:),rmse_fd(:,3), '*k'); hold on
plot(abs(slope_sd(:,3)).*RMSE(1090:1452,:),rmse_sd(:,3), '+m'); hold on
zz=std(Y(:,3)); plot((0:4), [zz,zz,zz,zz,zz], 'linewidth',3); grid on
xlabel('Measurement Precision, % SOM', 'FontSize',16)
ylabel('Standard Error of Prediction, % SOM', 'FontSize',16)
set(gca,'XTick',0:1:10); set(gca,'YTick',0:3:20);
plot([4,0], [0,4], '.-.');plot([4.5,0], [0,4.5], '.-.')
plot([5,0], [0,5], '.-.') ;plot([5.5,0], [0,5.5], '.-.');
plot([6,0], [0,6], '.-.')
plot(1.38, 2.14, 'dg', 'markersize',12)
plot(1.62, 4.18, 'or', 'markersize',12)
plot(1.67, 2.44, 'xb', 'markersize',12)
plot(1.31, 2.05, '*k', 'markersize',12)
plot(1.39, 2.75, '+m', 'markersize',12)
xlim([0 3]); ylim([0 7])
legend('Original', 'Smooth', 'First-Derivative', 'Second-Derivative', ...
 'Location','SouthEast')

figure(9)
tt= mean(avg1(1:363,:),2);
tt1=std(avg1(1:363,:)')';
plot(wavelengths, tt+tt1, '--g', 'markersize',5); hold on
plot(wavelengths, tt, '-k', 'markersize',5); hold on
plot(wavelengths, tt-tt1, '--g', 'markersize',5); hold on
text(1462, tt(190),'\leftarrow sand (best)','HorizontalAlignment','left')
text(1940,tt(296),' sand, clay (best) \rightarrow', 'HorizontalAlignment','right')
text(421,tt(2),'\leftarrow SOM ', 'HorizontalAlignment','left')
text(1747,tt(251),'\leftarrow SOM (best) ', 'HorizontalAlignment','left')
text(1605,tt(220),'SOM \rightarrow ', 'HorizontalAlignment','right')
text(590,tt(31),'\leftarrow SOM ', 'HorizontalAlignment','left')
text(722,tt(55),'\leftarrow SOM ', 'HorizontalAlignment','left')
```

```
set(gca,'XTick',400:300:2200); set(gca,'YTick',0:0.1:0.6);
xlabel('Wavelengths, nm', 'FontSize',16)
ylabel('Reflectance ', 'FontSize',16)
xlim([400 2200])
ylim([0.0 0.52])
grid on
```

D-2 Program to analyse *ex situ* collected vis-NIR spectral data (Chapter 6)

```
%% Import ex situ soil spectral data
exsitu1 = xlsread('VIS_NIR_Ex_Situ.xlsx', 'ALL_RW');exsitu1 =exsitu1';
exsitu2 = xlsread('VIS_NIR_Ex_Situ.xlsx', 'ALL_F');exsitu2 =exsitu2';
exsitu3 = xlsread('VIS_NIR_Ex_Situ.xlsx', 'ALL_FD');exsitu3 =exsitu3';
exsitu4 = xlsread('VIS_NIR_Ex_Situ.xlsx', 'ALL_SD');exsitu4 =exsitu4';

%% Combine factors
vis_nir_spectra=[exsitu1(6:end,2:end); exsitu2(6:end,2:end); exsitu3(6:end,2:end); ...
 exsitu4(6:end,2:end); exsitu1(2:5,2:end)];
vis_nir_spectra(:,6)=[]; vis_nir_spectra(:,25)=[]; vis_nir_spectra(:,44)=[];
vis_nir_factors=[exsitu1(6:end,1); exsitu2(6:end,1); ...
 exsitu3(6:end,1); exsitu4(6:end,1);];
wavelengths=vis_nir_factors(1:363,:);

%% Calculate average spectrum
j=1;
for i=1:1:19,
 avg1(:,j)=nanmean([vis_nir_spectra(:,i) vis_nir_spectra(:,i+19) vis_nir_spectra(:,i+38)],2);
 var1(:,j)=nanvar([vis_nir_spectra(:,i) vis_nir_spectra(:,i+19) vis_nir_spectra(:,i+38)]')';
 j=j+1;
end

%% Calculate some statistics
den= sqrt(mean(var1,2));
num=std(avg1')';
RSE=num./den;
VAR_DATA2=var(vis_nir_spectra')';
MST=var(avg1')'*3;
MSE=mean(var1,2);
TOTAL=VAR_DATA2*56;
SST=MST*19;
SSE=MSE*38;
F_MEAN=MST./MSE;
F_IND=F_MEAN/3;
RMSE=power(MSE, 1/2);

%% Plot the ex situ spectra
figure(1);
plot(vis_nir_factors(1:363,:), RSE(1:363,:), 'or'); hold on
plot(vis_nir_factors(364:726,:), RSE(364:726,:), 'xb'); hold on
plot(vis_nir_factors(727:1089,:), RSE(727:1089,:), '*k'); hold on
plot(vis_nir_factors(1090:1452,:), RSE(1090:1452,:), '+m'); grid on
xlabel('Wavelengths, nm', 'FontSize',12)
```

```
ylabel('RSE ', 'FontSize',12)
set(gca,'XTick',400:200:2200); set(gca,'YTick',0.0:1:12);
legend('Original', 'Smooth', 'First-Derivative', 'Second-Derivative 4', ...
 'Location','NorthWest')
xlim([400 2200])
ylim([0.0 12.0])

%% Calculate correlations with raw spectra
ll=1; hl=363;
Y=vis_nir_spectra(1453:1456,1:19)';

X=vis_nir_spectra(ll:hl,1:19)';

for i=1:1:363
  RHOt11(i,1) = corr(X(:,i),Y(:,1));
  RHOt12(i,1) = corr(X(:,i),Y(:,2));
  RHOt13(i,1) = corr(X(:,i),Y(:,3));
end

X=vis_nir_spectra(ll:hl,20:38)';

for i=1:1:363
  RHOt21(i,1) = corr(X(:,i),Y(:,1));
  RHOt22(i,1) = corr(X(:,i),Y(:,2));
  RHOt23(i,1) = corr(X(:,i),Y(:,3));
end

X=vis_nir_spectra(ll:hl,39:57)';

for i=1:1:363
  RHOt31(i,1) = corr(X(:,i),Y(:,1));
  RHOt32(i,1) = corr(X(:,i),Y(:,2));
  RHOt33(i,1) = corr(X(:,i),Y(:,3));
end

X=avg1(ll:hl,:)';
for i=1:1:363
  RHOM1(i,1) = corr(X(:,i),Y(:,1));
  RHOM2(i,1) = corr(X(:,i),Y(:,2));
  RHOM3(i,1) = corr(X(:,i),Y(:,3));
end

X=vis_nir_spectra(ll:hl,:)';
Y=vis_nir_spectra(1453:1456,:)';

for i=1:1:363
```

```
  RHOall1(i,1) = corr(X(:,i),Y(:,1));
  RHOall2(i,1) = corr(X(:,i),Y(:,2));
  RHOall3(i,1) = corr(X(:,i),Y(:,3));
end

corr_raw= [RHOall1, RHOall2 RHOall3, RHOM1, RHOM2, RHOM3, RHOt11, RHOt12,
RHOt13, RHOt21, RHOt22, RHOt23, RHOt31, RHOt32, RHOt33];

%% Calculate correlations with filtered spectra
ll=364; hl=726;
Y=vis_nir_spectra(1453:1456,1:19)';
X=vis_nir_spectra(ll:hl,1:19)';


for i=1:1:363
  RHOt11(i,1) = corr(X(:,i),Y(:,1));
  RHOt12(i,1) = corr(X(:,i),Y(:,2));
  RHOt13(i,1) = corr(X(:,i),Y(:,3));
end

X=vis_nir_spectra(ll:hl,20:38)';

for i=1:1:363
  RHOt21(i,1) = corr(X(:,i),Y(:,1));
  RHOt22(i,1) = corr(X(:,i),Y(:,2));
  RHOt23(i,1) = corr(X(:,i),Y(:,3));
end

X=vis_nir_spectra(ll:hl,39:57)';

for i=1:1:363
  RHOt31(i,1) = corr(X(:,i),Y(:,1));
  RHOt32(i,1) = corr(X(:,i),Y(:,2));
  RHOt33(i,1) = corr(X(:,i),Y(:,3));
end

X=avg1(ll:hl,:)';
for i=1:1:363
  RHOM1(i,1) = corr(X(:,i),Y(:,1));
  RHOM2(i,1) = corr(X(:,i),Y(:,2));
  RHOM3(i,1) = corr(X(:,i),Y(:,3));
end

X=vis_nir_spectra(ll:hl,:)';
Y=vis_nir_spectra(1453:1456,:)';
```

```
for i=1:1:363
  RHOall1(i,1) = corr(X(:,i),Y(:,1));
  RHOall2(i,1) = corr(X(:,i),Y(:,2));
  RHOall3(i,1) = corr(X(:,i),Y(:,3));
end
corr_filt= [RHOall1, RHOall2 RHOall3, RHOM1, RHOM2, RHOM3, RHOt11, RHOt12,
RHOt13,  RHOt21, RHOt22, RHOt23, RHOt31, RHOt32, RHOt33];

%% Calculate correlations with first derivative of spectra
ll=727; hl=1089;
Y=vis_nir_spectra(1453:1456,1:19)';
X=vis_nir_spectra(ll:hl,1:19)';


for i=1:1:363
  RHOt11(i,1) = corr(X(:,i),Y(:,1));
  RHOt12(i,1) = corr(X(:,i),Y(:,2));
  RHOt13(i,1) = corr(X(:,i),Y(:,3));
end

X=vis_nir_spectra(ll:hl,20:38)';

for i=1:1:363
  RHOt21(i,1) = corr(X(:,i),Y(:,1));
  RHOt22(i,1) = corr(X(:,i),Y(:,2));
  RHOt23(i,1) = corr(X(:,i),Y(:,3));
end

X=vis_nir_spectra(ll:hl,39:57)';

for i=1:1:363
  RHOt31(i,1) = corr(X(:,i),Y(:,1));
  RHOt32(i,1) = corr(X(:,i),Y(:,2));
  RHOt33(i,1) = corr(X(:,i),Y(:,3));
end

X=avg1(ll:hl,:)';
for i=1:1:363
  RHOM1(i,1) = corr(X(:,i),Y(:,1));
  RHOM2(i,1) = corr(X(:,i),Y(:,2));
  RHOM3(i,1) = corr(X(:,i),Y(:,3));
end

X=vis_nir_spectra(ll:hl,:)';
Y=vis_nir_spectra(1453:1456,:)';
```

```
for i=1:1:363
  RHOall1(i,1) = corr(X(:,i),Y(:,1));
  RHOall2(i,1) = corr(X(:,i),Y(:,2));
  RHOall3(i,1) = corr(X(:,i),Y(:,3));
end
corr_fd= [RHOall1, RHOall2 RHOall3, RHOM1, RHOM2, RHOM3, RHOt11, RHOt12,
RHOt13,  RHOt21, RHOt22, RHOt23, RHOt31, RHOt32, RHOt33];

%% Calculate correlations with second derivative spectra
ll=1090; hl=1452;
Y=vis_nir_spectra(1453:1456,1:19)';
X=vis_nir_spectra(ll:hl,1:19)';


for i=1:1:363
  RHOt11(i,1) = corr(X(:,i),Y(:,1));
  RHOt12(i,1) = corr(X(:,i),Y(:,2));
  RHOt13(i,1) = corr(X(:,i),Y(:,3));
end

X=vis_nir_spectra(ll:hl,20:38)';

for i=1:1:363
  RHOt21(i,1) = corr(X(:,i),Y(:,1));
  RHOt22(i,1) = corr(X(:,i),Y(:,2));
  RHOt23(i,1) = corr(X(:,i),Y(:,3));
end

X=vis_nir_spectra(ll:hl,39:57)';

for i=1:1:363
  RHOt31(i,1) = corr(X(:,i),Y(:,1));
  RHOt32(i,1) = corr(X(:,i),Y(:,2));
  RHOt33(i,1) = corr(X(:,i),Y(:,3));
end

X=avg1(ll:hl,:)';
for i=1:1:363
  RHOM1(i,1) = corr(X(:,i),Y(:,1));
  RHOM2(i,1) = corr(X(:,i),Y(:,2));
  RHOM3(i,1) = corr(X(:,i),Y(:,3));
end

X=vis_nir_spectra(ll:hl,:)';
Y=vis_nir_spectra(1453:1456,:)';
```

```
for i=1:1:363
  RHOall1(i,1) = corr(X(:,i),Y(:,1));
  RHOall2(i,1) = corr(X(:,i),Y(:,2));
  RHOall3(i,1) = corr(X(:,i),Y(:,3));
end
corr_sd= [RHOall1, RHOall2 RHOall3, RHOM1, RHOM2, RHOM3, RHOt11, RHOt12,
RHOt13,  RHOt21, RHOt22, RHOt23, RHOt31, RHOt32, RHOt33];

figure(2)
plot(RSE(1:363,:), corr_raw(:,1).^2, 'or'); hold on
plot(RSE(364:726,:), corr_filt(:,1).^2, 'xb'); hold on
plot(RSE(727:1089,:), corr_fd(:,1).^2, '*k'); hold on
plot(RSE(1090:1452,:), corr_sd(:,1).^2, '+m'); grid on
xlabel('RSE', 'FontSize',12)
ylabel('R ^2, %sand', 'FontSize',12)
set(gca,'XTick',0:1:20); set(gca,'YTick',0.0:0.10:1.0);
legend Original Smooth First-Derivative Second-Derivative

figure(3)
plot(RSE(1:363,:), corr_raw(:,2).^2, 'or'); hold on
plot(RSE(364:726,:), corr_filt(:,2).^2, 'xb'); hold on
plot(RSE(727:1089,:), corr_fd(:,2).^2, '*k'); hold on
plot(RSE(1090:1452,:), corr_sd(:,2).^2, '+m'); grid on
xlabel('RSE', 'FontSize',12)
ylabel('R ^2, %clay', 'FontSize',12)
set(gca,'XTick',0:1:20); set(gca,'YTick',0.0:0.10:1.0);
legend('Original', 'Smooth', 'First-Derivative', 'Second-Derivative 4', ...
 'Location','SouthEast')

figure(4)
plot(RSE(1:363,:), corr_raw(:,3).^2, 'or'); hold on
plot(RSE(364:726,:), corr_filt(:,3).^2, 'xb'); hold on
plot(RSE(727:1089,:), corr_fd(:,3).^2, '*k'); hold on
plot(RSE(1090:1452,:), corr_sd(:,3).^2, '+m'); grid on
xlabel('RSE', 'FontSize',12)
ylabel('R ^2, %carbon', 'FontSize',12)
set(gca,'XTick',0:1:20); set(gca,'YTick',0.0:0.10:1.0);
legend Original Smooth First-Derivative Second-Derivative

figure(5)
plot(wavelengths, avg1(1:363,:), '.-.'); grid on
xlabel('Wavelengths, nm', 'FontSize',12)
ylabel('Reflectance ', 'FontSize',12)
xlim([400 2200])
ylim([0.0 0.6])
```

```
%% Calculate precisions and accuracy parameters on raw spectra
ll=1; hl=363;
Y=vis_nir_spectra(1453:1456,:)';
X=vis_nir_spectra(ll:hl,:)';
c=length(Y);

for i=1:1:363
p = polyfit(X(:,i),Y(:,1),1); yhat=p(2)+p(1).*X(:,i); Pall1(i,1)=p(1);
RMSEall1(i,1)=sqrt(mean((Y(:,1)-yhat).^2));P2all1(i,1)=p(2);
SDEall1(i,1)=sqrt(sum((Y(:,1)-yhat).^2)/c);% SDE(i,1)=b=yhat-mean(yhat); a= X(:,i)-
mean(X(:,i)); sqrt((sum(b.^2) -((sum(a.*b)).^2/(sum(a.^2))))/c);
p = polyfit(X(:,i),Y(:,2),1); yhat=p(2)+p(1).*X(:,i); Pall2(i,1)=p(1);
RMSEall2(i,1)=sqrt(mean((Y(:,2)-yhat).^2));P2all2(i,1)=p(2);
SDEall2(i,1)=sqrt(sum((Y(:,2)-yhat).^2)/c);
p = polyfit(X(:,i),Y(:,3),1); yhat=p(2)+p(1).*X(:,i); Pall3(i,1)=p(1);
RMSEall3(i,1)=sqrt(mean((Y(:,3)-yhat).^2));P2all3(i,1)=p(2);
SDEall3(i,1)=sqrt(sum((Y(:,3)-yhat).^2)/c);
end

rmse_raw= [RMSEall1, RMSEall2 RMSEall3];
sde_raw=[SDEall1, SDEall2, SDEall3];
slope_raw=[Pall1, Pall2, Pall3];
intcpt_raw=[P2all1, P2all2, P2all3];

%% Calculate precisions and accuracy parameters on filtered spectra
ll=364; hl=726;
X=vis_nir_spectra(ll:hl,:)';

for i=1:1:363
p = polyfit(X(:,i),Y(:,1),1); yhat=p(2)+p(1).*X(:,i); Pall1(i,1)=p(1);
RMSEall1(i,1)=sqrt(mean((Y(:,1)-yhat).^2));P2all1(i,1)=p(2);
SDEall1(i,1)=sqrt(sum((Y(:,1)-yhat).^2)/c);
p = polyfit(X(:,i),Y(:,2),1); yhat=p(2)+p(1).*X(:,i); Pall2(i,1)=p(1);
RMSEall2(i,1)=sqrt(mean((Y(:,2)-yhat).^2));P2all2(i,1)=p(2);
SDEall2(i,1)=sqrt(sum((Y(:,2)-yhat).^2)/c);
p = polyfit(X(:,i), Y(:,3),1); yhat=p(2)+p(1).*X(:,i); Pall3(i,1)=p(1);
RMSEall3(i,1)=sqrt(mean((Y(:,3)-yhat).^2));P2all3(i,1)=p(2);
SDEall3(i,1)=sqrt(sum((Y(:,3)-yhat).^2)/c);
end

rmse_filt= [RMSEall1, RMSEall2 RMSEall3];
sde_filt=[SDEall1, SDEall2, SDEall3];
slope_filt=[Pall1, Pall2, Pall3];
intcpt_filt=[P2all1, P2all2, P2all3];

%% Calculate precisions and accuracy parameters on first derivative spectrum
```

```
ll=727; hl=1089;
X=vis_nir_spectra(ll:hl,:)';

for i=1:1:363
p = polyfit(X(:,i),Y(:,1),1); yhat=p(2)+p(1).*X(:,i); Pall1(i,1)=p(1);
RMSEall1(i,1)=sqrt(mean((Y(:,1)-yhat).^2));P2all1(i,1)=p(2);
SDEall1(i,1)=sqrt(sum((Y(:,1)-yhat).^2)/c);
p = polyfit(X(:,i),Y(:,2),1); yhat=p(2)+p(1).*X(:,i); Pall2(i,1)=p(1);
RMSEall2(i,1)=sqrt(mean((Y(:,2)-yhat).^2));P2all2(i,1)=p(2);
SDEall2(i,1)=sqrt(sum((Y(:,2)-yhat).^2)/c);
p = polyfit(X(:,i), Y(:,3),1); yhat=p(2)+p(1).*X(:,i); Pall3(i,1)=p(1);
RMSEall3(i,1)=sqrt(mean((Y(:,3)-yhat).^2));P2all3(i,1)=p(2);
SDEall3(i,1)=sqrt(sum((Y(:,3)-yhat).^2)/c);
end

rmse_fd= [RMSEall1, RMSEall2 RMSEall3];
sde_fd=[SDEall1, SDEall2, SDEall3];
slope_fd=[Pall1, Pall2, Pall3];
intcpt_fd=[P2all1, P2all2, P2all3];

%% Calculate precisions and accuracy parameters on second derivative spectra
ll=1090; hl=1452;
X=vis_nir_spectra(ll:hl,:)';

for i=1:1:363
p = polyfit(X(:,i),Y(:,1),1); yhat=p(2)+p(1).*X(:,i); Pall1(i,1)=p(1);
RMSEall1(i,1)=sqrt(mean((Y(:,1)-yhat).^2));P2all1(i,1)=p(2);
SDEall1(i,1)=sqrt(sum((Y(:,1)-yhat).^2)/c);
p = polyfit(X(:,i),Y(:,2),1); yhat=p(2)+p(1).*X(:,i); Pall2(i,1)=p(1);
RMSEall2(i,1)=sqrt(mean((Y(:,2)-yhat).^2));P2all2(i,1)=p(2);
SDEall2(i,1)=sqrt(sum((Y(:,2)-yhat).^2)/c);
p = polyfit(X(:,i), Y(:,3),1); yhat=p(2)+p(1).*X(:,i); Pall3(i,1)=p(1);
RMSEall3(i,1)=sqrt(mean((Y(:,3)-yhat).^2));P2all3(i,1)=p(2);
SDEall3(i,1)=sqrt(sum((Y(:,3)-yhat).^2)/c);
end

rmse_sd= [RMSEall1, RMSEall2 RMSEall3];
sde_sd=[SDEall1, SDEall2, SDEall3];
slope_sd=[Pall1, Pall2, Pall3];
intcpt_sd=[P2all1, P2all2, P2all3];


figure(6)
plot(abs(slope_raw(:,1)).*RMSE(1:363,:),rmse_raw(:,1), 'or'); hold on
plot(abs(slope_filt(:,1)).*RMSE(364:726,:),rmse_filt(:,1), 'xb'); hold on
plot(abs(slope_fd(:,1)).*RMSE(727:1089,:),rmse_fd(:,1), '*k'); hold on
```

```
plot(abs(slope_sd(:,1)).*RMSE(1090:1452,:),rmse_sd(:,1), '+m'); hold on
zz=std(Y(:,1)); plot((0:2:10), [zz,zz,zz,zz,zz,zz], 'linewidth',3); grid on
xlabel('Measurement error, %sand', 'FontSize',12)
ylabel('Prediction error, %sand ', 'FontSize',12)
set(gca,'XTick',0:1:20); set(gca,'YTick',0.0:2:20);
plot([14,0], [0,14], '.-.');
plot([15,0], [0,15], '.-.'); plot([16,0], [0,16], '.-.')
plot([17,0], [0,17], '.-.'); plot([18,0], [0,18], '.-.')
plot([19,0], [0,19], '.-.'); plot([20,0], [0,20], '.-.')
plot([21,0], [0,21], '.-.');plot([22,0], [0,22], '.-.')
plot([23,0], [0,23], '.-.'); plot([24,0], [0,24], '.-.');
plot(3.28, 5.74, 'db', 'markersize',12)
plot(3.68, 5.78, 'ob', 'markersize',12)
plot(3.12, 6.85, 'xb', 'markersize',12)
plot(3.74, 3.25, '*b', 'markersize',12)
plot(2.96, 3.78, '+b', 'markersize',12)
xlim([0 10]); ylim([0 20])
legend('Original', 'Smooth', 'First-Derivative', 'Second-Derivative', ...
 'Location','SouthEast')

figure(7)
plot(abs(slope_raw(:,2)).*RMSE(1:363,:),rmse_raw(:,2), 'or'); hold on
plot(abs(slope_filt(:,2)).*RMSE(364:726,:),rmse_filt(:,2), 'xb'); hold on
plot(abs(slope_fd(:,2)).*RMSE(727:1089,:),rmse_fd(:,2), '*k'); hold on
plot(abs(slope_sd(:,2)).*RMSE(1090:1452,:),rmse_sd(:,2), '+m'); hold on
zz=std(Y(:,2)); plot((0:5), [zz,zz,zz,zz,zz,zz], 'linewidth',3); grid on
xlabel('Measurement error, %clay', 'FontSize',12)
ylabel('Prediction error, %clay ', 'FontSize',12)
set(gca,'XTick',0:1:10); set(gca,'YTick',0:3:20);
plot([6,0], [0,6], '.-.'); plot([7,0], [0,7], '.-.')
plot([8,0], [0,8], '.-.'); plot([9,0], [0,9], '.-.')
plot([10,0], [0,10], '.-.'); plot([11,0], [0,11], '.-.')
plot([12,0], [0,12], '.-.'); plot([5,0], [0,5], '.-.');
plot([13,0], [0,13], '.-.'); plot([4,0], [0,4], '.-.');
plot(1.41, 3.25, 'db', 'markersize',12)
plot(1.50, 2.51, 'ob', 'markersize',12)
plot(1.23, 2.34, 'xb', 'markersize',12)
plot(0.64, 0.73, '*b', 'markersize',12)
plot(0.63, 0.84, '+b', 'markersize',12)
xlim([0 5]); ylim([0 12])
legend('Original', 'Smooth', 'First-Derivative', 'Second-Derivative', ...
 'Location','SouthEast')


figure(8)
plot(abs(slope_raw(:,3)).*RMSE(1:363,:),rmse_raw(:,3), 'or'); hold on
```

```
plot(abs(slope_filt(:,3)).*RMSE(364:726,:),rmse_filt(:,3), 'xb'); hold on
plot(abs(slope_fd(:,3)).*RMSE(727:1089,:),rmse_fd(:,3), '*k'); hold on
plot(abs(slope_sd(:,3)).*RMSE(1090:1452,:),rmse_sd(:,3), '+m'); hold on
zz=std(Y(:,3)); plot((0:4), [zz,zz,zz,zz,zz], 'linewidth',3); grid on
xlabel('Measurement error, %carbon', 'FontSize',12)
ylabel('Prediction error, %carbon', 'FontSize',12)
set(gca,'XTick',0:1:10); set(gca,'YTick',0:3:20);
plot([4,0], [0,4], '.-.');plot([4.5,0], [0,4.5], '.-.')
plot([5,0], [0,5], '.-.') ;plot([5.5,0], [0,5.5], '.-.');
plot([6,0], [0,6], '.-.');plot([6.5,0], [0,6.5], '.-.');
plot(1.20, 1.58, 'db', 'markersize',12)
plot(1.29, 2.24, 'ob', 'markersize',12)
plot(0.99, 1.46, 'xb', 'markersize',12)
plot(1.06, 1.48, '*b', 'markersize',12)
plot(0.99, 1.09, '+b', 'markersize',12)
xlim([0 3]); ylim([0 7])
legend('Original', 'Smooth', 'First-Derivative', 'Second-Derivative', ...
 'Location','SouthEast')

figure(9)
tt= mean(avg1(1:363,:),2);
tt1=std(avg1(1:363,:)')';
plot(wavelengths, tt+tt1, '--g', 'markersize',5); hold on
plot(wavelengths, tt, '-k', 'markersize',5); hold on
plot(wavelengths, tt-tt1, '--g', 'markersize',5); hold on
text(1413, tt(180),' Sand (best) \rightarrow','HorizontalAlignment','right')
text(1940,tt(296),'Sand, Clay (best) \rightarrow', 'HorizontalAlignment','right')
text(1919,tt(291),'\leftarrow Clay ', 'HorizontalAlignment','left')
text(556,tt(25),'\leftarrow Carbon ', 'HorizontalAlignment','left')
text(674,tt(46),'\leftarrow Carbon (best) ', 'HorizontalAlignment','left')
text(585,tt(30),'Carbon \rightarrow', 'HorizontalAlignment','right')
set(gca,'XTick',400:300:2200); set(gca,'YTick',0:0.1:0.6);
xlabel('Wavelengths, nm', 'FontSize',12)
ylabel('Reflectance ', 'FontSize',12)
xlim([400 2200])
ylim([0.0 0.52])
grid on
```

D-3 Function to convert GPS co-ordinates into metric units (Chapter 8)

```
function METRIC = GPS_TO_METER(DATA)

[n,p] = size(DATA); t = 1:n;

%% Assignment of variables
X = DATA(:,1); % LON
Y = DATA(:,2); % LAT
EC_SH = DATA(:,3); % EC-Shallow
EC_DP = DATA(:,4); % EC-Deep
Z = DATA(:,5); % Elevation

%% Paramters
a=6378137; b=6356752.3142;
f=0.003353; N=6388838;
M=6367382; W=0.998325;
phi=mean(Y); h=mean(Z);

%% Equations
F_long=(N+h)*cos(phi*pi/180)*pi/180; %,using stand alone
F_lat=(M+h)*pi/180; %,using stand alone

%% Distance formula
%sqrt((F_long*(x(i)-x(i-1))^2+(F_lat*(y(i)-y(i-1)))^2));

MIN_X=min(X);
MIN_Y=min(Y);
MIN_Z=min(Z);

for i=1:1:length(X)
 X(i)=F_long*(X(i)-MIN_X);
 Y(i)=F_lat*(Y(i)-MIN_Y);
 Z(i)=Z(i)-MIN_Z;
end

%% Return coverted variables to workspace
METRIC = [X Y EC_SH EC_DP Z];

%% Save workspace
save Metric
```

D-4 Function to average spatial data on grid basis (Chapter 8)

```
function NSA_DATA=GRID_AVG();

%% Load workspace variables returned by the previous funtion
load METRIC;

%% Specify the grid and offset parameters
gdsz=20;
hgd=gdsz/2;
lon= METRIC(:,1);
lat=METRIC(:,2);
londx= 0:gdsz:round(max(METRIC(:,1)));
latdx =0:gdsz:round(max(METRIC(:,2)));

%% Assign new variables
Y1= METRIC(:,1);
Y2= METRIC(:,2);
Y3= METRIC(:,3);
Y4= METRIC(:,4);
Y5= METRIC(:,5);

%% Code using loop to average data points withing each grid
for i=1:length(latdx)
 for j=1:length(londx)
  ind = find(lat<=latdx(i)+hgd & lat>latdx(i)-hgd & lon<=londx(j)+hgd & lon>londx(j)-hgd);
  avg_data1(i,j) = londx(j)+hgd;
  avg_data2(i,j) = latdx(i)+hgd;
  avg_data3(i,j) = nanmean(Y3(ind));
  avg_data4(i,j) = nanmean(Y4(ind));
  avg_data5(i,j) = nanmean(Y5(ind));
 end
end

%% Return the grid averaged variables to workspace
NSA_DATA= [avg_data1 avg_data2 avg_data3 avg_data4 avg_data5];


%% Save workspace
save NSA_DATA
```

D-5 Function to calculate the mean squared error on input data layers (Chapter 8)

```
function MSE = NSA_ERR(DATA,ZONE),

%% Identify parameters and assign values according to Zones
K=max(ZONE(:)); N=length(ZONE(:)) - length(ZONE(find(ZONE==0))); %N > K

if N>K,

WB=DATA; % Create a temporary matrice to perform rough work
for i=1:K,
  % Calculate and insert average of i zone values
 WB(find(ZONE==i))= mean(DATA(find(ZONE==i)));
end

%% Calculate the Mean Squared Error term
MSE=(sum(sum((DATA-WB).^2)))/(N-K); %return MSE value

else

"display error"

end
```

D-6 Function to display the output figures (Chapter 8)

```
function NSA_FIGURE()

%% Get and set settings related to screen resolution
set(0,'Units','normalized');
get(0,'ScreenSize');
set(0,'Units','pixels') ;
scnsize = get(0,'ScreenSize');

%% Get the first figure
fig1 = figure(1);

%% Perform calculations to set initial reference positions
position = get(fig1,'Position');
outerpos = get(fig1,'OuterPosition');
borders = outerpos - position;

%% Get the remaining figures
fig2 = figure(2);
fig3 = figure(3);

%% Assign thier position values with respect to screen size
edge = borders(1)/2;
pos1 = [edge,0,scnsize(3)/2 - edge,scnsize(4)];
pos2 = [scnsize(3)/2 + edge,pos1(4)/2,pos1(3),pos1(4)/2-edge];
pos3 = [scnsize(3)/2 + edge,0, pos1(3),pos1(4)/2-edge];

%% Apply settings to the presently displayed output
set(fig1,'OuterPosition',pos1)
set(fig2,'OuterPosition',pos2)
set(fig3,'OuterPosition',pos3)
%set(fig4,'OuterPosition',pos4)
```

D-7 Function to check if a chosen element has at least one immediate neighbour of present group (Chapter 8)

```
function NBR_VALID = NSA_NBR_VALID(IND, ZONE)

[rr cc]=size(ZONE);
S=[rr,cc]; I=1;

if (ZONE(IND-1)== max(ZONE(:))||ZONE(IND+1)== max(ZONE(:))|| ...
 ZONE(IND-rr)== max(ZONE(:))||ZONE(IND+rr)== max(ZONE(:)))

NBR_VALID=1;

else
 NBR_VALID=0;
end
```

D-8 Function to check if a element has 4 valid but immediate neighbours (Chapter 8)

```
function NHOOD_VALID = NSA_NHOOD_VALID(ZWINDOW),

NHOOD_VALID=prod(ZWINDOW(:));

end
```

D-9 Function to process multidimensional spatial data using NSA(Chapter 8)

```
function [DATA FZONE NSA_MSE NSA_MSE_dEC1 NSA_MSE_sEC1 NSA_MSE_ELEV]
= CLASSIFY()

%% Load the previously processed and saved data in to workspace
load NSA_DATA

%% Define some variables
a= 1; b=1; e=4; zzz=1; j=0;

%% Assign weights to enable and disable the spatial variables
Z1MSE = NSA_ERR(DATA,ZONE1)^a *NSA_ERR(DATA1,ZONE1)^b
*(NSA_ERR(ELEV,ZONE1))^e;

%% Some code to calculate the initial error and the performance indices
NSA_MSE(1)=Z1MSE; zzz=zzz+1;
NSA_MSE_dEC1(1)=NSA_ERR(DATA1,ZONE1); % Individual MSE
NSA_MSE_sEC1(1)=NSA_ERR(DATA,ZONE1);% Individual MSE
NSA_MSE_ELEV(1)=NSA_ERR(ELEV,ZONE1); % Individual MSE

[rr cc]=size(DATA); S=[rr,cc]; I=1; ZONE=ZONE1; K=2; GG=0;
FZMSE=Z1MSE; BB=0; ZGROWTH=0;GET_ZONE_VAL=0;

PRODR2= (1- NSA_MSE_dEC1(end)/NSA_MSE_dEC1(1))^b * ...
 (1- NSA_MSE_sEC1(end)/NSA_MSE_sEC1(1))^a * ...
(1- NSA_MSE_ELEV(end)/NSA_MSE_ELEV(1))^e;

%% Display histogram of the enabled input data layers
figure(4)
subplot(3,1,1);hist(ELEV(find(ELEV~=0)))
subplot(3,1,2);hist(DATA1(find(DATA1~=0)))
subplot(3,1,3);hist(DATA(find(DATA~=0)))
waitforbuttonpress

%% Display the input layers, error/performance and the output figures
figure(1)
view(2)
subplot (5,1,1); surf(LON,LAT,ELEV)
axis([min(LON(:)) max(LON(:)) min(LAT(:)) ...
 max(LAT(:)) min(ELEV(:)) max(ELEV(:))])
colorbar
view([0,90])
NSA_FIGURE()
```

```
figure(1)
subplot (5,1,2); surf(LON,LAT,DATA)
axis([min(LON(:)) max(LON(:)) min(LAT(:)) ...
 max(LAT(:)) min(DATA(:)) max(DATA(:))])
view([0,90])
colorbar
NSA_FIGURE()

figure(1)
subplot (5,1,3); surf(LON,LAT,DATA1)
axis([min(LON(:)) max(LON(:)) min(LAT(:)) ...
 max(LAT(:)) min(DATA1(:)) max(DATA1(:))])
view([0,90])
colorbar
NSA_FIGURE()

figure(3)
plot(1,0,'--rs','LineWidth',2,...
    'MarkerEdgeColor','b',...
    'MarkerFaceColor','b',...
    'MarkerSize',10);
   axis([1 2 0 1])
grid on; hold on
ylabel (' Product R^2' )
xlabel( 'Number of Dynamic Classes')

 %% Start the big iteration loop

while (GG~=1),
 j=1;
 ELMT=find(ZONE<=K & ZONE~=0); % Find all points in data that are not 0
for i=1:1:length(ELMT)
 IND=ELMT(i); [row,col]=ind2sub(S,IND);

 ZWINDOW=ZONE(row-I:row+I,col-I:col+I);
  NHOOD_VALID = NSA_NHOOD_VALID(ZWINDOW); % Check for valid elements,
          % if valid element then 1 else 0

 if NHOOD_VALID > 0,
  ZONE(row-I:row+I,col-I:col+I)=K;
  MSE = NSA_ERR(DATA,ZONE)^a *NSA_ERR(DATA1,ZONE)^b
*(NSA_ERR(ELEV,ZONE))^e;
  E(i,:)=[IND MSE];
  ZONE(row-I:row+I,col-I:col+I)= ZWINDOW;
 end
end %% Code to look for cluster centers.
```

```
CAND=sortrows(E,2);clear E % Sort in order
CAND((find(CAND(:,1)==0)),:)=[];
IND=CAND(1); % Pick the smallest element
[row,col]=ind2sub(S,IND);
ZWINDOW1=ZONE(row-I:row+I,col-I:col+I);
ZONE(row-I:row+I,col-I:col+I)=K;
MSE= NSA_ERR(DATA,ZONE)^a *NSA_ERR(DATA1,ZONE)^b d
*(NSA_ERR(ELEV,ZONE))^e;
if MSE >= FZMSE,
 disp([K MSE FZMSE])
 ZONE(row-I:row+I,col-I:col+I)= ZWINDOW1;
 break;
else
 disp('Opened a new uniform zone consisting of at least 9 cells')
 K
disp('ZONE NEWMSE OLDMSE')
disp([K MSE FZMSE])
end

% The above part of this code is to search for a new uniform zone within
%  The area of the present zone and to establish its conditional base

%% The next part of the code is to find all the neighbourhood indices of the new uniform zone

%% Start the smaller iteration loop
ZGROW =1;
while (ZGROW==1),
ZGROWTH=0;
TEMP_IND=find(ZONE~=K & ZONE~=0);
STR_NBRS=0;
for i=1:1:length(TEMP_IND)
 IND=TEMP_IND(i);
 NBR_VALID = NSA_NBR_VALID(IND, ZONE);
 if NBR_VALID ==1,
  STR_NBRS(i)=IND;
 end
  STR_NBRS(STR_NBRS==0)=[]; % store potential neighbors

end

% The next part of the code is to check the value of stored neighbor
% Indices and to decide the strategy to grow the cluster

%% Now its time to define what are the strategies
for i=1:1:length(STR_NBRS)
```

```
 if ZONE(STR_NBRS(i))==1,
 GS1MSE= NSA_ERR(DATA,ZONE)^a *NSA_ERR(DATA1,ZONE)^b
*(NSA_ERR(ELEV,ZONE))^e; %Error before zone growth
 ZONE(STR_NBRS(i))=K;
 GS1E= NSA_ERR(DATA,ZONE)^a *NSA_ERR(DATA1,ZONE)^b
NSA_ERR(ELEV,ZONE))^e; %Error after zone growth

 if GS1E > GS1MSE
 ZONE(STR_NBRS(i))=1;
 FZMSE=GS1MSE;
 else
  FZMSE=GS1E;
  ZGROWTH=ZGROWTH+1;
 end

 elseif ZONE(STR_NBRS(i))== K-1 & ZONE(STR_NBRS(i))~=1,
  GS2MSE= NSA_ERR(DATA,ZONE)^a *NSA_ERR(DATA1,ZONE)^b
*(NSA_ERR(ELEV,ZONE))^e; %Error before zone growth
  TEMPKS2=find(ZONE==K);
  ZONE(find(ZONE==K))=K-1;
  GS2E= NSA_ERR(DATA,ZONE)^a *NSA_ERR(DATA1,ZONE)^b
*(NSA_ERR(ELEV,ZONE))^e;

%Error after zone growth
  if GS2E > GS2MSE,
   ZONE(TEMPKS2)= K;
   FZMSE=GS2MSE;
  else
   FZMSE=GS2E;
   ZGROWTH=ZGROWTH+1;
   K=K-1;
   break,
  end

 elseif ZONE(STR_NBRS(i))<K-1 & ZONE(STR_NBRS(i))~=1,
  GS3MSE=NSA_ERR(DATA,ZONE)^a *NSA_ERR(DATA1,ZONE)^b
*(NSA_ERR(ELEV,ZONE))^e; %Error before zone growth
  TEMPK=ZONE(STR_NBRS(i));
  TEMPKS3= find (ZONE == K);
  ZONE(find(ZONE==K))=TEMPK;
  GS3E= NSA_ERR(DATA,ZONE)^a *NSA_ERR(DATA1,ZONE)^b
*(NSA_ERR(ELEV,ZONE))^e; %Error after zone growth
  if GS3E > GS3MSE,
   ZONE(TEMPKS3)= K;
   FZMSE=GS3MSE;
```

```
 else
  FZMSE=GS3E;
  ZGROWTH=ZGROWTH+1;
  K=K-1;
  break
 end
 end

 if i==length(STR_NBRS),
    if ZGROWTH ==0,
   if length(ZONE(find(ZONE==K)))<10,
   ZONE(find(ZONE==K))=max(ZWINDOW1(:));
   BB=1; K=K-1;
   end
  ZGROW=0;
  end
 end

 %% Revise the zone map
figure(2)
surf(LON,LAT,ZONE)
axis([min(LON(:)) max(LON(:)) min(LAT(:)) ...
 max(LAT(:)) min(ZONE(:)) max(ZONE(:))])
view([0,90])
NSA_FIGURE()
end
% End

%% Calculate the error and the performance parameters
NSA_MSE(zzz)=FZMSE;
NSA_MSE_dEC1(zzz)=NSA_ERR(DATA1,ZONE);
NSA_MSE_sEC1(zzz)=NSA_ERR(DATA,ZONE);
NSA_MSE_ELEV(zzz)=NSA_ERR(ELEV,ZONE);

PRODR2= (1- NSA_MSE_dEC1(end)/NSA_MSE_dEC1(1))^b * ...
 (1- NSA_MSE_sEC1(end)/NSA_MSE_sEC1(1))^a * ...
(1- NSA_MSE_ELEV(end)/NSA_MSE_ELEV(1))^e;
 zzz=zzz+1;
 end

figure(3)

if BB==1,
 BB=0;
 GG=1;
else
```

```
%% Display error and performance parameters from the present time
figure(3)
plot(K,PRODR2,'--rs','LineWidth',2,...
   'MarkerEdgeColor','k',...
   'MarkerFaceColor','g',...
   'MarkerSize',10);
plot(K, (1- NSA_MSE_dEC1(end)/NSA_MSE_dEC1(1)), ...
   '--rs','LineWidth',2,...
   'MarkerEdgeColor','k',...
   'MarkerFaceColor','r',...
   'MarkerSize',10);
plot(K, (1- NSA_MSE_sEC1(end)/NSA_MSE_sEC1(1)), ...
   '--rs','LineWidth',2,...
   'MarkerEdgeColor','k',...
   'MarkerFaceColor','c',...
   'MarkerSize',10);
plot(K, (1- NSA_MSE_ELEV(end)/NSA_MSE_ELEV(1)), ...
   '--rs','LineWidth',2,...
   'MarkerEdgeColor','k',...
   'MarkerFaceColor','b',...
   'MarkerSize',10);
axis([1 K 0 1])
grid on; hold on
ylabel (' Product R^2' )
xlabel( 'Number of Dynamic Classes')

K=K+1;

end
end

%% Revise the display of error and the performance parameters
figure(3)
plot(K,PRODR2,'--rs','LineWidth',2,...
   'MarkerEdgeColor','k',...
   'MarkerFaceColor','g',...
   'MarkerSize',10);
plot(K, (1- NSA_MSE_dEC1(end)/NSA_MSE_dEC1(1)), ...
   '--rs','LineWidth',2,...
   'MarkerEdgeColor','k',...
   'MarkerFaceColor','r',...
   'MarkerSize',10);
plot(K, (1- NSA_MSE_sEC1(end)/NSA_MSE_sEC1(1)), ...
   '--rs','LineWidth',2,...
   'MarkerEdgeColor','k',...
```

```
    'MarkerFaceColor','c',...
    'MarkerSize',10);
plot(K, (1- NSA_MSE_ELEV(end)/NSA_MSE_ELEV(1)), ...
    '--rs','LineWidth',2,...
    'MarkerEdgeColor','k',...
    'MarkerFaceColor','b',...
    'MarkerSize',10);
axis([1 K 0 1])
grid on; hold on
ylabel (' Product R^2' )
xlabel( 'Number of Dynamic Classes')
legend('Elevation','Product','DeepECa1','ShallowECa1')

%% Return the output variables to workspace
FZONE=ZONE; % The groups matrix
NSA_MSE; % The over all error vector
NSA_MSE_dEC1 % The error vector for first input layer
NSA_MSE_sEC1 % The error vector for second input layer
NSA_MSE_ELEV % The error vector for fifth input layer
```