

A frequency response-based dielectric sensor for real-time sensing and long-term monitoring of  
soil water content, nitrogen, and salinity

by

Mohammed Mezher Hasan

B.S., University of Baghdad, Iraq, 2001

M.S., University of Baghdad, Iraq, 2005

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Carl and Melinda Helwig Department of Biological & Agricultural Engineering  
Carl R. Ice College of Engineering

KANSAS STATE UNIVERSITY  
Manhattan, Kansas

2020

## **Abstract**

Practical sensors capable of in-field, simultaneous, real-time sensing and long-term monitoring of several soil properties that are important to production agriculture and environment, including water content, fertility, and salinity, are not available. Worldwide research has concentrated on optical (spectral), electrochemical, and dielectric types of sensors that measure only one or two properties. In the Instrumentation and Control Laboratory of the Department of Biological and Agricultural Engineering at Kansas State University, a frequency response-based dielectric sensor (“FR sensor”) was developed in 2004 and it has been used to measure multiple properties of various types of dielectric materials, including soil, water, air, and fuel. The early work showed that the sensor is multifunctional, inexpensive, and rugged for field applications. However, the potential of the sensor for use in the field for simultaneous measurements of soil water and nutrients is unknown.

This dissertation reports work since 2017, when a new sensor probe was designed specifically for “buried-in-soil” applications. Tests showed that the sensor could conduct simultaneous measurements of soil water content, density, and salinity. When salinity was replaced with a nitrate-N fertilizer, an ammonium-N fertilizer, or an organic N fertilizer, the sensor reported good results. The sensor, in general, had difficulties in separately measuring two types of the nitrogen fertilizer when both existed in soil.

To simulate the in-field applications, corn was planted in two large pots. Two sensor probes were placed in each of them, and the pots were placed outdoors for more than three months. Dielectric spectral data from the sensors were measured every other day, and, at the same time, soil samples were taken for analysis of water content, nitrate-N, and ammonium-N using standard instrument and at the KSU Soil Testing Laboratory to provide reliable references.

Computer programs were written to analyze the data and to develop prediction models using the partial least squares regression method. Prediction results were compared with the reliable references. For water content and nitrate-N, good results were obtained using data from single sensors as well as data from across multiple sensors. For ammonium-N, good results were obtained using single-sensor data, but the results from across-sensor data were not satisfactory.

From the partial least squares regression analyses, a limited number of frequencies - the “signature frequencies” – were identified for each soil properties. With the reduced number of frequencies, the cost and measurement time can be reduced, while the measurement accuracy can be improved.

In the outdoor experiment, one fifth of the data were lost due to broken wires. Thus, improvements in the design of the probe, control box, and wire connections are needed to insure longevity and reliability of the sensor.

A frequency response-based dielectric sensor for real-time sensing and long-term monitoring of  
soil water content, nitrogen, and salinity

by

Mohammed Mezher Hasan

B.S., University of Baghdad, Iraq, 2001

M.S., University of Baghdad, Iraq, 2005

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Carl and Melinda Helwig Department of Biological & Agricultural Engineering  
Carl R. Ice College of Engineering

KANSAS STATE UNIVERSITY  
Manhattan, Kansas

2020

Approved by:

Major Professor  
Dr. Naiqian Zhang

## **Abstract**

Practical sensors capable of in-field, simultaneous, real-time sensing and long-term monitoring of several soil properties that are important to production agriculture and environment, including water content, fertility, and salinity, are not available. Worldwide research has concentrated on optical (spectral), electrochemical, and dielectric types of sensors that measure only one or two properties. In the Instrumentation and Control Laboratory of the Department of Biological and Agricultural Engineering at Kansas State University, a frequency response-based dielectric sensor (“FR sensor”) was developed in 2004 and it has been used to measure multiple properties of various types of dielectric materials, including soil, water, air, and fuel. The early work showed that the sensor is multifunctional, inexpensive, and rugged for field applications. However, the potential of the sensor for use in the field for simultaneous measurements of soil water and nutrients is unknown.

This dissertation reports work since 2017, when a new sensor probe was designed specifically for “buried-in-soil” applications. Tests showed that the sensor could conduct simultaneous measurements of soil water content, density, and salinity. When salinity was replaced with a nitrate-N fertilizer, an ammonium-N fertilizer, or an organic N fertilizer, the sensor reported good results. The sensor, in general, had difficulties in separately measuring two types of the nitrogen fertilizer when both existed in soil.

To simulate the in-field applications, corn was planted in two large pots. Two sensor probes were placed in each of them, and the pots were placed outdoors for more than three months. Dielectric spectral data from the sensors were measured every other day, and, at the same time, soil samples were taken for analysis of water content, nitrate-N, and ammonium-N using standard instrument and at the KSU Soil Testing Laboratory to provide reliable references.

Computer programs were written to analyze the data and to develop prediction models using the partial least squares regression method. Prediction results were compared with the reliable references. For water content and nitrate-N, good results were obtained using data from single sensors as well as data from across multiple sensors. For ammonium-N, good results were obtained using single-sensor data, but the results from across-sensor data were not satisfactory.

From the partial least squares regression analyses, a limited number of frequencies - the “signature frequencies” – were identified for each soil properties. With the reduced number of frequencies, the cost and measurement time can be reduced, while the measurement accuracy can be improved.

In the outdoor experiment, one fifth of the data were lost due to broken wires. Thus, improvements in the design of the probe, control box, and wire connections are needed to insure longevity and reliability of the sensor.

# Table of Contents

List of Figures .....	x
List of Tables .....	xiii
Acknowledgements .....	xv
Chapter 1 - Introduction and Objective .....	1
1.1 Needs for in-situ soil health sensor .....	2
1.2 Soil health properties to be measured .....	3
1.2.1 Soil fertility .....	4
1.2.2 Soil moisture .....	5
1.2.3 Soil salinity .....	7
1.3 Current soil fertility sensing technologies and our previous work .....	9
1.3.1 Optical sensors .....	10
1.3.2 Electrochemical sensors .....	10
1.3.3 Dielectric sensors .....	11
1.4 Permittivity .....	14
1.5 Our previous work on dielectric sensors .....	16
1.6 Need for sensors buried in soil .....	17
1.7 Partial least squares regression .....	18
1.8 Objectives .....	20
Chapter 2 - Materials and Methods .....	21
2.1 Basic measurement principle - the frequency-response method .....	21
2.2 The control box .....	23
2.3 Probes .....	24
2.4 Experimental procedure .....	25
2.4.1 Laboratory experiments .....	25
2.4.1.1 Test materials .....	26
2.4.1.2 Test apparatus .....	26
2.4.1.3 Sample preparation .....	28
2.4.1.4 Preliminary observation of the dielectric spectral data .....	29
2.4.1.4.1 Soil density .....	29

2.4.1.4.2 Soil moisture .....	30
2.4.1.4.3 Soil salinity .....	31
2.4.1.4.4 Soil fertility (single nitrogen fertilizer), VWC, and density. ....	32
2.4.1.4.5 Soil fertility (combined two nitrogen fertilizers), moisture, and density.....	34
2.4.2 Outdoor tests .....	35
2.4.2.1 Test apparatus .....	35
2.4.2.1.1 Pots.....	35
2.4.2.1.2 Probes.....	37
2.4.2.1.3 Other instruments.....	37
2.4.2.2 Sample preparation .....	38
2.4.2.3 Sampling schedule .....	40
2.4.2.4 Removal of soil from the pots.....	40
2.4.2.4.1 Conditions of the plants upon removal from the pots.....	41
2.4.2.4.2 Conditions of the sensors upon removal from the pots.....	41
2.4.2.5 Conditions of the sensor signals due to potential problems caused by broken connections .....	43
Sensor 1.....	43
Sensor 2.....	44
Sensor 3.....	44
Sensor 4.....	45
Chapter 3 - Results and Discussion .....	46
3.1 Laboratory experiment.....	46
3.1.1 Samples and PLS analysis of dielectric spectral data .....	47
3.1.2 Results of the PLS regression analysis .....	50
3.1.2.1 Soil salinity .....	50
3.1.2.2 Effect of bulk density.....	50
3.1.2.3 One type of N-fertilizer.....	52
3.1.2.3.1 Ammonium sulfate.....	52
3.1.2.3.2 Effect of bulk density.....	52
3.1.2.3.3 Ammonium nitrate .....	53
3.1.2.3.4 Effect of bulk density.....	54



3.1.2.3.5 Urea fertilizer .....	54
3.1.2.3.6 Effect of bulk density .....	55
3.1.2.4 Two types of N-fertilizers .....	56
3.1.2.4.1 Ammonium sulfate and ammonium nitrate .....	56
3.1.2.4.2 Ammonium sulfate and urea .....	57
3.1.2.4.3 Ammonium nitrate and urea .....	57
3.2 Outdoor experiment .....	58
3.2.1 Sample groups .....	58
3.2.2 Sample range .....	59
3.2.3 Coverage Index .....	60
3.2.4 Matlab programs for data preparation and PLS regression analysis .....	65
3.2.5 Prediction results .....	68
3.2.5.1 Models for individual sensors .....	68
3.2.5.2 Models across multiple sensors .....	77
“Universal” models for measuring VWC .....	78
“Universal” models for measuring NO <sub>3</sub> .....	79
“Universal” models for measuring NH <sub>4</sub> ” .....	81
3.2.6 Selection of signature frequencies .....	84
Chapter 4 - Conclusions and Recommendations .....	87
4.1 Conclusions .....	87
4.2 Recommendations for further studies .....	89
References .....	90
Appendix A - Soil Texture .....	96
Appendix B - Matlab Program for preliminary observation .....	97
Appendix C - Outdoor test planter .....	99
Appendix D - Datalogger code .....	100
Appendix E - Matlab program for the laboratory tests .....	102
Appendix F - Matlab program for data preparation and preprocessing for the outdoor tests .....	106
Appendix G - Matlab program for PLS regression for the outdoor tests .....	116
Appendix H - VWC, NO <sub>3</sub> , and NH <sub>4</sub> values for outdoor test samples .....	125

## List of Figures

Figure 2.1 Measurement principle of the FR sensor .....	21
Figure 2.2 Simulation vs. measured frequency responses (a) gain, (b) phase .....	22
Figure 2.3 The control box.....	24
Figure 2.4 Dimensions of the probe.....	24
Figure 2.5 The probe and the coaxial cable for laboratory tests .....	25
Figure 2.6 The laboratory tests container .....	26
Figure 2.7 The camera stand, piston, and container .....	27
Figure 2.8 (a) Gain and (b) Phase frequency responses at different density levels .....	29
Figure 2.9 (a) Gain and (b) Phase frequency responses at different GWC levels .....	30
Figure 2.10 Conductivity meter .....	31
Figure 2.11(a) Gain and (b) Phase frequency responses to different soil salt levels .....	32
Figure 2.12(a) Gain and (b) Phase frequency responses to different ammonium sulfate levels ..	33
Figure 2.13 (a) Gain and (b) Phase frequency responses to different ammonium nitrate levels .....	33
Figure 2.14 (a) Gain and (b) Phase frequency responses to different urea levels .....	33
Figure 2.15 (a) Gain and (b) Phase frequency responses at different ammonium sulfate & urea levels .....	34
Figure 2.16 (a) Gain and (b) Phase frequency responses with different ammonium sulfate & ammonium nitrate levels.....	34
Figure 2.17 (a) Gain and (b) Phase frequency responses with different ammonium nitrate levels & urea.....	35
Figure 2.18 Dimensions of the pot.....	36
Figure 2.19 The outdoor tests pots (a) the window (b) the holes .....	37
Figure 2.20 The outdoor test probe and coaxial cable .....	37
Figure 2.21 Instruments used in the outdoor experiment (a) humidity and temperature meter and light meter, (b) datalogger, and (c) Decagon 5TE soil moisture sensor .....	38
Figure 2.22 Insert sensors and seeds in soil (top view). (a) lower sensor, (b) upper sensor, and (c) seeds .....	39

Figure 2.23 The outdoor tests (a) a pot with the four germinated corn seeds at the beginning of the experiment, (b) a pot after thinning to two seedlings, and (c) the two pots with one plant each at the end of the experiment.....	39
Figure 2.24 Some roots around the sensor area .....	40
Figure 2.25 (a) Sensors before being removed from the pots. (b) Roots and porous areas between plates .....	41
Figure 2.26 Dielectric spectral data collected from sensor 1 between 2 October to 20 October (a) Gain, (b) Phase.....	44
Figure 2.27 Dielectric spectral data collected from sensor 2 between 2 October to 20 October (a) Gain, (b) Phase.....	44
Figure 2.28 Dielectric spectral data collected from sensor 3 between 2 October to 20 October (a) Gain, (b) Phase.....	45
Figure 2.29 Dielectric spectral data collected from sensor 4 between 2 October to 20 October (a) Gain, (b) Phase.....	45
Figure 3.1 Percent variance explained in predictor variables (X) and response variable (Y) vs number of PLS components.....	49
Figure 3.2 Fitted response (calibration) vs observed response for the training data sets with (a) 1, (b) 8, and (c) 15 PLS components .....	49
Figure 3.3 Fitted response (calibration) vs observed response for the validation data sets with (a) 1, (b) 8, and (c) 15 PLS components .....	49
Figure 3.4 Five cases of validation data coverage .....	61
Figure 3.5 Performance of NH <sub>4</sub> prediction model using two sets of samples for sensor 3 with an out-of-range sample (sample 8) included .....	62
Figure 3.6 Performance of NH <sub>4</sub> prediction model using two sets of samples for sensor 3 with an out-of-range sample (sample 8) removed .....	63
Figure 3.7 Performance of NH <sub>4</sub> prediction model using three sets of samples for sensor 3 with an out-of-range sample (sample 8) included .....	63
Figure 3.8 Performance of NH <sub>4</sub> prediction model using three sets of samples for sensor 3 with an out-of-range sample (sample 8) removed .....	64
Figure 3.9 Performance of NH <sub>4</sub> prediction model using four sets of samples for sensor 3 with an out-of-range sample (sample 8) included .....	64

Figure 3.10 Performance of NH <sub>4</sub> prediction model using four sets of samples for sensor 3 with an out-of-range sample (sample 8) removed.....	65
Figure 3.11 Performance of VWC prediction model using four sets of samples across four sensors (a) training data, (b) validation data .....	79
Figure 3.12 Residual values for the training and validation data for VWC by using four sets of samples across four sensors .....	79
Figure 3.13 Performance of NO <sub>3</sub> prediction model using four sets of samples across four sensors (a) training data, (b) validation data.....	80
Figure 3.14 Residual plots for training and validation data for NO <sub>3</sub> by using four sets of samples across four sensors .....	81
Figure 3.15 Performance of NH <sub>4</sub> prediction model using four sets of samples from four sensors (a) training data, (b) validation data.....	82
Figure 3.16 Performance of NH <sub>4</sub> prediction model using four sets of samples from four sensors (a) training data, (b) validation data.....	82
Figure 3.17 Performance of NH <sub>4</sub> prediction model using four sets of samples for all sets in all sensors with samples removed (a) training data, (b) validation data .....	83
Figure 3.18 Residual plot for training and validation data for NH <sub>4</sub> by using all data for all sensors after removing some out-of-range samples.....	83
Figure 3.19 Average VIP scores derived from 74 PLS regression runs for measuring VWC: (a) Gain spectra, (b) Phase spectra (Two distinguished peaks (DC and 30 MHz) in the magnitude spectra) .....	86
Figure 3.20 Average VIP scores derived from 74 PLS regression runs for measuring NO <sub>3</sub> : (a) Gain spectra, (b) Phase spectra (One distinguished peak (DC) in the magnitude spectra)...	86
Figure 3.21 Average VIP scores derived from 74 PLS regression runs for measuring NH <sub>4</sub> : (a) Gain spectra, (b) Phase spectra (One distinguished peak (27 MHz) in the phase spectra)...	86

## List of Tables

Table 2.1 Soil sample size for different density .....	29
Table 2.2. Water and soil weight at different GWC levels .....	30
Table 2.3 Amount of NaCl and soil to achieve different salinities.....	31
Table 3.1. Effect of number of PLS components on salinity measurement using the phase data	48
Table 3.2 R-squared values for predicting soil salinity, VWC, and bulk density.....	50
Table 3.3 R-squared values for VWC prediction at different bulk densities.....	51
Table 3.4 R-squared values for soil salinity prediction at different bulk densities.....	51
Table 3.5 R-squared value for VWC, ammonium sulfate, and bulk density .....	52
Table 3.6 R-squared value for ammonium sulfate prediction at different bulk densities.....	52
Table 3.7 R-squared value for VWC prediction at different bulk densities .....	53
Table 3.8 R-squared values for ammonium nitrate, bulk density, and VWC .....	53
Table 3.9 R-squared values for ammonium nitrate prediction at different bulk densities.....	54
Table 3.10 R-squared values for VWC prediction at different bulk densities.....	54
Table 3.11 R-squared value for urea, bulk density, and VWC .....	55
Table 3.12 R-squared values for urea prediction at different bulk densities .....	55
Table 3.13 R-squared values for VWC prediction at different bulk densities.....	55
Table 3.14 R-squared values for ammonium nitrate + ammonium sulfate, bulk density, and VWC .....	56
Table 3.15 R-squared values for ammonium sulfate + urea, bulk density, and VWC .....	57
Table 3.16 R-squared values for ammonium nitrate + urea, bulk density, and VWC.....	58
Table 3.17 Ranges of NO <sub>3</sub> , NH <sub>4</sub> , and VWC for each sample set.....	59
Table 3.18 The effect of Coverage Index (CI) on the performance of prediction models .....	62
Table 3.19 Performance of PLS prediction models for VWC established using various combinations of the sample data from sensor1 .....	69
Table 3.20 Performance of PLS prediction models for NO <sub>3</sub> established using various combinations of the sample data from sensor1 .....	70
Table 3.21 Performance of PLS prediction models for NH <sub>4</sub> established using various combinations of the sample data from sensor 1 .....	71

Table 3.22 Performance of PLS prediction models for VWC established using various combinations of the sample data from sensor 2 .....	72
Table 3.23 Performance of PLS prediction models for NO <sub>3</sub> established using various combinations of the sample data from sensor 2 .....	72
Table 3.24 Performance of PLS prediction models for NH <sub>4</sub> established using various combinations of the sample data from sensor 2 .....	73
Table 3.25 Performance of PLS prediction models for VWC established using various combinations of the sample data from sensor 3 .....	74
Table 3.26 Performance of PLS prediction models for NO <sub>3</sub> established using various combinations of the sample data from sensor 3 .....	74
Table 3.27 Performance of PLS prediction models for NH <sub>4</sub> established using various combinations of the sample data from sensor 3 .....	75
Table 3.28 Performance of PLS prediction models for VWC established using various combinations of the sample data from sensor 4 .....	76
Table 3.29 Performance of PLS prediction models for NO <sub>3</sub> established using various combinations of the sample data from sensor 4 .....	76
Table 3.30 Performance of PLS prediction models for NH <sub>4</sub> established using various combinations of the sample data from sensor 4 .....	77
Table 3.31 Performance of PLS prediction models for VWC established using various combinations of sensor data from all sets .....	78
Table 3.32 Performance of PLS prediction models for NO <sub>3</sub> established using various combinations of sensor data from all sets .....	80
Table 3.33 Performance of PLS prediction models for NH <sub>4</sub> established using various combinations of sensor data from all sets .....	81

## **Acknowledgements**

I would first like to express my deep and sincere appreciation for my major professor, Dr. Naiqian Zhang, for his continuous help, support, and encouragement as I have worked on my doctoral degree. I am very thankful to have learned so much from him.

My sincere thanks to Mingqiang Han for helping in the installation of the sensors and in taking measurements and in analyzing data from the outdoor tests, I greatly appreciate his help.

I would like to thank my committee members, Dr. M.B. Kirkham, Dr. Danny Rogers, Dr. Ajay Sharda, and Dr. Hongyu Wu. Dr. Kirkham also provided valuable guidance on my research and provided me seeds and soil for the field tests.

Jacob Thomas, Assistant Scientist, Agronomy Soil Testing Laboratory, deserves many thanks for providing test results for the soil.

I am very grateful to Dr. Colby Moorberg, Assistant Professor - Agronomy, for valuable guidance in planting the corn in the pots.

Thanks to Dustan Ridder, Assistant Agronomist-Agronomy, for his help in getting the soil from the field that was used in the outdoor experiment.

I would like to acknowledge Dr. Joe Harner, Department Head, Biological and Agricultural Engineering, and the BAE office staff: Jamie Boeckman, Arlene Jacobson, and Barbara Moore.

I appreciate the help of Jonathan Zeller, Research Technician, who helped to build the probes and to modify the pots in the workshop.

I would like to acknowledge the students in the Instrumentation and Control Lab for their help and friendship: Yong Wei and the Chinese visiting scholars.

Finally, I wish to express my deep gratitude to my family for their love and support. I especially want to thank my parents and my wife who have always been there when I need them.

## **Chapter 1 - Introduction and Objective**

Agricultural sustainability and world food security depend on new crop varieties, soil health, and farm management. Numerous new varieties of crops with disease-, insect-, and drought-resistant traits are available, and the best field management techniques are widely promoted. Nevertheless, the assessment of soil health using soil sensors requires further research.

Soil is composed of mineral matter, organic matter, water, and air. It has three phases: solid, liquid, and gas. The mineral matter in soil (solid phase) has different sizes, shapes, and chemical composition that affect the physical and chemical soil properties. .

Physical and chemical properties of soil are determined by many factors. Soil moisture is one of the main factors. It has an effect on plant growth directly. Soil moisture is important, because it can dissolve nutrients or chemical ions that are needed for crop growth. (Stewart and Nielsen, 1990).

Soil is a heterogeneous system whose processes and mechanisms are complex and difficult to fully understand. Many analytical techniques are used in an attempt to establish the relationship between soil physical and chemical properties, often disregarding their complexity. In addition, soil chemical extractions that alter the equilibrium between the phases may further complicate the interpretation of results (Viscarra Rossel and McBratney, 1998a).

Soil health is the continued capacity of soil to function as a vital living ecosystem that sustains plants, animals, and humans (NRCS, 2019). This capacity is the result of the status and interaction of soil biological, chemical, and physical properties. Although farmers and environmental consultants have soil chemical properties analyzed to make fertilizer recommendations and when reclaiming contaminated soils, continuous monitoring information about soil physical, chemical, and biological properties is not readily available, because most



research laboratories do not have the equipment and expertise to carry out the complex and costly analyses of all these properties. Therefore, there is a need for real-time soil sensing for timely management and long-term, continuous monitoring to understand the dynamic trends in the soil.

## **1.1 Needs for in-situ soil health sensor**

Methods currently in use to define and quantify soil health move beyond standard soil fertility testing to incorporate soil physical and biological properties. Haney (2014) and Haney et al. (2018) presented a tool for soil health evaluation, which incorporated soil physical and biological parameters to be used in conjunction with standard soil-fertility testing. The process allowed for an estimation of nutrient mineralization, as nutrients became inorganic and plant-available by microbial actions. The Haney method includes CO<sub>2</sub> respiration for an estimate of microbial activity, C:N ratio of soil, organic C, and organic N to develop a soil health score. The Cornell method, known as the Comprehensive Assessment of Soil Health (CASH), is another standard soil health tool (Fine et al., 2017; McGowen et al., 2018). The Alabama Soil Health Index consists of a routine soil test, estimated cation exchange capacity (CEC), soil organic matter (SOM) content, soil respiration, and aggregate stability (Alabama Extension, 2019). Microbial respiration is directly related to soil nutrient cycling. It can give an estimate of how much of a nutrient may be available to the crop through decomposition of soil organic matter (SOM). Therefore, SOM is the backbone of soil health, because it improves physical properties, regulates chemical properties, and enhances biological properties of the soil. Nearly 10 million soil samples are tested annually in North America, and the majority are tested for plant available N, P and K to make fertilizer recommendations and to protect the environment. Inadequate N and P application will negatively impact crop yields and farm profit, but over applying those nutrients will result in air and water quality degradation. Both N and P should be applied at the right rate, time, and place

with the right source. If not, for example, ammonium-N can be volatilized into the air and nitrate can be leached into groundwater. Excess N and P in the soil can also be transported to surface water via runoff and trigger algal blooms (Sharpley et al., 1996).

Methods to measure soil organic matter and plant available nutrients, however, involve field sample collections, laboratory analyses, and interpretation of results. It is often time consuming, costly, and does not capture the dynamic nature of the changing soil properties. An in-situ, continuous monitoring device is needed to provide site-specific and timely indications of soil health properties. Sensors that are installed at a field site and wirelessly transmit information can provide timely information for farmers to make informed decisions on soil management.

## **1.2 Soil health properties to be measured**

To better understand and assess soil health, important soil physical, chemical, and biological properties need to be measured. The Soil Health Institute, established in 2017 as an independent nonprofit organization to encourage and enhance research focusing on soil health, recently proposed 19 Tier 1 indicators of soil health, including soil organic carbon (OC), pH, bulk density, available water, macro- and micro-nutrients, and others, but it does not include any direct measurement of biological indicators (Soil Health Institute, 2019). A meta-analysis by Stewart et al. (2018) stated that most currently used measurement protocols have little consistency, and some of the indicators showed low response to management practices.

### 1.2.1 Soil fertility

Accurate measurements of soil fertility are needed for efficient agricultural production, including site-specific crop management, where fertilizer application rates are adjusted spatially based on local demands. Conventional methods of measuring soil fertilizer levels are tedious, time-consuming, and expensive.

Soil fertility management is one of the important areas in which precision agriculture has been applied. Fertilizers are one of the greatest costs involved in agricultural production (Thomasson et al., 2001). The high cost and environmental pollution caused by excess nitrogen applications increase the concern of using nitrogen fertilizers. There are many ways to determine the amount of nitrogen in crops, but they are expensive, time-consuming, inaccurate, and require specialists to operate the tools (Saberioon et al, 2014).

Plants absorb nitrogen from the soil solution in the forms of  $\text{NO}_3^-$  or  $\text{NH}_4^+$  (Mengel and Viro, 1978). The nitrogen form used by plants depends on their growth stage. For example, in rice,  $\text{NH}_4^+$  is absorbed more during the vegetation growth stage than other stages. However,  $\text{NO}_3^-$  is absorbed more during the reproductive growth stage (Takenaga, 1995).

Ammonium sulfate  $[(\text{NH}_4)_2\text{SO}_4]$  and urea  $[\text{CO}(\text{NH}_2)_2]$  are generally considered effective fertilizer compounds (Gaudin and Dupuy, 1999).

Many electrochemical methods have been used to evaluate soil fertility, including ion-selective electrodes (glass or polymer membrane) and ion-selective field-effect transistors (ISFET). These methods measure potential differences related to the concentration of specific ions (e.g.,  $\text{H}^+$ ,  $\text{K}^+$ ,  $\text{NO}_3^-$ ).

### 1.2.2 Soil moisture

Increasing agricultural production with less land and water use is the goal of agriculture in many parts of the world. Thus, technologies that monitor water use will greatly improve agricultural production (Robert et al., 2013).

Soil water content is commonly expressed in gravimetric units as the ratio of the mass of water to the mass of the dry soil

$$\text{GWC} = M_w / M_s \quad (1.1)$$

where

GWC is the gravimetric moisture content,

$M_w$  is the mass of soil water and

$M_s$  is the mass of the soil.

The volumetric moisture content is the ratio of the volume of soil water to the volume of the total soil

$$\text{VWC} = V_w / V_t \quad (1.2)$$

where

VWC is the volumetric moisture content,

$V_w$  is the volume of soil water, and

$V_t$  is the total soil volume.

$$\text{GWC} = \text{VWC} / \rho \quad (1.3)$$

where

$\rho$  is soil bulk density

Classical soil gravimetric water content measurement techniques require removal of the moisture from the soil sample by evaporation. The thermo-gravimetric technique (oven-drying) is

widely used for measuring the soil water content and has been employed as the standard reference for determining soil water content (ASTM, 2008). The water content is calculated by subtracting the oven-dry weight from the initial field soil weight (Lunt et al., 2005).

The ions contained in the soil can be used to determine the soil water content. Higher conductivity of the soil caused by the flowing ions indicates an increase in water content due to a reduction in electric resistance. Therefore, measuring the voltage drop across the soil and calibrating it against the water content is a possible way to measure the water content (Soorya et al., 2013). The disadvantages of this technique are the requirement of calibration each time it is used and failure in saline soil (Robinson et al., 2008).

Soil bulk density is one of the soil properties that affects soil water content. Soil bulk density is defined as the ratio of the total mass of soil to the total volume of soil

$$\rho = M_t / V_t \quad (1.4)$$

where

$\rho$  is soil bulk density,

$M_t$  is the total mass of the soil, and

$V_t$  is the total volume of the soil.

Most soil water sensors that use microwave technology rely on a calibration procedure to measure the soil's bulk density and then assume that the bulk density of the soil remains constant throughout the measurement. In fact, for almost all applications, it is valid to assume that the soil's bulk density does remain the same throughout the measurement period. For long-term, in-situ monitoring, however, this assumption may no longer hold true. Physical or environmental conditions at the monitoring site change with time and the soil bulk density around the sensor will change.

For measuring soil volumetric water content, different designs of capacitance probes have been used. Capacitive probes have a pair of electrodes that are inserted into soil to form a capacitor. This capacitor and an inductor in an oscillation circuit generate oscillations. The frequency of the oscillation is determined by the capacitance. Changes of the capacitance due to changes in the volumetric water content change the oscillation frequency.

### **1.2.3 Soil salinity**

The solution ionic conductivity within soil water increases the permittivity, especially the imaginary part of the relative permittivity. The conductive behavior of soils has been used to measure bulk soil EC with a contact-type design, which places four electrodes in a Wenner array. The configuration of the electrodes on the immediate surface of the soil is a non-contact, nondestructive design that uses the electromagnetic induction principle. Wenner array sensors have been tested and were found highly accurate in measuring soil salinity, because the major factor determining bulk soil EC is soil salinity (Lee, 2005).

Soil water content also has a strong effect on bulk soil EC. Thus, measurement of soil salinity using this Wenner array design has to be made under known soil water conditions. Because of the interactions between the capacitive and conductive characteristics of soils, these sensors cannot separate the main factors that affect the bulk soil EC-soil water content and salinity (Zhang et al. 2004).

Soil salinity is one of the reasons for desertification. Soil salinity spoils large areas of farmlands, grasslands, and forests, especially in arid and semiarid regions (Yun et al. 2003). When plants are irrigated with saline water, they must extract fresh water from brackish water during transpiration, which leaves salts behind to accumulate in the soil. Consequently, soil salinity is increased (Skinner and Lambert, 2011).

Soil salinity is caused by several factors. Soils may become saline as a result of land use, including the use of irrigation water with high levels of salt. The intrusion of salty water into fresh water aquifers may occur when wells are close to the coast and water is pumped to the surface for various purposes, including irrigation. Irrigating from salt-impacted wells or saline industrial water may lead to the formation of saline soils.

A saline soil contains excess soluble salts that reduce the growth of most crops or ornamental plants. These soluble salts contain cations such as sodium ( $\text{Na}^+$ ), potassium ( $\text{K}^+$ ), calcium ( $\text{Ca}^{++}$ ), and magnesium ( $\text{Mg}^{++}$ ) along with anions chloride ( $\text{Cl}^-$ ), sulfate ( $\text{SO}_4^{--}$ ), nitrate ( $\text{NO}_3^-$ ), bicarbonate ( $\text{HCO}_3^-$ ), and carbonate ( $\text{CO}_3^{--}$ ).

In many countries around the world, especially arid and semi-arid countries, soil salinity has a great impact on soil productivity due its effect on soil fertility. It is also a critical environmental problem (Ochieng et al. 2013).

In Iraq, it has been estimated that 4% of the irrigated area is severely saline, 50% moderately saline, and 20% slightly saline (Al-Taie 1970). Other estimates indicate that the area of salt-affected soils in Iraq is about 6.7 Mha (Abrol et al. 1988); however, only 1.0 Mha are partially or totally reclaimed (Committee of Agriculture and Water Resources Sector 2009). The Food and Agricultural Organization (FAO) estimates that 70% of the total irrigated area of Iraq has been robbed of its production potential because of soil salinity. The FAO estimated that up to 30% of the farm land in Iraq has gone completely out of production (FAO 2011).

Many factors cause soil salinity, with the most significant ones being land use and use of irrigation water with high levels of salt. Salinity decreases the amount of available water for the plant. High salt levels make water absorption by the plant hard, causing physiological drought in the plant. The soil may contain an acceptable amount of water, but plant roots cannot absorb the

water, because of the high osmotic pressure or low osmotic potential. This is known as the osmotic or water-deficit effect of salinity. Plants are most sensitive to salinity during germination and early growth.

The accuracy of in situ measurements with dielectric sensors in saline and clay soils is problematic, although some experimental and theoretical studies have been presented to solve this problem (Marco 2011). Thus, a potential application of the sensor developed in this study is to monitor soil salinity, especially in the middle and south parts of Iraq, where most soil has a high level of salinity.

### **1.3 Current soil fertility sensing technologies and our previous work**

Most soil fertility sensing technologies for field applications are either optical, electrochemical, or dielectric in nature. While these have improved with time, there are a number of issues preventing these technologies from being adapted effectively for real-time sensing and long-term monitoring.

For measurement of macro-nutrients, N, P, and K, in soil, field sampling followed by laboratory tests has been traditionally considered the de-facto standard. Recent research has focused on fast, in-field sensing of these nutrients, especially on-the-go sensing technologies. The direct measurement of nitrogen by laboratory procedures is one of the most accurate and direct ways, but it is time-consuming and it needs a skilled operator (Gholizadeh et al. 2011).

Nitrogen is the most important macronutrient for crop growth, but nitrogen fertilizer residuals found in ground water or drinking water, as the result of leaching or runoff, damage the ecological system and the natural environment. Technologies for in-situ soil nitrogen sensing developed in past years have concentrated on three basic types: spectrophotometric/spectroscopic



(optical) sensing, electrochemical sensing, and dielectric sensing (Sinfield et al. 2010). Similar methods have been used for phosphorus and potassium sensing.

### **1.3.1 Optical sensors**

Included in the category of spectrophotometric/spectroscopic sensing are copper/cadmium reduction (CCR), near-infrared reflectance spectroscopy (NIRS), mid-infrared Fourier transform attenuated total reflectance (ATR) spectroscopy, and morphology-dependent simulated Raman scattering (MDSRS). The optical methods, such as NIRS, enable truly non-contact measurements. However, these sensors only measure properties immediately adjacent to the sensor (within a millimeter) and the measurement is strongly affected by other factors that change the spectral reflectance of the soil. Furthermore, these methods require considerable site-specific calibrations due to inherently variable optical features of different soil types. It was also found that the results were not satisfactory within the range the macronutrient concentrations usually applied through fertilizer (Kim et al. 2009).

### **1.3.2 Electrochemical sensors**

Electrochemical techniques included the nitrate ion-selective electrode (ISE), nitrate ion-selective field-effect transistor (ISFET), and a combination of CCR and ISE. The ISEs have proven to provide high accuracy with low limits of detection. However, the tedious micronutrient extraction procedure, long response time, limited durability of the electrodes, low robustness, high maintenance requirements, and the need for frequent calibration due to signal drift make the use of ISEs in in-situ, real-time applications difficult. Nevertheless, the electrochemical sensors have been the choice of researchers for real-time, on-the-go soil sensing for site-specific crop management. Several researchers used custom-designed soil samplers and commercially available ISEs for sensing nitrate and pH in soils (Kim et al. 2009). Although research for many years did

not lead to commercially available technologies, a soil pH mapping system is now commercially available (Adamchuk et al. 2014). In general, on-the-go sensing of soil fertility using the optical methods has been less successful mainly due to calibration and accuracy issues (Kim et al. 2009).

The optical and electrochemical sensors are generally expensive and not mechanically robust, making them unsuitable for long-term burial that is required to observe dynamic changes in the soil at specific locations over time.

### **1.3.3 Dielectric sensors**

Another technology with great potential in measuring soil fertility in-situ is dielectric spectroscopy (impedance spectroscopy) that measures the permittivity of soil. Different dielectric sensors have been developed for soil measurement, including time-domain reflectometry (TDR), frequency-domain reflectometry (capacitance probe), and the standing wave ratio (SWR) method. The use of dielectric sensors has mainly been restricted to measuring water content and electrical conductivity, and there are few reports in the literature concerning field experiments involving dielectric sensors that measure soil fertility. The enclosed design of some dielectric sensor probes also makes them unsuitable for underground sensing.

Many studies have shown the capability of dielectric sensors to measure water content and salinity. By measuring the complex dielectric constant of two types of soil using an L-band dielectric probe (1.25 GHz), concentrations of two types of salts were detected (Sreenivas et al. 1995). It was observed that, at this frequency, the dielectric constant is dependent on the soil texture and volumetric moisture content, while salinity does not have much influence on the dielectric constant. Furthermore, the dielectric increases with increases in salinity and moisture content in the soil. In dry soils, the soil conductivity and dielectric loss are low, because there exists more air inside the soil, which further reduces the dielectric loss. For wet soil, the

conductivity of soil increases with an increase in salt content, indicating increases in dielectric loss.

The dielectric constant is an indicator of the material's capability of storing energy in an electric field (electrical polarization), while the dielectric loss is an indicator of the material's capability of absorbing energy from the alternating electric field. The measurement of dielectric properties of soils related to moisture content and fertilizer content is useful in agriculture (Yadav et al. 2009). Measurement of soil moisture and electrical conductivity (salinity) is critical for agricultural, engineering, and environmental applications. (Suweis et al. 2010).

There are many methods for indirect, nondestructive, in-situ, dielectric measurement of soil moisture such as time domain reflectometry (TDR), capacitance-based frequency domain reflectometry (FDR), impedance-based amplitude domain reflectometry (ADR), phase transmission, and time-domain transmission (TDT) (Muñoz-Carpena et al. 2005a). These dielectric methods all estimate soil moisture and bulk electrical conductivity by measuring soil dielectric properties (Topp et al. 1980).

Accurate estimation of soil salinity is an important issue, especially in arid regions, where the salinity of the soil may increase significantly and become a danger to plants (Hindrikx et al. 2002). Measurement of soil salinity is complicated, because it is affected by many factors that produce measurement errors, including temperature, moisture, and texture of the soil. Therefore, the development of new measurement methods and equipment for evaluating soil salinity is an active field of research. An ideal measurement tool for soil salinity should read the majority of the influencing factors at the same time and same location as quickly as possible. This could be done by incorporating various sensors in a single unit for insertion into the measured material or by selective analysis of a sensor's outputs for discrimination of various quantities like soil water

content and electrical conductivity, as in the case of time-domain reflectometry (Boldaji et al. 2011).

Time domain reflectometry (TDR), the capacitance technique, and frequency domain reflectometry (FDR) have been the major techniques that use the dielectric property of the soil to measure soil moisture content. The significant difference in the dielectric constant between dry soil and pure water is the basis of these dielectric technique. The major disadvantages of TDR are its high cost, loss of reflection in highly saline soil, and increase in conductivity with wetting of the soil mass. The device may give a good waveform in a very dry saline soil, but not a good one in a wet saline soil (Marco, 2011).

Capacitance-based techniques have an oscillating circuit and a sensing part that are placed inside the soil. This technique determines the dielectric permittivity/dielectric constant of a medium (soil) by measuring the charge time of the capacitor (Minet et al. 2010). The soil moisture content determines the electrical capacitance of the capacitor. When this capacitor become a part of an oscillator, forming an electrical circuit, changes in the frequency of the oscillation indicate changes in the capacitance, hence change in soil moisture content. Thus, the soil water content is measured (Muñoz-Carpena et al. 2004). This is a non-destructive method (Zazueta and Xin 1994).

## 1.4 Permittivity

Each dielectric material has a specific conductivity and a specific ability to polarize in an electric field, hence a specific permittivity. When a contaminant is added to a dielectric material, water as an example, the permittivity of the contaminated water will be different from that of the pure water. That is why a material's permittivity can be an indicator of chemical composition changes of the material (Scholte et al. 2002).

Interpretation of dielectric permittivity is complicated in saline soil and in fine-textured soil, because of the effect of ionic conductivity and the clay-water-ion interactions, which depend on the frequency that affect the permittivity reading (Kelleners et al. 2005).

Permittivity is the behavior of a material in resisting the formation of an electric field in the material. When a current flows, an electromagnetic wave is applied to the material. Capacitive and conductive behaviors of the material determine the amount of total current flowing in the material. Polarization hinders current flow by storing electrical energy that is released once the application of the field is stopped. It occurs when dipole molecules in the material are aligned in the opposite directions at low frequencies, when the polarity of the electric field changes slowly. (Robinson et al., 2003).

Dielectric relaxation is the dissipation of applied energy, which occurs at high frequencies when the dipole orientation cannot follow the change in the polarity of the electric field due to the binding force between atoms. Permittivity can be expressed in a complex form due to this relaxation process, because it gives rise to a phase lag between the imposed field and the material's response to it. This phase lag is a function of the frequency of the imposed field (Topp et al. 2000).

Zhang et al. (2004) pointed out that, at a sufficiently high-frequency range, the effects of the conductive behaviors on the capacitive behaviors of soils can be minimized.

Hillhorst (1998) reported that, at low frequencies, electric currents in soil due to conductance are higher than those due to capacitance, so the soil EC can be measured more accurately than dielectric permittivity. At low frequencies, the effects of the capacitive behaviors on the conductive behaviors of soils can be minimized.

There are many factors, including frequency, temperature, and moisture content, that affect the dielectric properties of materials. Dielectric properties can vary significantly with frequency. The frequency-dependent trend of dielectric properties can provide important information about the material characteristics (Wang et al. 2003).

Permittivity is a quantity used to describe dielectric properties that influence the reflection of electromagnetic waves at interfaces between materials and the attenuation of wave energy within materials.

In the frequency domain, the complex relative permittivity  $\epsilon^*$  of a material to that of free space can be expressed in the following form:

$$\epsilon^* = \epsilon' - j \epsilon'' \quad (1.5)$$

The real part  $\epsilon'$  is referred to as the dielectric constant and represents stored energy when the material is exposed to an electric field, while the dielectric loss factor  $\epsilon''$ , which is the imaginary part, influences energy absorption and attenuation, and  $j = \sqrt{-1}$  (Afsar et al. 1986).

The imaginary part of permittivity is called the loss factor and is a measure of how dissipative the material is to an external electric field. The imaginary part is always greater than zero and is usually smaller than the real part. The loss factor includes the effect of both dielectric loss and conductivity. The real part of permittivity is a measure of how much energy from the external electric field is stored in the material. Permittivity can change with frequency, temperature, orientation, pressure, and molecular structure of the material.

## 1.5 Our previous work on dielectric sensors

Dr. Zhang's group at Kansas State University has developed a unique, frequency-response (FR)-based permittivity sensor to measure simultaneously multiple properties of dielectric materials. Permittivity describes the response of a dielectric material, such as soil, to externally applied alternating electric fields. The response is characterized by *polarization* of the dipole molecules at lower frequencies, *relaxation* due to the binding force between atoms at higher frequencies, and *conduction* in both solid and liquid (ionic) phases. Thus, permittivity strongly depends on the composition of the material at the molecular, atomic, and ionic levels. Furthermore, because the polarization, relaxation, and conduction are all frequency dependent, information specific to each component in soil may be extracted at specific frequencies or frequency ranges ("signature frequencies"). Due to the complexity in soil composition, statistical methods are often used to extract these "signature frequencies".

The FR-based soil sensor generates sinusoidal waves at multiple frequencies within a wide frequency range. Dr. Zhang's group has used from near 0 MHz or direct current, DC, to 400 MHz. It measures the responses of the sensor probe filled with the dielectric material to these waves, hence forming "dielectric spectra". Because of the use of multiple frequencies, prediction models to measure different properties can be established based on the same spectra. As a result, multiple properties of the materials may be measured simultaneously. In the past, the sensor has been used to measure simultaneously soil VWC, density, and salinity (Zhang et al. 2004; Lee et al. 2007<sub>1,2</sub>), fossil fuel/biofuel mixing ratio (Xi and Zhang 2011), air pollutants – glycerol, ethanol, and ammonia (Ware 2012), nutrients in water (Shultz 2009), and sediment concentration in water (Utley et al. 2012). However, these sensors had not been tested in measuring soil fertility before 2017.

Noticing the work done by Zhang and his group on frequency-response based dielectric sensors, Chighladze et al. (2011) developed a similar sensor to measure soil solution nitrate concentration. The sensor measured frequency responses of the soil and used the partial least squares (PLS) method for multi-variate analysis. Instead of using an open probe, they used a closed, cylindrical sample holder to hold soil samples. For soil with relatively high VWC and low salinity, they successfully measured  $\text{NO}_3\text{-N}$  concentration in soil. Using similar methods, Pluta and Hewitt (2009) measured soil moisture content and density, and Pandey et al. (2013) estimated individual components in soil, including air, water, and nitrate.

## **1.6 Need for sensors buried in soil**

For practical, in-situ measurement, the sensor must meet the following requirements: (1) is capable of taking measurements at different depths, especially in the crop root zone, (2) has a rugged sensor probe to sustain harsh outdoor and underground conditions and mechanical damage, and (3) has an open probe that is in constant contact with soil so that dynamic variations in soil properties can be tracked. Additionally, sensors having the following capabilities are preferred for field application: (1) capable of real-time sensing to allow timely management decisions, such as fertilizer prescription, (2) capable of long-term deployment with minimum maintenance for continuous monitoring, (3) capable of measuring multiple properties, and (4) capable of having measurement data easily downloaded. With all these requirements considered, the buried-in-soil FR sensor developed in Dr. Zhang's laboratory seems to be the best candidate.

Similar to many other soil sensors, the FR sensor measurement is influenced by multiple soil properties, of which soil type, soil texture, and organic matter may have the strongest effects. Thus, the sensor may need frequent calibration. However, when the sensor probe is buried in soil at a fixed location and a fixed depth for long-term monitoring, the type and texture of the soil



surrounding the sensor probe probably do not significantly change. As a result, the sensor should not need frequent calibration. Moreover, if the soil bulk density at the sensor location does not vary significantly, the effect of bulk density on the measurement would also be minimized. Under these conditions, only water content and macronutrients have significant effects on the impedance spectra; hence, more accurate measurement of water content and macronutrients can be expected. Additionally, because of the simple and rugged mechanical design, the low sensitivity of the FR dielectric assessment to probe surface contact-resistance changes, and the anti-corrosion material selected for the probe, it is believed that the FR sensor is more suitable for long-term monitoring of dynamic soil changes and for effective detection of adverse conditions such as drought, nutrient deficiencies, and compaction.

## **1.7 Partial least squares regression**

The partial least squares (PLS) regression is a good prediction method when there are a large number of correlated x-variables (Numerical Algorithms Group 2007). The PLS regression method combines principal component regression (PCR) with multiple linear regression (MLR). The PCR method finds factors that maximize the variance of the independent (predictor) variables, and MLR finds a variable to maximize the correlation between the independent (predictor) and dependent (response) variables. Thus, PLS is an improvement over MLR and PCR, because it uses information from both the predictor and response variables to form a model. As a result, PLS finds factors that maximize the variance and correlation (Wise et al. 2006).

The PLS regression is particularly suited when the matrix of predictors has more variables than observations (samples) and when there is multi-collinearity among the predictor variable values (Wikipedia 2020), which are the features of our dielectric spectral data. Thus, this method is selected as the main analytic tool for this study.

The R-squared value is a statistical measure on how close the sample data are to the fitted regression line. It is also known as the coefficient of determination, or the coefficient of multiple determination for multiple regression. One hundred percent indicates that the model explains all the variability of the response data around its mean. The root mean square error (RMSE) is the standard deviation of the residuals (prediction errors). Residuals are a measure of how far the data point are from the regression line. The RMSE is a measure of how spread out the residuals are. In other words, it gives the indication on how concentrated the data are around the line of best fit. In this study, we will use these two statistical indices to evaluate the performances of the prediction models developed through the PLS regression analyses.

## 1.8 Objectives

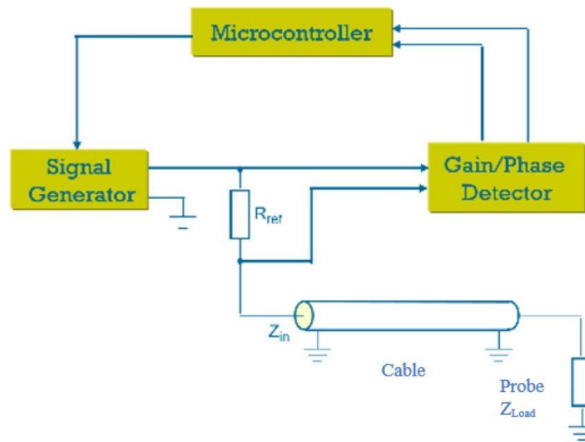
The specific goals of this study were:

1. To design a probe for the already-developed frequency response-based (FR) dielectric sensor that can be buried in soil for real-time measurement and long-term monitoring of multiple soil properties.
2. To conduct a laboratory experiment to evaluate the FR sensor with the new soil probe in their ability to measure simultaneously soil water content, salinity, bulk density, and one or two types of nitrogen fertilizer.
3. To conduct a long-term, outdoor experiment to exam the FR sensor with multiple probes buried in the root zone of corn plants in their ability to measure simultaneously soil water content, bulk density, a nitrate-N fertilizer, an ammonium-N fertilizer, and urea.
4. To study the correct procedure for calibration of the sensor.
5. To identify the strengths and potential difficulties of the sensor in agricultural applications.

## Chapter 2 - Materials and Methods

### 2.1 Basic measurement principle - the frequency-response method

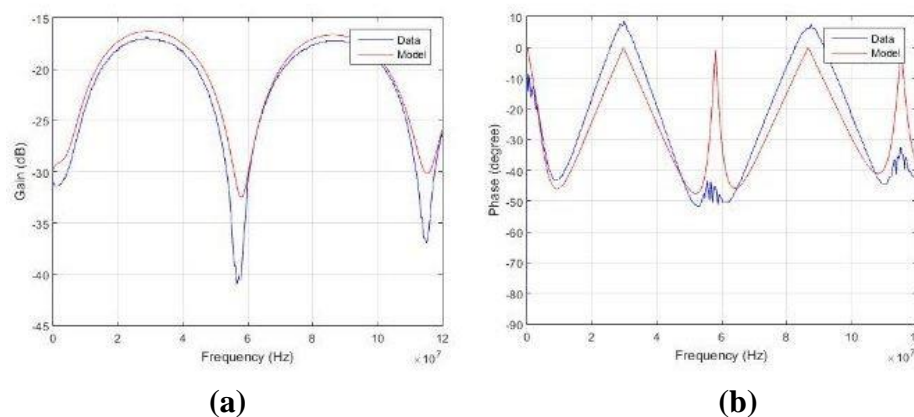
The basic measurement principle for the frequency-response (FR) sensor is to consider the probe inserted in soil as an impedance load,  $Z_{Load}$ , which forms a voltage divider with a constant, precision resistor  $R_{ref}$  (Figure 2.1). When a sinusoidal signal is applied to the voltage divider as the input signal, the output signals measured at  $Z_{Load}$  will have a lower amplitude and a phase delay, although still at the same frequency. The magnitude reduction the gain and the phase delay are measured by a gain-phase detector. Because the gain and phase responses of the probe filled with soil (as the medium) are determined by the complex permittivity of the soil, the gain and phase measurement is related to the factors that determine the conductive and dielectric nature of the soil. By measuring the gain and phase delay at multiple frequencies, dielectric gain and phase frequency response curves, or “spectra”, can be obtained. Considering the geometric parameters of the probe, spectra for the real and imaginary parts of the soil permittivity can be derived. Various methods for spectral data analysis can then be used to extract “signature frequencies” for specific soil properties. Because different soil properties have different conductive or dielectric effects in different frequency regions, multiple properties of soil may be measured.



**Figure 2.1 Measurement principle of the FR sensor**

For field measurement, the sensor probe and the control unit have to be connected through a proper cable. With a cable inserted between  $R_{\text{ref}}$  and  $Z_{\text{Load}}$ , the voltage divider is no longer formed by  $R_{\text{ref}}$  and  $Z_{\text{Load}}$ . Instead, it is now formed by  $R_{\text{ref}}$  and  $Z_{\text{in}}$  (Figure 2.1). Thus, the frequency response measured by the system is a representation of  $Z_{\text{in}}$ , not  $Z_{\text{Load}}$ . In order to study the characteristics of  $Z_{\text{Load}}$ , an impedance transformation model between  $Z_{\text{in}}$  and  $Z_{\text{Load}}$  needs to be studied based on transmission line theory. During the past, we have established the impedance transformation model and used it to simulate gain and phase responses of different load impedances and co-axial cable length.

. Figure 2.2 shows the gain and phase responses to an impedance load, which is a parallel connection between a 2,000 pF capacitor and a  $10\Omega$  resistor, with a 60 cm long co-axial cable. The repeated patterns of the responses shown in the figure are the result of signal reflections within the cable length. It can be seen that the simulation matched the measured responses very well, which validated the impedance transformation model (Tang 2009). However, this simulation did not include the impedance model of the probe. Because of the irregular shape of the capacitive probe, an impedance model of the probe must be developed to allow for prediction of frequency responses of the sensor to various medium properties. Future studies will consider the impedance modeling of the probe.



**Figure 2.2 Simulation vs. measured frequency responses (a) gain, (b) phase**

The principle of the FR sensor is considered in parallel with that of time-domain reflectometry (TDR). In TDR, time responses of the probe-medium to an impulse input are measured, whereas in the FR sensor, frequency responses of the sensor to sinusoidal inputs are measured. Because measuring time response requires equipment with high-time-resolution, the TDR sensor imposes higher technical difficulties in the design and manufacture. Nevertheless, from the literature, the frequency-response measurements require expensive impedance meters. Practical and inexpensive FR sensors, such as the one developed in our laboratory, are still unique.

To date, the FR sensor has been tested in various dielectric media using frequencies of up to 400 MHz, including soil, air, water, and fuel. We also studied the effect of temperature on the measurement (Lee 2005).

## **2.2 The control box**

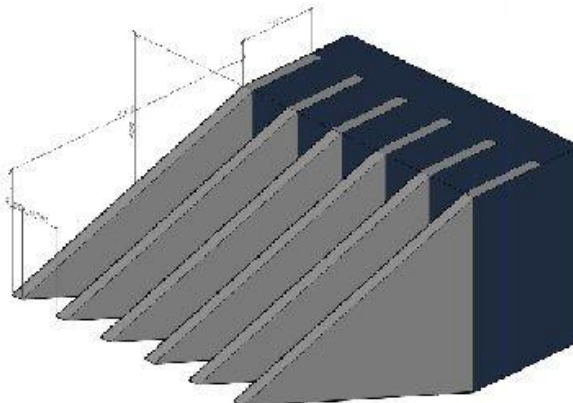
The control box (Figure 2.3) measures the gain and phase responses of the probe/soil sample at 635 frequencies, ranging from 50 Hz to 120 MHz. It is programmed to test at each frequency three times. After measurement, data are uploaded into a PC computer via a serial port. The control box is powered by a power supply of 9V. The electronic circuit in the control box is composed of an Infineon CR167 microcontroller (Infineon Technologies, Neubiberg, Germany), a signal generator, and a gain and phase detector. The signal generator generates sinusoidal signals as the input of the voltage divider. The gain and phase detector measures the magnitude and phase changes between the input (the signal generated by the signal generator) and the output (the signal reflected back from the probe through the cable) of the voltage divider (Tang, 2009).



**Figure 2.3 The control box**

## **2.3 Probes**

The probe was specifically designed for buried-in-soil use. It was constructed from six parallel stainless steel plates. Each plate is electrically insulated from the neighboring plates. The odd numbered plates were electrically connected, and so were the even-numbered plates. During measurement, the probe is immersed or buried in the medium material to be measured, such as soil. For the soil tests, the plates were made in a triangular shape so that it can be penetrated into soil easily. Each plate is 2 mm thick, and the dimensions are shown in Figure 2.4. The probe has a rectangular base with dimension 1x4 cm. The six plates are mechanically installed on the base with proper insulations. The spaces between adjacent plates were 8 mm and the area of each plate was 8 cm<sup>2</sup>.



**Figure 2.4 Dimensions of the probe**

A 60-cm coaxial cable connected the probe with the control box. The base was assembled on a T-shaped mounting structure so that the probe can be pushed into the soil samples (Figure 2.5).



**Figure 2.5 The probe and the coaxial cable for laboratory tests**

## **2.4 Experimental procedure**

### **2.4.1 Laboratory experiments**

The main goal of the laboratory experiments was to observe the frequency response of the sensor to soil samples at different combinations of density, volumetric water contents (VWC), salinity, and concentrations of several types of fertilizers. Three experiments were conducted in the Instrumentation and Control Laboratory of the Biological and Agricultural Engineering Department of Kansas State University:

- 1) Tests on soil samples with different levels of VWC, density, and salinity.
- 2) Tests on soil samples with different levels of VWC, density, and concentration of a single type of nitrogen fertilizer - urea ( $\text{CO}(\text{NH}_2)_2$ ), ammonium sulfate  $[(\text{NH}_4)_2\text{SO}_4]$ , and ammonium nitrate ( $\text{NH}_4\text{NO}_3$ ).



3) Tests on soil samples with different levels of VWC, density, and concentration of two nitrogen fertilizers combined.

#### **2.4.1.1 Test materials**

Smolan silty clay loam was used in the tests. The soil was collected from Riley County, KS, using the North American Proficiency Testing Program. It was oven-dried and ground.

The textural composition of the soil was determined by the Soil Testing Laboratory at Kansas State University. Particle size analysis was done using the hydrometer method. (The textural analysis is given in Appendix A).

Soil samples with five levels of salinity levels (0.1, 0.2, 0.4, 0.8, and 1.6 % salt by weight), five levels of GWC (12, 15, 18, 21, and 24 g/g), and three levels of bulk density (0.8, 1.0, 1.2 g/cm<sup>3</sup>) were prepared in the laboratory.

#### **2.4.1.2 Test apparatus**

A steel container in a rectangular shape was used to contain the soil sample. The internal dimensions of the container were 9.2 by 9.2 cm. The depth of the container was 10 cm. A steel “piston” of 9.1 by 9.1 cm was made to push the sample within the container to achieve different bulk densities (Figure 2.6).



**Figure 2.6 The laboratory tests container**

The piston was attached to a camera stand (Figure 2.7). The piston can move on the stand vertically and the movement can be adjusted with the help of a height scale on the stand. For each test, the piston was pressed into the sample to a calculated depth. Knowing the cross-sectional area, the penetration depth, and the weight of the soil sample, the density of the soil can be calculated. After each test, the probe was lifted from the soil sample, rinsed in distilled water, and dried with air. The coaxial cable that connected the probe with the control box remained connected between tests.



**Figure 2.7 The camera stand, piston, and container**

During the test, the control box was connected to the serial port of a computer. Microsoft HyperTerminal 5.1 program was used to collect the data from the control box and save the data in .txt files.

#### **2.4.1.3 Sample preparation**

To achieve different GWC and salinity levels, a certain amount of salt (NaCl) was first dissolved in a certain amount of distilled water, and the NaCl solution was then added to the soil sample. To allow the samples to reach the equilibration condition, measurements were initiated at least 12 hours after the NaCl solution was added.

Thus, the procedure to prepare soil sample was as follows:

- (1) Prepare the NaCl solution based on the desirable levels of GWC and salinity.
- (2) Place the soil onto a thin plastic sheet and spray the prepared NaCl solution into the soil.
- (3) Mix the soil by hand gently, until a high degree of uniformity was achieved.
- (4) Place the mixed soil into the container.
- (5) Cover the container with a plastic bag tightly.
- (6) Keep the sample untouched for at least 12 hours.

Thus, to achieve combinations of 5 GWC and 5 salinities, a total of 25 samples were prepared.

During the test, the piston mounted on the camera stand was moved downwards to reach the lowest density to take the first measurement. The piston was then moved to two more positions to achieve two higher densities.

The three bulk densities selected for the laboratory tests were 0.8, 1.0, 1.2 g/cm<sup>3</sup>. These densities were achieved by changing the height of the soil sample in the container, which changed the volume of the sample (Table 2.1)

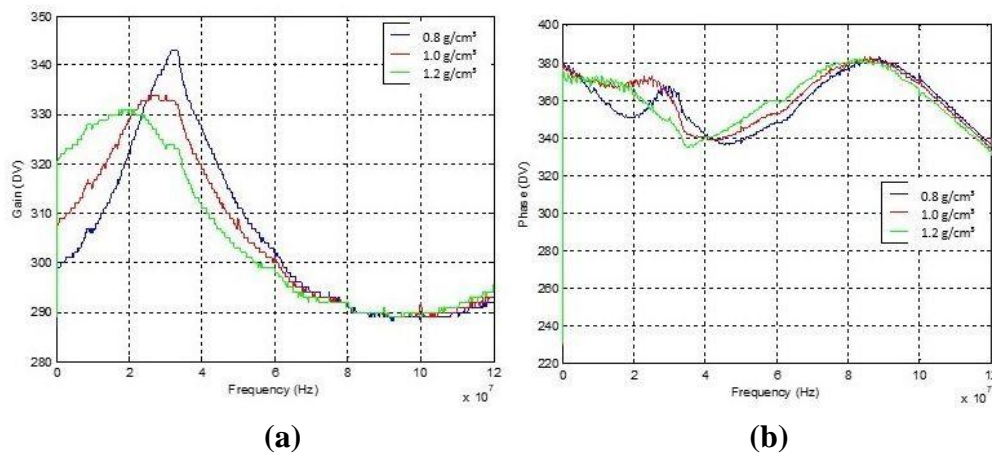
**Table 2.1 Soil sample size for different density**

Dry soil (g)	Area (cm <sup>2</sup> )	Height (cm)	Volume (cm <sup>3</sup> )	Density (g/cm <sup>3</sup> )
600	84.64	8.5	719.44	0.8
600	84.64	7	592.48	1.0
600	84.64	6	507.84	1.2

#### 2.4.1.4 Preliminary observation of the dielectric spectral data

##### 2.4.1.4.1 Soil density

Preliminary observations of the obtained dielectric spectra, including the gain spectra and phase spectra (Appendix B), show significant differences caused by changes in the bulk density levels (Figure 2.8).



**Figure 2.8 (a) Gain and (b) Phase frequency responses at different density levels**

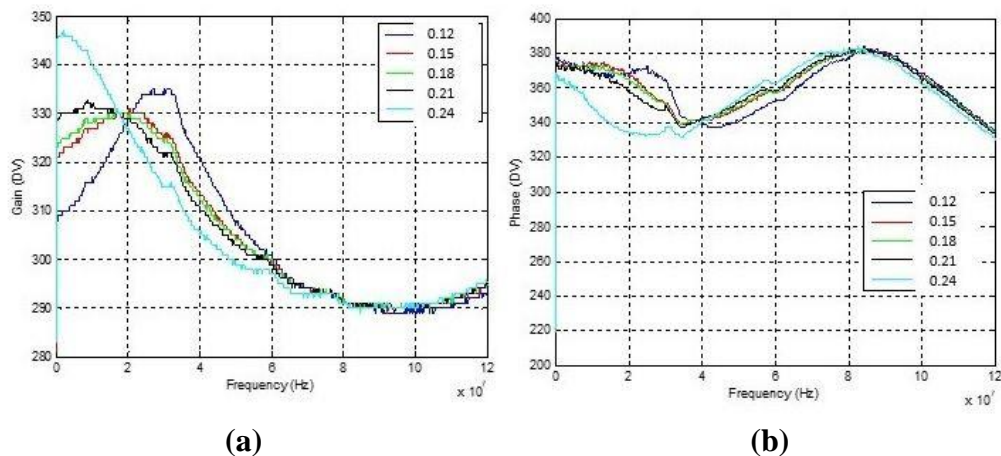
#### 2.4.1.4.2 Soil moisture

Table 2.2 gives the amount of water added to the soil sample to achieve the desirable GWC.

**Table 2.2. Water and soil weight at different GWC levels**

Dry soil (g)	Water (g)	GWC (g/g)
600	72	0.12
600	90	0.15
600	108	0.18
600	126	0.21
600	144	0.24

Initial observation of the gain and phase spectra also showed significant differences among different GWC levels (Figure 2.9).



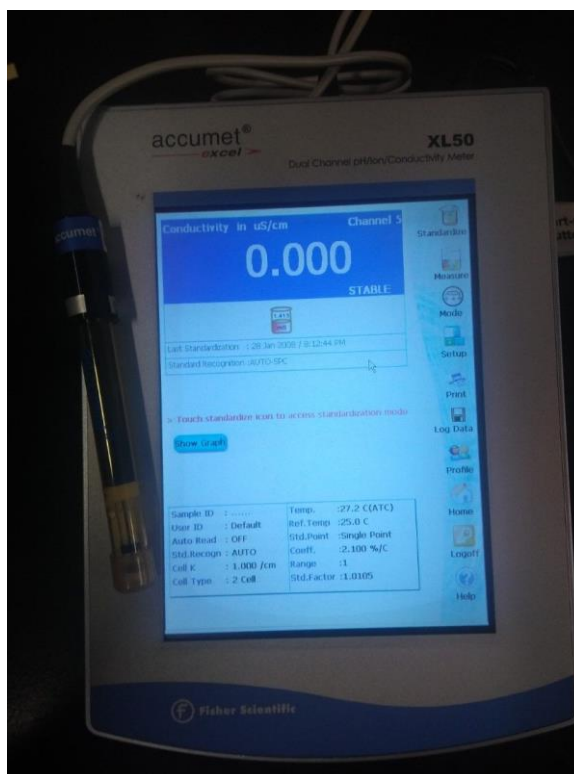
**Figure 2.9 (a) Gain and (b) Phase frequency responses at different GWC levels**

#### 2.4.1.4.3 Soil salinity

Table 2.3 shows that amount of NaCl salt needed to achieve the desirable salinity levels. Soil salinity was measured with Fisher Scientific Accumet Excel XL 50 Series Meter (Figure 2.10).

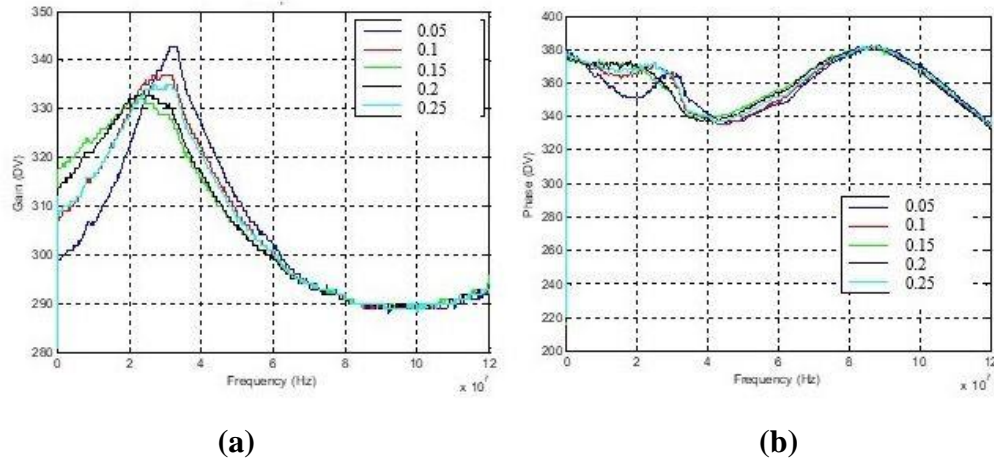
**Table 2.3 Amount of NaCl and soil to achieve different salinities**

Soil (g)	NaCl (g)	Conductivity (mS/cm)
600	0.3	1.93
600	0.6	2.51
600	0.9	3.25
600	1.2	4.53
600	1.5	6.34



**Figure 2.10 Conductivity meter**

Figure 2.11 shows significant changes in the dielectric spectra due to the change in salinity levels in the soil sample.



**Figure 2.11(a) Gain and (b) Phase frequency responses to different soil salt levels**

#### ***2.4.1.4.4 Soil fertility (single nitrogen fertilizer), VWC, and density.***

For these tests, the same procedure was used to vary soil VWC and density. Instead of NaCl, one of the three fertilizers was added to achieve at six different levels: 22.4, 33.6, 44.8, 56, 67.2 and 78.4 kg/ha. Thus, for three types of fertilizer, a total of  $5 \times 6 \times 3 = 90$  soil samples were prepared. With each sample pressed to three densities, 270 spectral data in total were obtained. The three types of nitrogen fertilizers were:

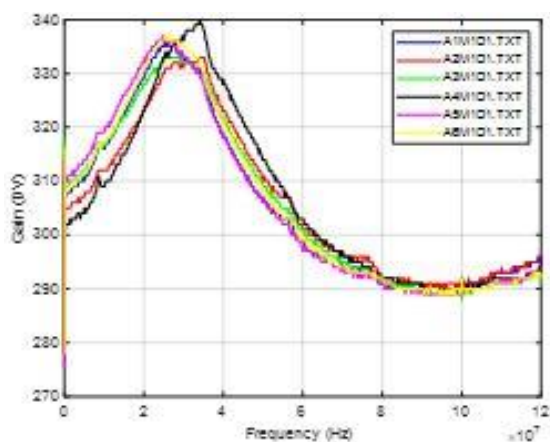
Urea                                       $\text{CO}(\text{NH}_2)_2$  (46% N)    (PotashCoro 46-0-0)

Ammonium sulfate     $(\text{NH}_4)_2 \text{SO}_4$  (20% N) (Raw nitrogen 20-0-0 by NPK Industries)

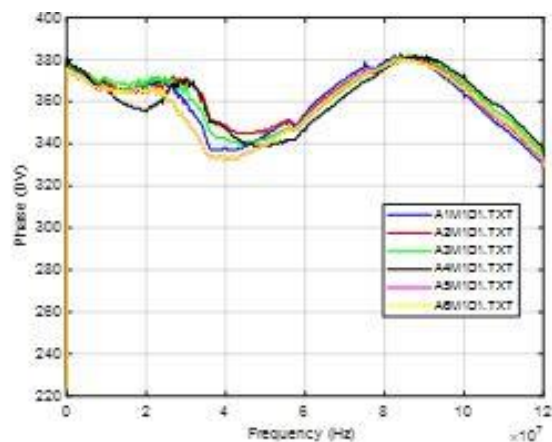
Ammonium nitrate         $\text{NH}_4\text{NO}_3$  (98% N) (Fisher Chemical)

Initial observation of the gain and phase spectra showed significant differences related to nitrogen contents (Figures 2.12– 2.14), which is a good indicator for the effectiveness of the sensor in soil nitrogen sensing.



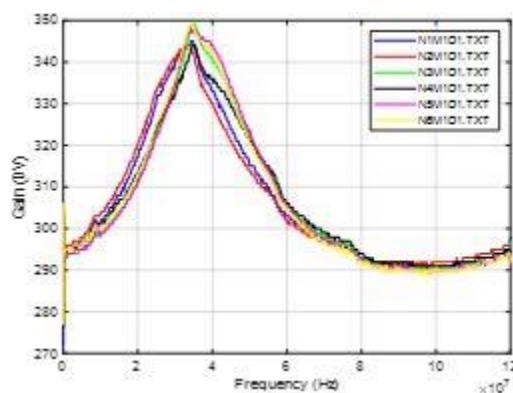


(a)

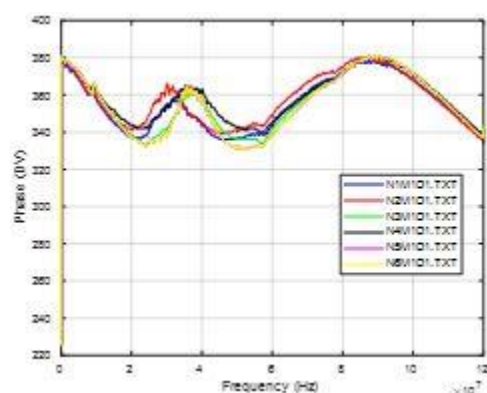


(b)

Figure 2.12(a) Gain and (b) Phase frequency responses to different ammonium sulfate levels

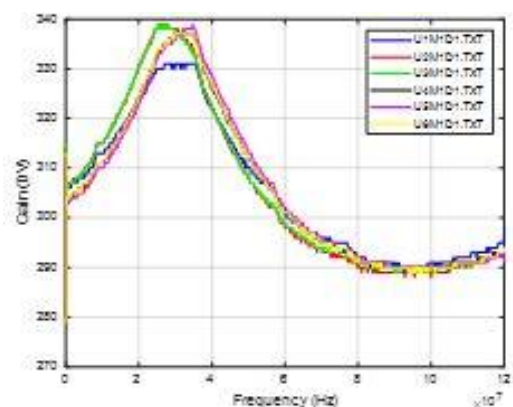


(a)

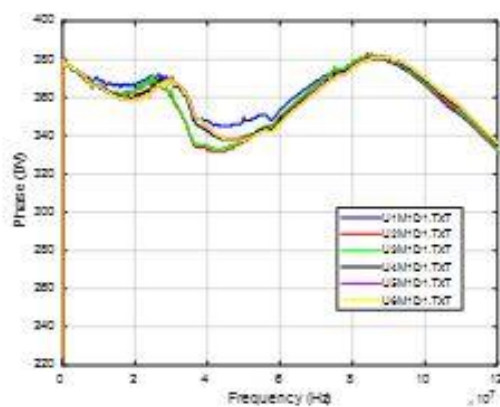


(b)

Figure 2.13 (a) Gain and (b) Phase frequency responses to different ammonium nitrate levels



(a)



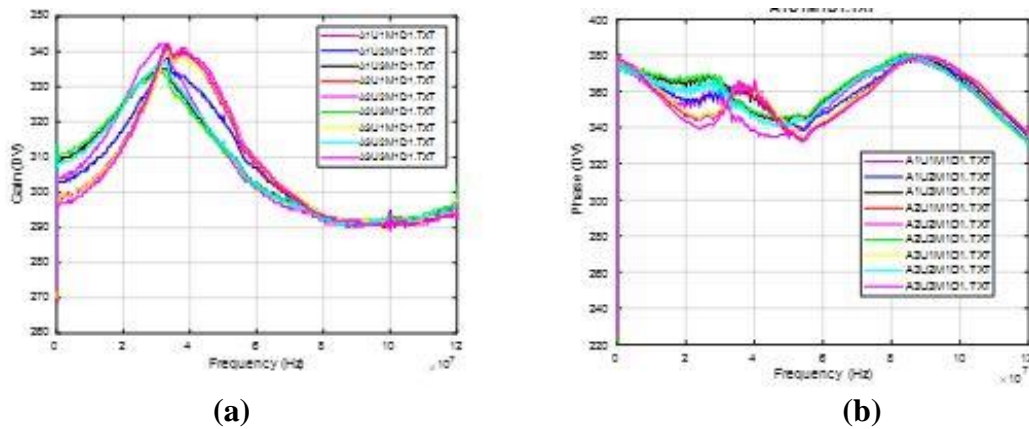
(b)

Figure 2.14 (a) Gain and (b) Phase frequency responses to different urea levels

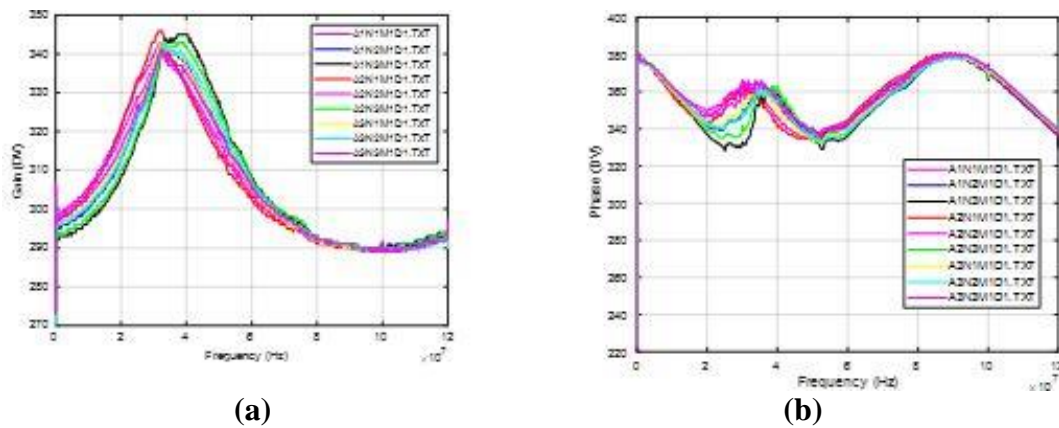


#### 2.4.1.4.5 Soil fertility (combined two nitrogen fertilizers), moisture, and density.

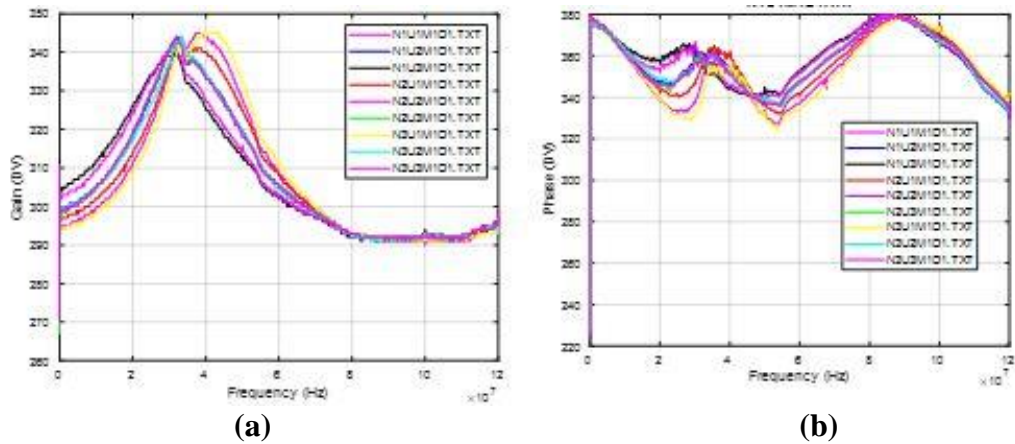
For this test, five GWC and three density levels were achieved using the same method. Among the three nitrogen fertilizers, combinations of two, each at 3 concentration levels – 22.4, 44.8, and 67.2 kg/ha -- were prepared. Later in the experiment, combinations of three more fertilizer concentrations – 11.2, 33.6, and 56 kg/ha were added. Thus, the total number of soil samples prepared for each two-fertilizer test was  $5 \times (3 \times 3 + 3 \times 3) \times 3 = 270$ . For three two-fertilizer combinations, the total number of samples were  $270 \times 3 = 810$ . Figures 2.15-2.17 display the examples of the spectral data obtained in these tests.



**Figure 2.15 (a) Gain and (b) Phase frequency responses at different ammonium sulfate & urea levels**



**Figure 2.16 (a) Gain and (b) Phase frequency responses with different ammonium sulfate & ammonium nitrate levels**



**Figure 2.17 (a) Gain and (b) Phase frequency responses with different ammonium nitrate levels & urea**

## 2.4.2 Outdoor tests

For the outdoor experiment, different test apparatus and procedures were prepared.

### 2.4.2.1 Test apparatus

#### 2.4.2.1.1 Pots

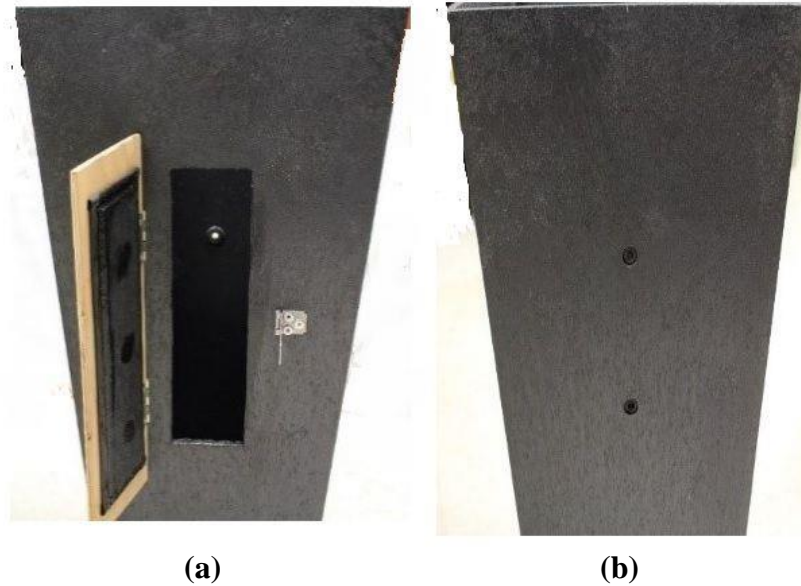
For the outdoor tests, two large pots were planted with a corn plant in each of the pots (Appendix C). The size of the pots was 14 in. x 27.5 in (Figure 2.18). The material was rubber.



**Figure 2.18 Dimensions of the pot**

On one side of the pot, a lockable window (Figure 2.19a) would be opened to allow soil samples from the pot to be taken for chemical tests. It also allowed measurement of soil volumetric water content (VWC) using a commercial sensor (Decagon 5TE, Pullman, WA). In addition, a small hole on the opposite side of the window (Figure 2.19b) was opened that allowed the coaxial cable of the sensor to be connected to the control box so that the dielectric spectral data could be acquired. When samples and data were not taken, the window was firmly locked to avoid leaking of the soil from the pot.

In order to obtain the vertical profile of the soil property information, two sensors were buried in each pot. The intention was to place both sensors within the root zone of the corn plant. Thus, the two sensors were buried at 12.5 cm and 25 cm from the soil surface, respectively.



**Figure 2.19 The outdoor tests pots (a) the window (b) the holes**

#### ***2.4.2.1.2 Probes***

The same probe designed for the laboratory tests was used for the outdoor tests. The only difference was that the probes were buried in soil and were never moved during the entire experiment. Thus, the T-shaped mounting structure was not needed. For signal acquisition, each sensor had a 12-inch long coaxial cable (Figure 2.20).

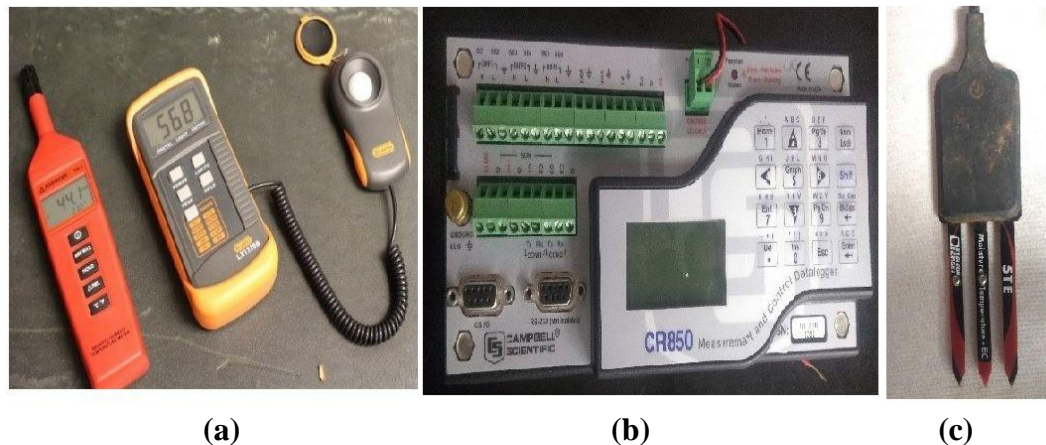


**Figure 2.20 The outdoor test probe and coaxial cable**

#### ***2.4.2.1.3 Other instruments***

Other instruments used in the experiment included a temperature and humidity meter (AMPROBE TE-3; Everett, WA), and a light meter (Dr.meter LX1330B; Shenzhen, People's

Republic of China). A Campbell Scientific (Logan, Utah) CR850 datalogger was used to collect data from a Decagon 5TE soil moisture sensor (Figure 2.21).



**Figure 2.21 Instruments used in the outdoor experiment (a) humidity and temperature meter and light meter, (b) datalogger, and (c) Decagon 5TE soil moisture sensor**

#### **2.4.2.2 Sample preparation**

In each pot, soil was added to the desired depth of the lower sensor; then we placed the lower sensor at the center of the pot. Then more soil was added until it reached the desired depth of the upper sensor. Then the upper sensor was placed at the center of the pot, and added more soil until it reached the desired height of soil surface, which was 5 cm from the top of the pot (Figure 2.22).

On 27 July 2019, soil and sensors were placed in both pots. On the same day, 4 corn seeds were planted in each pot. After all seeds germinated and emerged, two weaker seedlings in each pot. Another one, was later removed, leaving only the strongest seedling were removed from each pot (Figure 2.23).

The soil used is taxonomically identified as Smolan silty clay loam from the KSU North Farm. Samples were sent to the KSU Soil Testing Laboratory to examine the texture. The corn



seeds (DEKALB, Corn Seed Blend, DKC51-20RIB) were acquired from Professor Kirkham's laboratory.



(a)

(b)

(c)

**Figure 2.22** Insert sensors and seeds in soil (top view). (a) lower sensor, (b) upper sensor, and (c) seeds

The two pots were placed outside the west wall of the BAE Instrumentation and Control Laboratory. The positions of the four sensors were marked outside the pots (Figure 2.22a and b).



(a)

(b)

(c)

**Figure 2.23** The outdoor tests (a) a pot with the four germinated corn seeds at the beginning of the experiment, (b) a pot after thinning to two seedlings, and (c) the two pots with one plant each at the end of the experiment

#### **2.4.2.3 Sampling schedule**

After the corn seeds were planted, the measurements were taken every two days. The measurement included 1) reading the dielectric spectral, 2) reading the VWC using the Decagon 5TE soil water sensor, and 3) taking soil samples within the pots near the sensor locations and sending them to the KSU Soil Testing Laboratory to measure the concentrations of  $\text{NO}_3$  and  $\text{NH}_4$ . Over the period of 27 July – 28 October 28 2019, 36 measurements were collected for each sensor, totaling 144 measurements for all four sensors.



**Figure 2.24 Some roots around the sensor area**

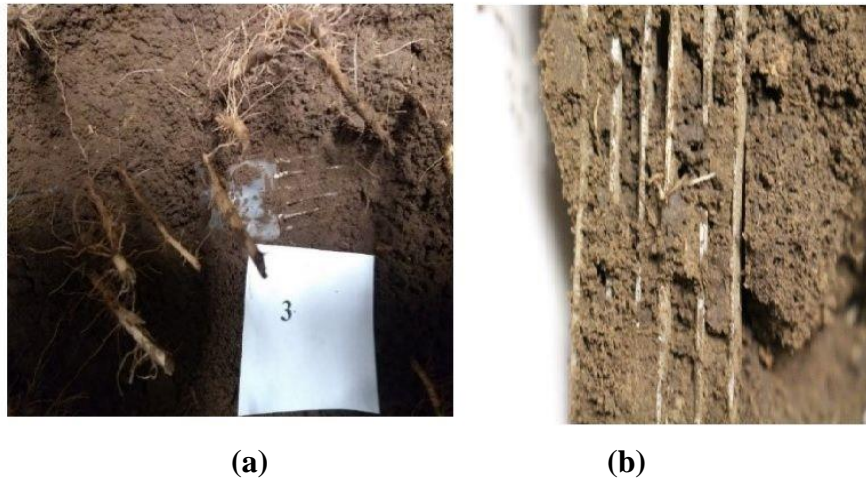
For the Decagon 5TE soil moisture and temperature sensor, a Campbell Scientific datalogger program was written to calculate the VWC (Appendix D). The two pots were irrigated as needed. Fertilizer (ammonium nitrate) was added to the pots every 8-10 days. During the measurements, many roots were observed growing along the perimeters of the pots (Figure 2.24). This probably indicated that the pot size was smaller than what the corn roots needed.

#### **2.4.2.4 Removal of soil from the pots**

On 7 November 2019, the soil from the pots was removed and the sensors checked.

It was very difficult to remove the soil surrounding the sensors because the soil was very dry. Many roots were around the sensors (Figure 2.25a) and some of them grew into the space

between the plates on the probe, also some porous area were observed between the plates (Figure 2.25b).



**Figure 2.25 (a) Sensors before being removed from the pots. (b) Roots and porous areas between plates**

#### ***2.4.2.4.1 Conditions of the plants upon removal from the pots***

The plants were dry and yellow in color. This is probably because that the temperature reached the freezing point in the two weeks before plant removal from the pots.

The weight of the corn ear in pot 1, where sensors 1 and 2 were located, was 176.6 g. In pot 2, where sensors 3 and 4 were located, it was 256.8 g.

The weight of the plant (excluding the roots) in pot 1 was 203.5 g. It was 160.4 g in pot 2.

The height of the plant in pot 1 was 170 cm. It was 174 cm in pot 2.

While the top soil layer was very dry, the moisture content of soil surrounding the sensor was 16.7-16.8 cm<sup>3</sup>/cm<sup>3</sup> in both pots.

#### ***2.4.2.4.2 Conditions of the sensors upon removal from the pots***

After the soil was removed, the conditions of the four sensors were examined.

##### **Sensor 1**

- No short circuits were observed between the odd and even plates on the probe.



- Plate #2 connection was broken.
- Roots were found in the space between the plates, with the largest being 0.25 mm in diameter.
- The sensor had moved down slightly, probably due to soil movement caused by heavy rains in August.

**Sensor 2:**

- No short circuits were observed between the odd and even plates on the probe.
- Plate #6 connection was broken.
- The sensor had moved down slightly due to soil movement, which was probably caused by heavy rains in August
- Roots were found in the space between the plates, with the largest being 0.25 mm in diameter.
- The space between plates was slightly enlarged on the lower portion of the probe.

**Sensor 3:**

- No short circuits were observed between the odd and even plates on the probe.
- The sensor had moved down 4 cm due to soil movement, which was probably caused by heavy rains in August.
- Plates 1, 2, 4 connections were broken.
- Roots were found in the spaces between the plates.

**Sensor 4:**

- No short circuits between the odd and even plates were observed on the probe.
- The sensor had moved down 5 cm due to soil movement, which was probably caused by heavy rains in August

- Plate #1 connection was broken.
- The space between plates was slightly enlarged on the lower portion of the probe.
- Roots were found in the space between the plates, with the largest being 0.25 mm in diameter.

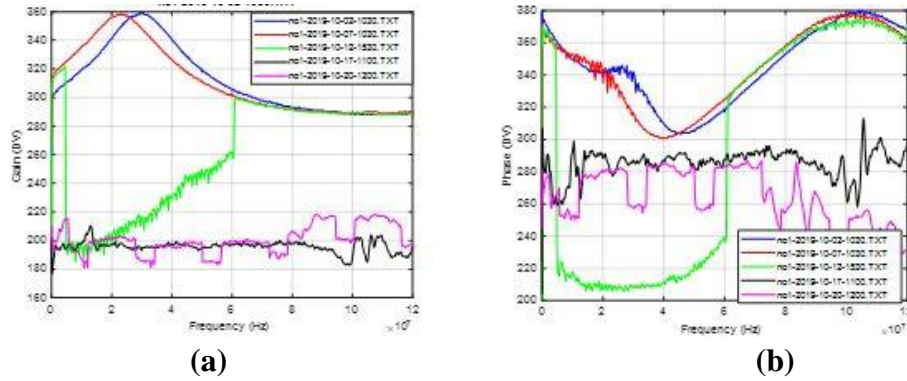
#### **2.4.2.5 Conditions of the sensor signals due to potential problems caused by broken connections**

The findings of broken connection were a concern. An effort was made to find 1) when the plates were disconnected and what was the cause? 2) How many data sets were affected by this? In review of the moving history of the pots, it was found that the wire disconnection occurred, on 5 October 2019, when the two pots were moved by the KSU mowing personnel without our knowledge. While the pots were moved twice by ourselves prior to October, it was done with extreme care. The unforeseen move by the mowing crew may not have been done with proper care. To verify this assumption, the spectral data collected before and after 5 October were compared. Good data was collected from 27 July to 4 October, which included the first 4 sets of 29 spectral data for each sensor, with a total of 116 observations.

The sensor signals were checked, and the following results were found:

##### ***Sensor 1***

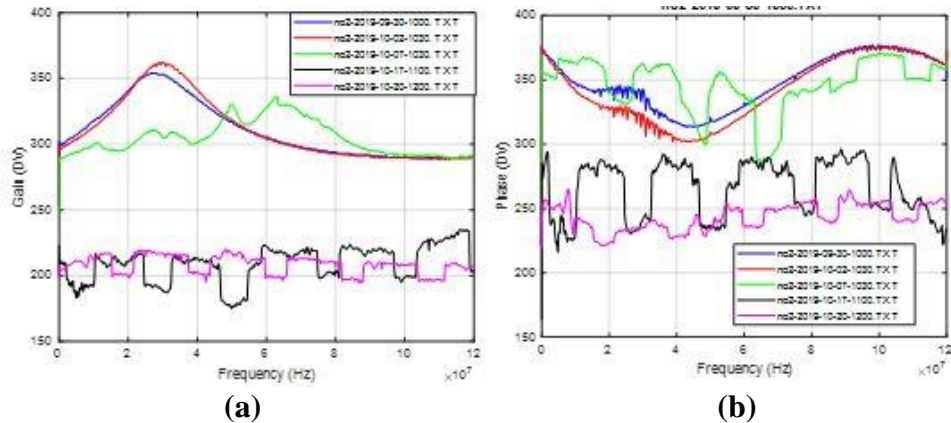
From Figure 2.26, it can be seen that the data collected on 2 October and 7 October were normal. The data collected on and after 12 October had problems.



**Figure 2.26 Dielectric spectral data collected from sensor 1 between 2 October to 20 October**  
**(a) Gain, (b) Phase**

### *Sensor 2*

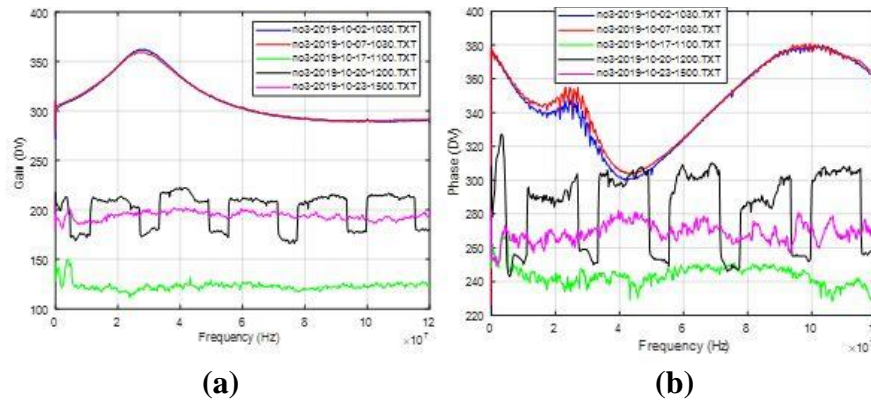
The data collected on 30 September and 2 October were normal. They became problematic on 7 October.



**Figure 2.27 Dielectric spectral data collected from sensor 2 between 2 October to 20 October**  
**(a) Gain, (b) Phase**

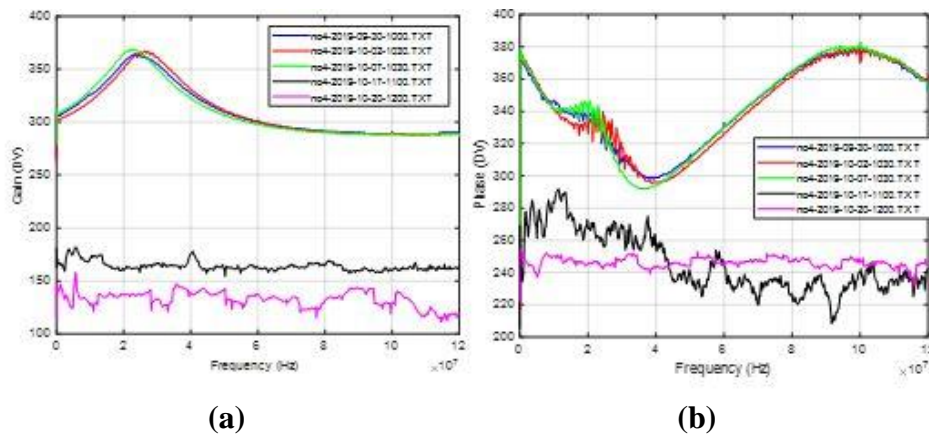
### *Sensor 3*

A similar trend can be observed for the dielectric spectral data collected by Sensors 3 and 4 (Figures 2.28 and 2.29).



**Figure 2.28 Dielectric spectral data collected from sensor 3 between 2 October to 20 October (a) Gain, (b) Phase.**

#### *Sensor 4*



**Figure 2.29 Dielectric spectral data collected from sensor 4 between 2 October to 20 October (a) Gain, (b) Phase**

After carefully studying the spectral data, we believe that the wire connection problem occurred on 5 October when the mowing crew moved the two pots. Accordingly, we decided to abandon all data taken after 5 October. This leaves us with good data collected from 27 July to 4 October, which included the first 4 sets of 29 spectral data for each sensor. For the four sensors we had a total of 116 spectral data were available.

## **Chapter 3 - Results and Discussion**

Results of the laboratory and outdoor experiments are reported in this chapter.

### **3.1 Laboratory experiment**

This section reports the performance of the FR sensor under laboratory conditions while (1) simultaneously measuring soil VWC and salinity, (2) simultaneously measuring soil VWC and one type of nitrogen fertilizer, and (3) simultaneously measuring VWC and two types of nitrogen fertilizers. These measurements were conducted under different soil bulk densities. Hence, the effect of soil bulk density on the measurement of soil VWC, salinity, and N-fertilizers was also studied.

When preparing the samples, water, salt, and fertilizers were added to soil in a weight basis – they were all weighed using a balance. However, except the situation when the sample was pressed using a piston to change the bulk density, the total volume of the sample was not changed. The water added to the sample at a given bulk density only changes the volumetric water content (VWC), not the gravitational water content (GWC). This is consistent to the water content that is interested by the field researchers – rainfall, irrigation, runoff, and leaching all can be conveniently quantified using changes in VWC. GWC becomes important only when compaction is the concern. On the other hand, the nature of the FR sensor also makes VWC one of the primary factors that affects the frequency responses. Within a given soil volume, the distributions of the solid, liquid, and gas phases determine the conductive and dielectric behaviors of the soil – conduction, polarization, and relaxation. Conductive ions in the liquid phase strongly affects the conductivity and more water (liquid phase) filling the pores (gas phase) improves the mobility of the ions, hence increasing the conductivity. In the same time, the composition of the liquid phase greatly affects the dielectric constant, hence changing the dielectric behavior. The mineral composition of the

solid phase strongly affects the conductivity. These behaviors all can be explained in a volume basis.

Variations in GWC definitely would change the conductive/dielectric behaviors of the soil as well. However, these variations represent a combined effects of several primary factors. With these considerations, we converted GWC to VWC in data analysis (Equation 1.3), and used soil bulk density as an additional factor to specifically study its effect on the measurement of VWC, salinity, and N-fertilizers.

### **3.1.1 Samples and PLS analysis of dielectric spectral data**

The total number of the spectral data collected for the VWC-salinity-bulk density tests was 75. The number for the VWC- bulk density-one fertilizer tests was 180, and for the VWC-bulk density-two fertilizers tests, the number was 810. For each spectral data, there were 635 predictor variables (635 frequencies). These dielectric spectral data were analyzed using the PLS regression method. The program was written in Matlab (Appendix E). In the PLS regression analysis, an important parameter is the number of PLS components used. The PLS regression method is capable of modeling a response variable with a large number of predictor variables, some of which are highly correlated or even collinear, as in the case of our dielectric spectral data. To make the model simple and more effective, the PLS regression method constructs new predictor variables, known as components, as linear combinations of the original predictor variables, to explain the observed variability in the predictor variables, taking the response variable into account. Thus, the PLS regression method often leads to models that are able to fit the response variable with fewer components.

A preliminary study was conducted to select the number of PLS components to establish the prediction models. Table 3.1 shows the effect of the number of components on the R-squared

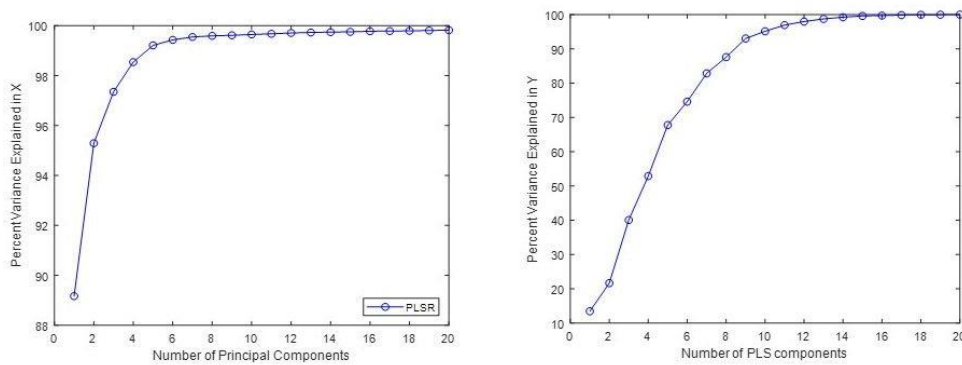
value of the regression models in measuring salinity. We can see that the highest R-squared values for both the training and validation data sets were reached when 12 components were used. However, after 9 components, the increase in the R-squared value for the validation data is insignificant. Thus, 9 components were selected for these data.

**Table 3.1. Effect of number of PLS components on salinity measurement using the phase data**

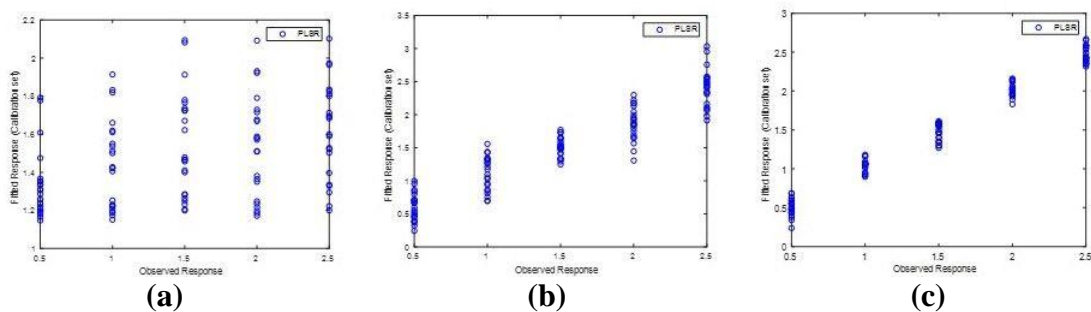
PLS components	R <sup>2</sup> for training	R <sup>2</sup> for validation
1	0.13	0.16
2	0.22	0.24
3	0.40	0.42
4	0.53	0.56
5	0.68	0.70
6	0.75	0.72
7	0.83	0.78
8	0.88	0.83
9	0.93	0.87
10	0.95	0.87
11	0.97	0.87
12	0.98	0.88
13	0.98	0.88
14	0.99	0.88
15	0.99	0.88

It was noticed that the “optimum” number of components is not the same for all spectral data sets. To identify the best number of components, we changed the Matlab program (Appendix G) to perform the PLS regression analysis using different numbers of PLS components and selected the number of the components that gave the largest R-squared value for the validation data sets. Figure 3.1 shows the percent variance explained in the predictor variables (X) and the percent variance explained in the response variable (Y). Figure 3.2 compares the fitted response with the observed response for the training data sets. In this case, 15 PLS components gave the best prediction. A similar comparison for the validation data sets is given in Figure 3.3. For the

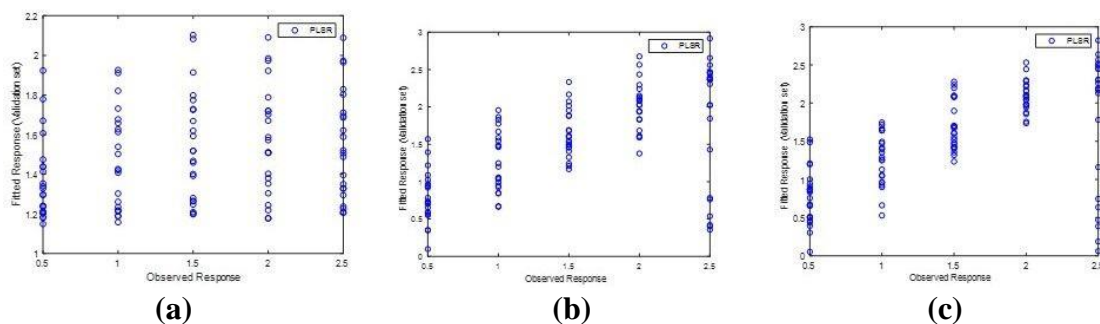
PLS regression analyses, the odd numbered data sets were used as the training data sets, and the even numbered the validation data sets. Thus, the ratio between the training and validation data sets was 50%:50%.



**Figure 3.1 Percent variance explained in predictor variables (X) and response variable (Y) vs number of PLS components**



**Figure 3.2 Fitted response (calibration) vs observed response for the training data sets with (a) 1, (b) 8, and (c) 15 PLS components**



**Figure 3.3 Fitted response (calibration) vs observed response for the validation data sets with (a) 1, (b) 8, and (c) 15 PLS components**



### 3.1.2 Results of the PLS regression analysis

#### 3.1.2.1 Soil salinity

During these tests, three factors - VWC, bulk density, and soil salinity - were controlled. The PLS calibration models for predicting these properties were established using the training data set. The models were then evaluated using the validation data set. The magnitude spectra alone, phase spectra alone, and both magnitude and phase spectra were used to establish the calibration models.

The results for predicting soil salinity, VWC, and bulk density are shown in Table 3.2. The R-squared values for VWC was 0.85-0.89 for the validation data sets. The values for salinity and bulk density were 0.82-0.85 and 0.87-0.90, respectively.

**Table 3.2 R-squared values for predicting soil salinity, VWC, and bulk density**

Data set	FR data used	VWC (cm <sup>3</sup> /cm <sup>3</sup> )	Bulk density (g/cm <sup>3</sup> )	Salinity (mS/cm)
Training	Phase	0.92	0.93	0.88
	Magnitude	0.94	0.94	0.92
	Phase and magnitude	0.92	0.92	0.89
Validation	Phase	0.88	0.87	0.83
	Magnitude	0.85	0.90	0.82
	Phase and magnitude	0.89	0.89	0.85

#### 3.1.2.2 Effect of bulk density

Tables 3.3-3.4 compare the performance of the PLS models in predicting VWC and soil salinity using magnitude, phase, and both magnitude and phase spectral data obtained with combinations of all VWC and salinity levels at different bulk densities. The R<sup>2</sup> values achieved by the models in predicting VWC and salinity at different bulk densities were similar. By comparing Tables 3.3-3.4 with Table 3.2, it can be seen that, when bulk density did not change, we obtained

a better model to predict VWC and salinity. However, in general, the effect of varying bulk density did not significantly reduce the R-squared values for the VWC and salinity predictions.

From Tables 3.3 and 3.4, it also can be observed that the  $R^2$  values increased slightly with increasing bulk density. This probably is related to the decrease of air volume in the soil at higher densities.

**Table 3.3 R-squared values for VWC prediction at different bulk densities**

Data set	FR data used	Bulk density (g/cm <sup>3</sup> )		
		0.8	1.0	1.2
Training	Phase	0.99	0.99	0.99
	Magnitude	0.99	0.99	0.99
	Phase and magnitude	0.99	0.99	0.99
Validation	Phase	0.92	0.94	0.96
	Magnitude	0.90	0.93	0.94
	Phase and magnitude	0.93	0.96	0.96

**Table 3.4 R-squared values for soil salinity prediction at different bulk densities**

Data set	FR data used	Bulk density (g/cm <sup>3</sup> )		
		0.8	1.0	1.2
Training	Phase	0.98	0.99	0.999
	Magnitude	0.98	0.999	0.99
	Phase and magnitude	0.98	0.99	0.98
Validation	Phase	0.87	0.90	0.95
	Magnitude	0.80	0.92	0.94
	Phase and magnitude	0.89	0.91	0.95

### 3.1.2.3 One type of N-fertilizer

#### 3.1.2.3.1 Ammonium sulfate

The results for predicting ammonium sulfate, bulk density, and VWC are given in Table

3.5. It seems that using the magnitude spectra alone gave better results.

**Table 3.5 R-squared value for VWC, ammonium sulfate, and bulk density**

Data set	FR data used	VWC (cm <sup>3</sup> /cm <sup>3</sup> )	Bulk density (g/cm <sup>3</sup> )	Ammonium sulfate (ppm)
Training	Phase	0.88	0.82	0.95
	Magnitude	0.97	0.88	0.98
	Phase and magnitude	0.86	0.82	0.96
Validation	Phase	0.76	0.70	0.84
	Magnitude	0.85	0.73	0.84
	Phase and magnitude	0.82	0.72	0.84

#### 3.1.2.3.2 Effect of bulk density

Tables 3.6-3.7 give the results of predicting ammonium sulfate and VWC at two different bulk densities. It can be seen that, in general, the prediction results are slightly better than those obtained with varying bulk densities (Table 3.5). However, the difference is not significant. At the same time, the R-squared values obtained at a higher bulk density are slightly higher than those at a lower bulk density. Again, this may be explained by amount of air in the soil. Air has a very different permittivity from other materials. More air in soil can cause changes in permittivity.

**Table 3.6 R-squared value for ammonium sulfate prediction at different bulk densities**

Data set	FR data used	Bulk density (g/cm <sup>3</sup> )	
		1.0	1.2
Training	Phase	0.98	0.98
	Magnitude	0.98	0.99
	Phase and magnitude	0.98	0.98
Validation	Phase	0.77	0.87
	Magnitude	0.81	0.82
	Phase and magnitude	0.79	0.85

**Table 3.7 R-squared value for VWC prediction at different bulk densities**

Data set	FR data used	Bulk density (g/cm <sup>3</sup> )	
		1.0	1.2
Training	Phase	0.99	0.96
	Magnitude	0.99	0.99
	Phase and magnitude	0.99	0.97
Validation	Phase	0.87	0.92
	Magnitude	0.90	0.92
	Phase and magnitude	0.88	0.95

**3.1.2.3.3 Ammonium nitrate**

The results for predicting ammonium nitrate, VWC, and bulk density are shown in Table 3.8. Tables 3.9-3.10 give the prediction results for ammonium nitrate and VWC at constant densities. Finding from these tables are similar to those for ammonium sulfate.

**Table 3.8 R-squared values for ammonium nitrate, bulk density, and VWC**

Data set	FR data used	VWC (cm <sup>3</sup> /cm <sup>3</sup> )	Bulk density (g/cm <sup>3</sup> )	Ammonium Nitrate (ppm)
Training	Phase	0.97	0.90	0.89
	Magnitude	0.98	0.95	0.97
	Phase and magnitude	0.97	0.91	0.87
Validation	Phase	0.92	0.71	0.81
	Magnitude	0.89	0.74	0.85
	Phase and magnitude	0.92	0.73	0.81

#### 3.1.2.3.4 Effect of bulk density

**Table 3.9 R-squared values for ammonium nitrate prediction at different bulk densities**

Data set	FR data used	Bulk density (g/cm <sup>3</sup> )	
		1.0	1.2
Training	Phase	0.93	0.99
	Magnitude	0.99	0.99
	Phase and magnitude	0.95	0.98
Validation	Phase	0.83	0.87
	Magnitude	0.80	0.85
	Phase and magnitude	0.84	0.86

**Table 3.10 R-squared values for VWC prediction at different bulk densities**

Data set	FR data used	Bulk density (g/cm <sup>3</sup> )	
		1.0	1.2
Training	Phase	0.98	0.99
	Magnitude	0.99	0.99
	Phase and magnitude	0.98	0.99
Validation	Phase	0.93	0.93
	Magnitude	0.90	0.90
	Phase and magnitude	0.92	0.94

#### 3.1.2.3.5 Urea fertilizer

The R-squared values in predicting urea concentration are slightly lower. This is probably due to the fact that both ammonium nitrate and ammonium sulfate are inorganic salts that are formed by anions and cations, which greatly affect the conductivity of the soil, especially when more water exists in the soil. Urea, on the other hand, is an organic matter. While it may contribute more to the dielectric behavior of the soil, its influence on conductivity is not strong.

**Table 3.11 R-squared value for urea, bulk density, and VWC**

Data set	FR data used	VWC (cm <sup>3</sup> /cm <sup>3</sup> )	Bulk density (g/cm <sup>3</sup> )	Urea (ppm)
Training	Phase	0.85	0.86	0.88
	Magnitude	0.91	0.92	0.97
	Phase and magnitude	0.86	0.86	0.90
Validation	Phase	0.63	0.56	0.71
	Magnitude	0.69	0.65	0.68
	Phase and magnitude	0.66	0.59	0.79

### 3.1.2.3.6 Effect of bulk density

Trends for the influence of bulk density on the measurements of VWC and urea can be observed in Tables 3.12-3.13.

**Table 3.12 R-squared values for urea prediction at different bulk densities**

Data set	FR data used	Bulk density (g/cm <sup>3</sup> )	
		1.0	1.2
Training	Phase	0.94	0.96
	Magnitude	0.88	0.99
	Phase and magnitude	0.95	0.96
Validation	Phase	0.56	0.77
	Magnitude	0.72	0.73
	Phase and magnitude	0.66	0.85

**Table 3.13 R-squared values for VWC prediction at different bulk densities**

Data set	FR data used	Bulk density (g/cm <sup>3</sup> )	
		1.0	1.2
Training	Phase	0.97	0.99
	Magnitude	0.99	0.99
	Phase and magnitude	0.96	0.98
Validation	Phase	0.82	0.91
	Magnitude	0.87	0.88
	Phase and magnitude	0.83	0.92

#### ***3.1.2.4 Two types of N-fertilizers***

For the two-fertilizer tests, we applied different amounts of two types of N-fertilizers in the soil samples, and established PLS prediction models to predict each type of fertilizer. The two types of N-fertilizers were chosen from the three studied in the experiment: ammonium nitrate, ammonium sulfate, and urea.

##### ***3.1.2.4.1 Ammonium sulfate and ammonium nitrate***

When both ammonium nitrate and ammonium sulfate were applied to the soil, the model predicting VWC still performed well, resulting in high R-squared values for both the training and validation data. However, the model predicting bulk density seemed to have some difficulties unless only the phase data were used to establish the model. However, when the two types of N-fertilizer were measured individually, both the magnitude and phase spectral data need to be used to obtain good results.

**Table 3.14 R-squared values for ammonium nitrate + ammonium sulfate, bulk density, and VWC**

Data set	FR data used	VWC (cm <sup>3</sup> /cm <sup>3</sup> )	Bulk density (g/cm <sup>3</sup> )	Ammonium Sulfate (ppm)	Ammonium Nitrate (ppm)	Ammonium Sulfate with Ammonium Nitrate (ppm)
Training	Phase	0.98	0.91	0.84	0.93	0.90
	Magnitude	0.95	0.95	0.88	0.87	0.59
	Phase and magnitude	0.98	0.93	0.85	0.91	0.89
Validation	Phase	0.91	0.80	0.40	0.62	0.44
	Magnitude	0.92	0.76	0.39	0.59	0.39
	Phase and magnitude	0.93	0.82	0.45	0.67	0.50

#### 3.1.2.4.2 Ammonium sulfate and urea

When urea was added with ammonium sulfate in the soil, the prediction results, in general, worsened. Although the VWC could still be measured with an R-squared value of higher than 0.8, measurement of the bulk density became more difficult. Among the two fertilizers, urea could be measured with a higher R-squared value, but it was unsuccessful in measuring ammonium sulfate. While still believing that the ammonium salts mainly influence the conductive behavior of the soil, the effect of organic urea on permittivity was still unclear in our minds. This is a topic that needs to be researched in the future.

**Table 3.15 R-squared values for ammonium sulfate + urea, bulk density, and VWC**

Data set	FR data used	VWC (cm <sup>3</sup> /cm <sup>3</sup> )	Bulk density (g/cm <sup>3</sup> )	Ammonium Sulfate (ppm)	Urea (ppm)	Ammonium Sulfate with Urea (ppm)
Training	Phase	0.87	0.59	0.34	0.84	0.82
	Magnitude	0.91	0.78	0.36	0.84	0.86
	Phase and magnitude	0.87	0.94	0.34	0.84	0.82
Validation	Phase	0.82	0.53	0.21	0.73	0.74
	Magnitude	0.80	0.62	0.24	0.72	0.74
	Phase and magnitude	0.83	0.57	0.28	0.73	0.75

#### 3.1.2.4.3 Ammonium nitrate and urea

When different amounts of ammonium nitrate and urea were applied in the soil, the performance of all prediction models, perhaps except the ones for VWC and bulk density, was reduced (Table 3.16). It can also be noticed that, among the two fertilizers, urea still had better prediction results than ammonium nitrate. At this time, we do not have an explanation for this finding.



**Table 3.16 R-squared values for ammonium nitrate + urea, bulk density, and VWC**

Data set	FR data used	VWC (cm <sup>3</sup> /cm <sup>3</sup> )	Bulk density (g/cm <sup>3</sup> )	Ammonium Nitrate (ppm)	Urea (ppm)	Ammonium Nitrate with Urea (ppm)
Training	Phase	0.91	0.96	0.58	0.92	0.88
	Magnitude	0.94	0.87	0.19	0.69	0.69
	Phase and magnitude	0.92	0.97	0.60	0.96	0.88
Validation	Phase	0.86	0.53	0.15	0.46	0.29
	Magnitude	0.85	0.64	0.09	0.29	0.22
	Phase and magnitude	0.87	0.66	0.15	0.50	0.34

## 3.2 Outdoor experiment

### 3.2.1 Sample groups

In total, 116 effective dielectric spectrum samples were taken by the four sensors. Each sensor had 29 samples. Sensors 1 and 2 were buried in pot 1 and sensors 3 and 4 in pot 2. Sensors 1 and 3 were buried at the same depth (12.5 cm from the soil surface) and sensors 2 and 4 were 25 cm from the surface. Five sets of samples were taken from each sensor. Set 1 samples were taken from 27 July to 12 August. Eight samples were taken from each sensor. Thus, Set 1 had 32 samples. Sets 2-4 each had 28 samples with 7 from each sensor. Set 2 samples were collected from 14 August to 29 August. Set 3 samples were collected from 2 September to 14 September, and Set 4 (28 samples) from 17 September to 2 October.

An additional set of samples (Set 5) were collected from 7 October to 28 October. However, this set of samples could not be used in the analysis because of broken wire connections, probably caused by an unexpected movement of the pots by the university's lawn mowing crew, which made these data unusable.

For each sample set, the VWC, nitrate-N, and ammonium-N were measured. The VWC was measured using a Decagon 5TE soil moisture sensor and all the samples were sent to the KSU

Soil Testing Laboratory to measure the nitrate-N and ammonium-N. For convenience in writing and reading, the nitrate-N and ammonium-N will be referred to as “NO<sub>3</sub>” and “NH<sub>4</sub>” hereafter in this dissertation.

Two Matlab programs were developed to analyze the data. The first prepared the data for analysis (Appendix F). It gave the flexibility of sample set selection and various ways of data preprocessing. The second performed the PLS regression analysis (Appendix G). It provided the flexibility of selecting the optimum number of PLS components and the number of most significant frequencies used in the regression.

### 3.2.2 Sample range

Table 3.17 gives the ranges of the variables measured (VWC, nitrate-N, and ammonium-N) for each sample set.

**Table 3.17 Ranges of NO<sub>3</sub>, NH<sub>4</sub>, and VWC for each sample set**

Sensor	Set	Period	File numbers	VWC (cm <sup>3</sup> /cm <sup>3</sup> )	Nitrate-N (ppm)	Ammonium-N (ppm)
1	1	7/27-8/12	1-8	12.7-28.9	17.8-59.6	5.6-10.4
	2	8/14-8/29	33-39	21.0-31.4	17.2-48.3	3.4-16.4
	3	9/2-9/14	61-67	21.2-30.5	1.2-9.0	3.3-6.3
	4	9/17-10/2	89-95	21.3-24.8	1.8-51.3	4.4-11.2
2	1	7/27-8/12	9-16	11.6-37.3	19.2-85.8	5.7-9.2
	2	8/14-8/29	40-46	29.6-36.2	12.7-92.4	3.5-12.4
	3	9/2-9/14	68-74	23.6-37.7	1.1-6.2	3.6-7.2
	4	9/17-10/2	96-102	15.5-26.0	2.0-4.5	4.9-9.1
3	1	7/27-8/12	17-24	11.2-30.8	17.6-76.0	6.4-10.1
	2	8/14-8/29	47-53	24.4-33.0	28.3-42.7	2.9-12.1
	3	9/2-9/14	75-81	24.0-36.0	0.8-10.2	3.3-7.0
	4	9/17-10/2	103-109	19.9-23.7	1.3-8.2	4.5-8.7
4	1	7/27-8/12	25-32	13.5-33.8	16.8-116.3	6.4-11.0
	2	8/14-8/29	54-60	29.0-36.4	21.4-102.5	4.5-15.4
	3	9/2-9/14	82-88	23.0-37.4	1.0-12.2	4.0-11.1
	4	9/17-10/2	110-116	16.8-24.5	1.2-7.4	4.1-8.3

From Table 3.17, it can be seen that the ranges of the variables to be measured within different sample sets are different. For example, the range of NO<sub>3</sub> varied from 1.1-6.2 ppm (Sensor 2, sample set 3) to 21.4-102.5 ppm (Sensor 4, sample set 2). Obviously, the prediction model trained using the data in a certain sample set may not allow correct prediction for a validation sample set if the coverages of the training and validation sample sets are very different. To describe the matching between the ranges of a training sample set and a validation sample set, we defined a “Coverage Index” (“CI”) as follows.

### 3.2.3 Coverage Index

Coverage Index (CI) is an indicator of the convergence of the range of the training data with that of the validation data. It can be calculated from multiplying the coverage ratio (CR) by un-coverage ratio (UR)

$$CI = CR * UR \quad (3.1)$$

$$CR = c_v / r_t \quad (3.2)$$

where  $c_v$  is the portion of the validation data range ( $c_v$ ) covered by the training data range

$r_t$  is the range of the training data (  $t_{\max} - t_{\min}$  ) (Figure 3.4).

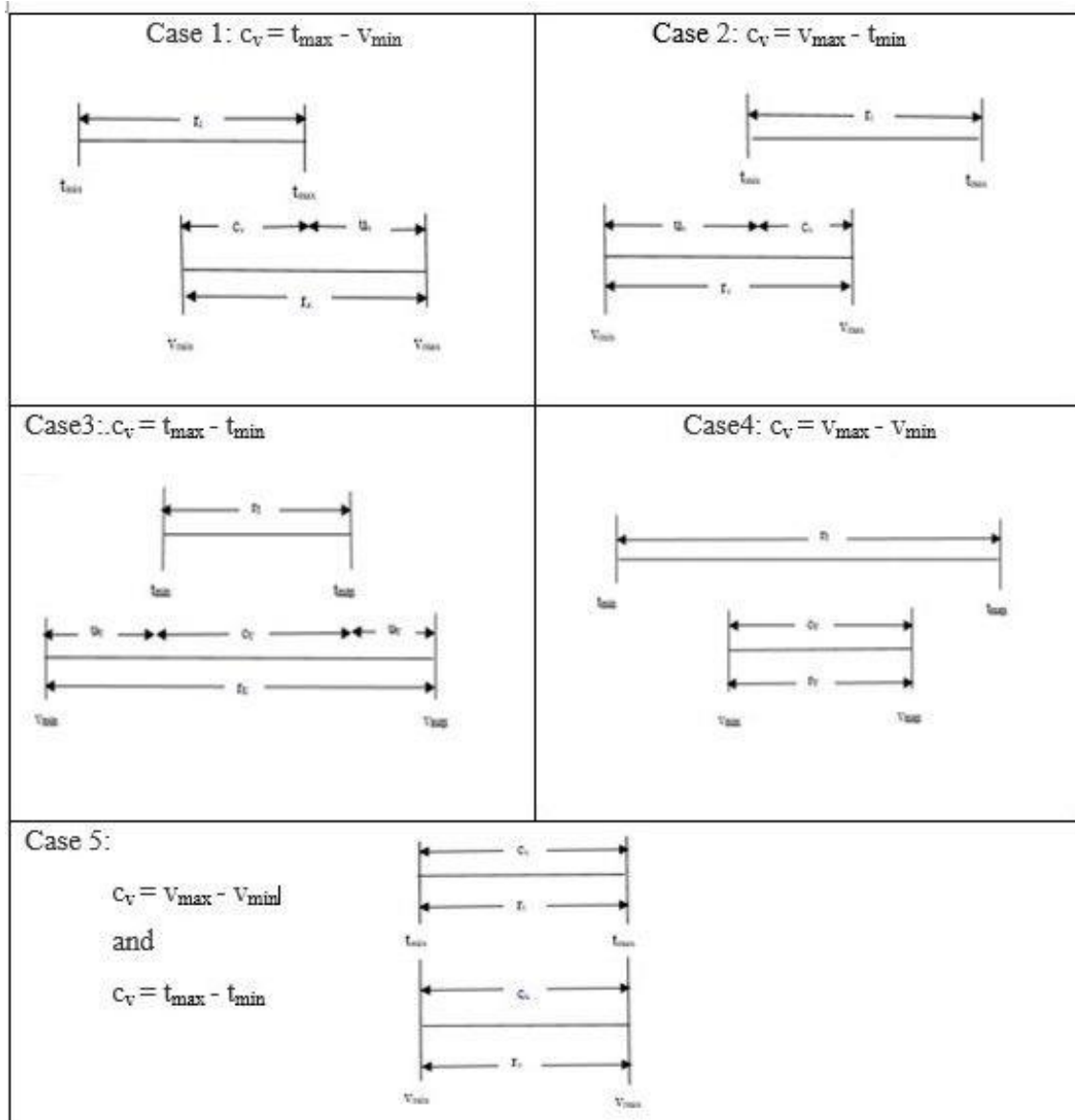
$$UR = c_v / r_v \quad (3.3)$$

where  $r_v$  is the range of the validation data (  $v_{\max} - v_{\min}$  )

Therefore

$$CI = (c_v / r_t) (c_v / r_v) = c_v^2 / (r_t * r_v) \quad (3.4)$$

The coverage of validation data range ( $c_v$ ) is different depending on both the training and validation data and there are 5 cases (Figure 3.4):



**Figure 3.4 Five cases of validation data coverage**

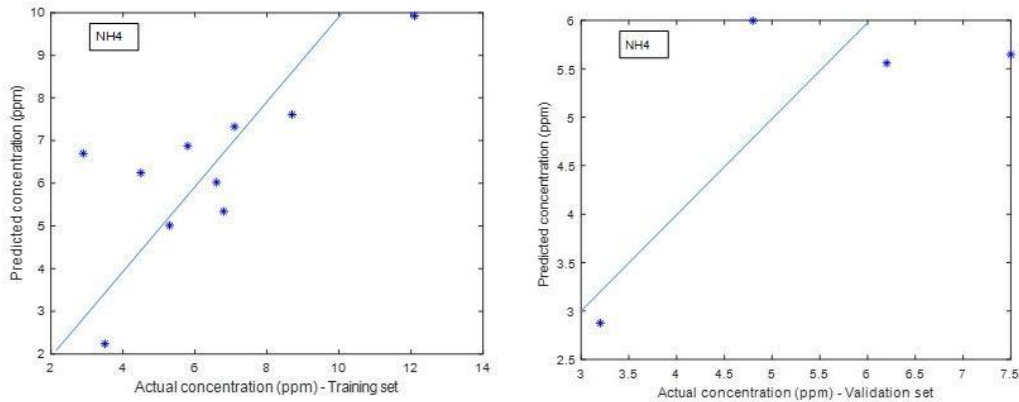
Obviously, case 5 gives the best matching between the training and validation data ranges and it has a perfect CI value of 1.0.

CI has a strong effect on the PLS regression models. Table 3.18 compares the regression models trained and validated using 2, 3, and 4 sets of data for Sensor 3 in predicting the concentration of  $\text{NH}_4$ . The range of the  $\text{NH}_4$  concentration was 2.9-12.1 ppm. However, among the 29 samples, 28 had  $\text{NH}_4$  concentrations between 2.9-10.1 ppm, and only one sample (sample

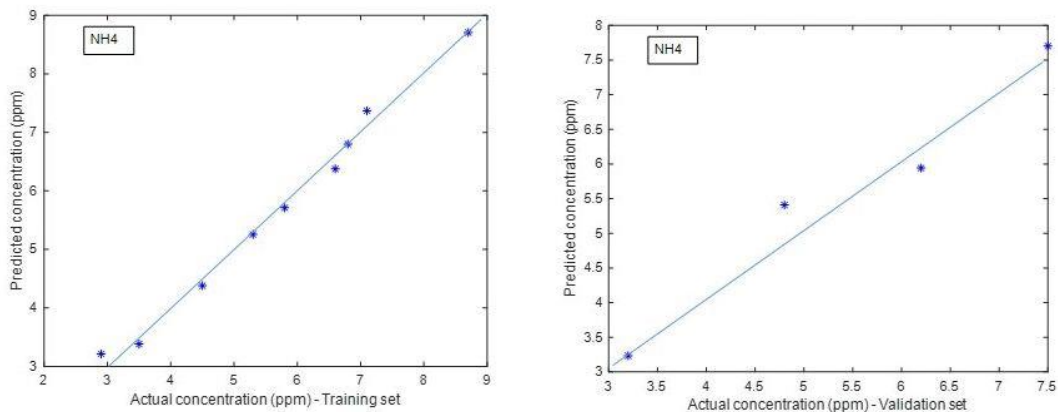
8) had a concentration of 12.1 ppm. Including this sample in the PLS regression analysis strongly affected the CI, and in-turn affected the accuracy of the trained PLS models. Figures 3.5-10 demonstrate the improvement in prediction of NH<sub>4</sub> after sample 8, the out-of-range sample, was removed from the training data sets for the 2-, 3-, and 4-set data prediction models for sensor 3. From Table 3.18, the cause-effect relationship between CI and R<sup>2</sup> value is clearly displayed by the prediction results.

**Table 3.18 The effect of Coverage Index (CI) on the performance of prediction models**

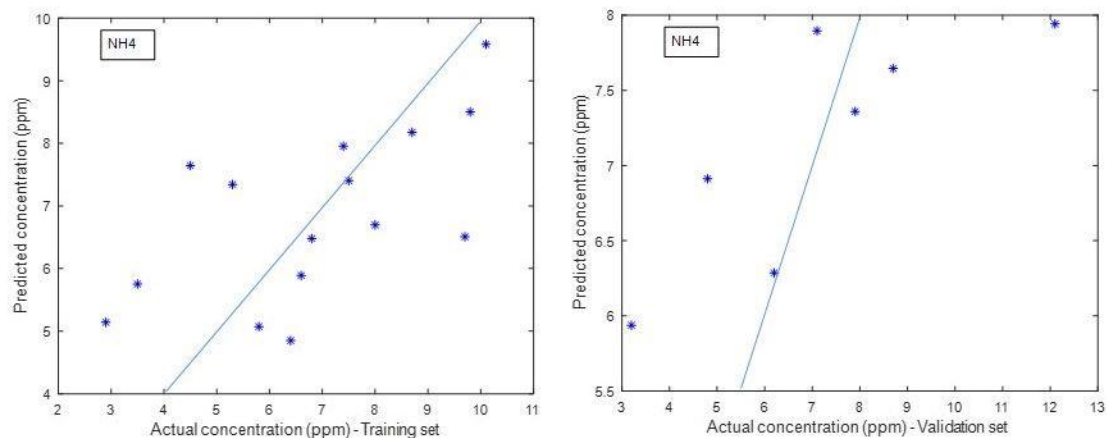
Model parameter		Out-of-range sample (sample 8) included	Out-of-range sample (sample 8) removed
Two sample sets	CI	0.47	0.74
	R <sup>2</sup> (training/validation)	0.55/0.48	0.99/0.95
	RMSE (training/validation)	1.69/1.16	0.17/0.35
Three sample sets	CI	0.74	0.76
	R <sup>2</sup> (training/validation)	0.39/0.37	1.00/0.94
	RMSE (training/validation)	1.68/2.11	0.03/0.47
Four sample sets	CI	0.75	0.91
	R <sup>2</sup> (training/validation)	1.00/0.64	0.76/0.76
	RMSE (training/validation)	0.00/1.24	0.97/1.01



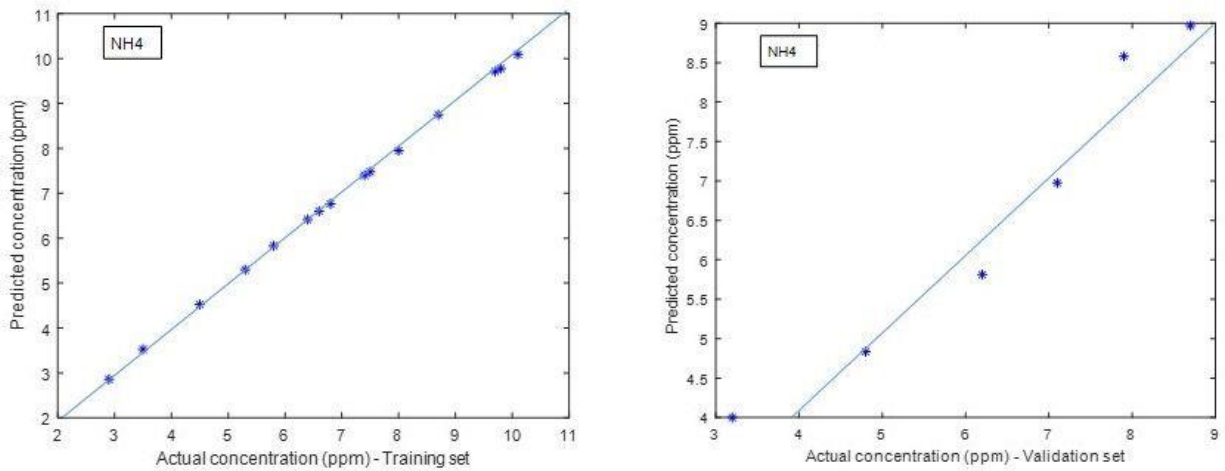
**Figure 3.5 Performance of NH<sub>4</sub> prediction model using two sets of samples for sensor 3 with an out-of-range sample (sample 8) included**



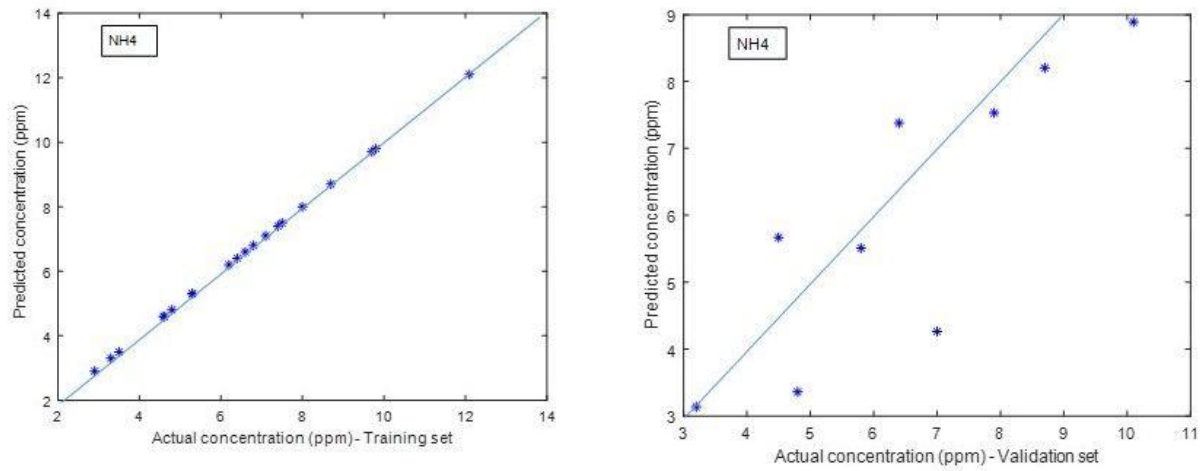
**Figure 3.6 Performance of NH4 prediction model using two sets of samples for sensor 3 with an out-of-range sample (sample 8) removed**



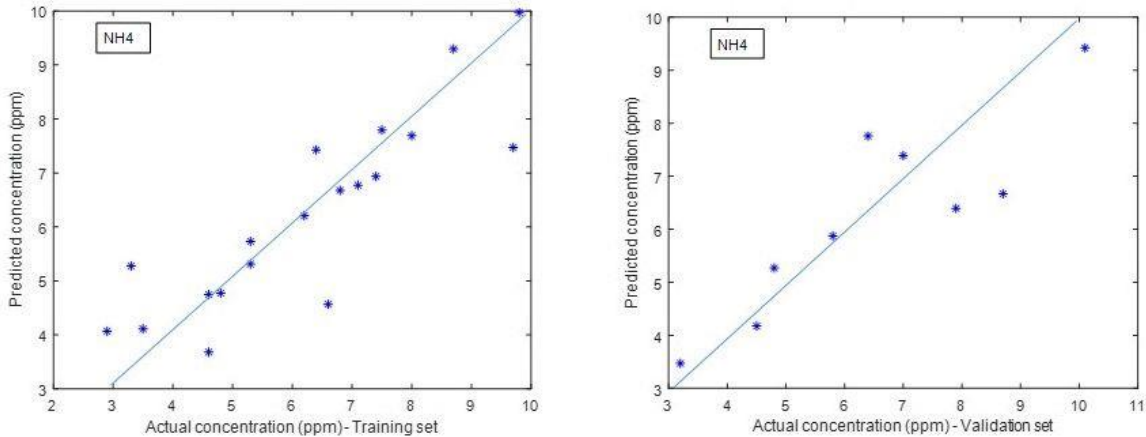
**Figure 3.7 Performance of NH4 prediction model using three sets of samples for sensor 3 with an out-of-range sample (sample 8) included**



**Figure 3.8 Performance of NH<sub>4</sub> prediction model using three sets of samples for sensor 3 with an out-of-range sample (sample 8) removed**



**Figure 3.9 Performance of NH<sub>4</sub> prediction model using four sets of samples for sensor 3 with an out-of-range sample (sample 8) included**



**Figure 3.10 Performance of NH<sub>4</sub> prediction model using four sets of samples for sensor 3 with an out-of-range sample (sample 8) removed**

### 3.2.4 Matlab programs for data preparation and PLS regression analysis

The computer programs handling the processing and analytics of the dielectric spectral data were written in Matlab. There are two programs. The first prepares the data for PLS regression analysis and the second searches the best combination of the PLS components and signature frequencies, performs the PLS regression, and reports the results in various ways.

The data preparation program provides options for data selection and data preprocessing. For training and validation data selection, the options are:

- Separately selecting data sets for training and validation,
- Selecting certain data sets for training and validation and the program will randomly separate the data into training and validation data sets based on a user-selected ratio (the recommended ratio is 70%:30%).
- Selecting certain data sets for training and validation and the program will sort the selected data sets by the values of the response variable and separate the sorted data sets into training and validation data sets based on a user-selected ratio (the recommended ratio is



70%:30%). The third option was made available for spectral data that have insufficient range and low distribution uniformity for the response variable.

The options for data preprocessing include:

- Selection of the response variable: For the outdoor experiment, the response variable can be VWC, nitrate-N, and ammonium-N. The user may select to run one, two, or three response variables.
- Selection of the predictor variables: Available options are 1) only use the magnitude spectra, 2) only use the phase spectral data, and 3) use both the magnitude and phase spectral data.
- Selection of the frequencies of the spectral data to be used in the PLS regression analysis: Options available include 1) User defines the maximum frequency and the program will eliminate all the spectral data beyond this frequency. 2) Compression of the spectral data by averaging them with a user-selected number of frequencies. For example, if the user selects to compress the data to  $\frac{1}{4}$  of the original data, spectral data of every 4-frequency group would be averaged.
- Selection to smooth the spectral data: Options include 1) The user can select to average the three runs in a spectral data set. This would reduce the number of data sets to  $\frac{1}{3}$ . 2) The user may select smoothing the spectral through moving-averaging. This would not reduce the length of the spectral data but would make the spectral data look smoother, and it would eliminate some abrupt changes in the spectral data (Appendix F).

The program that handles the PLS regression analysis considered the following important principles: 1). the number of PLS components have an important role in deciding the performance of the regression model. An insufficient number of the PLS components could not fully represent the variance in the predictor variables and corresponding response variables. This would result in

under fitting of the prediction models and inaccurate prediction for the validation data. On the other hand, an excessive number of the PLS components may result in overfitting, hence less accurate prediction for the validation data. 2). The Variable Importance in Projection (VIP) scores estimate the importance of each predictor variable in the projection used in a PLS model and is often used for variable selection. A variable with a VIP score close to or greater than 1 (one) can be considered important in a given model. Variables with VIP scores significantly less than 1 are less important and might be good candidates for exclusion from the model.

Based on these principles, the program that handles the PLS regression analysis included the following functions:

- Calculate the ranges of the response variables of the training and validation data sets and the Coverage Index (CI). This helps the user decide whether the data sets selected for training and validation are appropriate.
- Allow the user to select the maximum number of PLS components N and calculate the R-squared values for the validation data set using all available predictors (all frequencies) and 1 to N PLS components.
- Select the number of PLS components that produce the highest R-squared value and run the PLS regression analysis using this number of components, again to derive the VIP scores of all predictor variables (frequencies).
- Sort the predictor variables (frequencies) based on the VIP scores to determine the importance of the predictor variables (frequencies).
- Allow the user to select the maximum number of the predictor variables (frequencies) M and run the PLS regression analysis using the optimum number of the PLS components and 1-M predictor variables (frequencies) selected from the top of the sorted VIP list.

The optimum number of the predictor variables (frequencies) and the optimum number of the PLS components that produced the maximum R-squared value (or the minimum RMSE) for the validation data set are then reported in the results.

- The frequencies within the optimum number of predictor variables (frequencies) are then selected as the “signature frequencies” (Appendix F).

### **3.2.5 Prediction results**

First, the sample data obtained from two, three, and four sets of individual sensors were used to train and validate the PLS models for predicting VWC, NO<sub>3</sub>, and NH<sub>4</sub>. Then, to study the possibility of establishing “universal” prediction models for these measurements, sample data across sensors were combined to train and validate the models. The results are summarized in this section.

#### **3.2.5.1 Models for individual sensors**

The performances of PLS models established to predict VWC, NO<sub>3</sub>, and NH<sub>4</sub> for individual sensors are given in Tables 3.19-3.21.

##### **Sensor 1**

The performance of the prediction models established for sensor 1 in predicting VWC using different combinations of the training and validation data is given in Table 3.19. It was noticed that the first sample for VWC was taken before fertilizer and water were added to pot 1, Thus, this sample had a very low VWC (12.7%), which resulted in low CI values (0.37-0.47). Removing this sample improved the CI and, as a result, improved the R<sup>2</sup> values and RMSE, as can be seen in Table 3.19

**Table 3.19 Performance of PLS prediction models for VWC established using various combinations of the sample data from sensor1**

Sample sets used	CI	R <sup>2</sup> (training/validation)	RMSE (cm <sup>3</sup> /cm <sup>3</sup> ) (training/validation)
1+2	0.41	1.00/0.97	0.00/0.71
1+3	0.39	1.00/0.95	0.27/0.75
1+4	0.37	0.92/0.91	1.08/0.80
2+3	0.79	0.84/0.52	1.21/2.13
2+4	0.69	0.96/0.87	0.63/0.96
3+4	0.63	0.84/0.82	1.04/0.91
1+2+3	0.47	0.92/0.17	1.20/3.05
1+2+3*	0.74	1.00/0.55	0.01/1.85
1+2+4	0.46	0.91/0.45	1.23/2.44
1+2+4*	0.69	0.74/0.73	1.58/1.28
1+3+4	0.39	0.95/0.55	0.87/1.98
1+3+4*	0.71	0.85/0.83	1.03/0.88
2+3+4	0.71	0.90/0.66	1.02/1.46
1+2+3+4	0.51	0.39/0.36	3.03/2.51
1+2+3+4*	0.89	0.80/0.69	1.26/1.74

\*with an out-of-range sample removed

For NO<sub>3</sub>, we found an out-of-range sample that had a NO<sub>3</sub> concentration value of 51.3 ppm, when we used sample sets 3+4, whereas the range of remaining samples was 1.2-9.1 ppm. From Table 3.20 it can be seen that removing this sample improved the prediction results.

**Table 3.20 Performance of PLS prediction models for NO<sub>3</sub> established using various combinations of the sample data from sensor1**

Sample sets used	CI	R <sup>2</sup> (training/validation)	RMSE(ppm) (training/validation)
1+2	0.82	1.00/0.52	0.44/11.3
1+3	0.91	0.86/0.74	6.84/11.3
1+4	0.86	0.72/0.63	10.2/13.3
2+3	0.54	0.83/0.75	6.60/4.96
2+4	0.72	0.92/0.82	4.71/5.96
3+4	0.15	0.38/0.17	11.2/3.25
3+4*	0.98	0.63/0.62	1.57/1.84
1+2+3	0.91	0.92/0.82	5.04/8.09
1+2+4	0.86	0.89/0.87	5.61/6.93
1+3+4	0.91	1.00/0.77	0.00/9.89
2+3+4	0.51	0.99/0.58	1.40/5.99
1+2+3+4	0.91	0.99/0.71	1.59/9.89

\*with an out-of-range sample removed

The prediction results using data from two data sets and four data sets for training and validation are shown in Table 3.21. The best results for predicting NH<sub>4</sub> were found when using sets 1+2, sets 1+3, and sets 1+2+3+4, which gave R squared value 0.99, 1.00, and 0.91 for training and 0.93, 0.83, and 0.85 for validation data, respectively. For NH<sub>4</sub>, set 2 samples had the largest range (3.4 -16.4 ppm) whereas set 3 recorded the lowest NH<sub>4</sub> concentration (3.3 ppm). After removing this out-of-range sample, prediction model for set 3+4 was improved.

**Table 3.21 Performance of PLS prediction models for NH<sub>4</sub> established using various combinations of the sample data from sensor 1**

Sample sets used	CI	R <sup>2</sup> (training/validation)	RMSE(ppm) (training/validation)
1+2	0.39	0.99/0.93	0.15/1.03
1+3	0.72	1.00/0.83	0.00/0.89
1+4	0.77	0.96/0.64	0.40/1.38
2+3	0.34	0.99/0.36	0.31/1.28
2+4	0.59	1.00/0.10	0.00/2.50
2+4*	0.65	0.99/0.64	0.03/1.05
3+4	0.25	0.52/0.15	1.48/1.92
3+4*	0.35	0.99/0.68	0.14/0.41
1+2+3	0.53	0.82/0.67	0.85/2.23
1+2+4	0.49	0.78/0.71	1.02/2.02
1+3+4	0.75	0.42/0.17	1.52/2.05
1+2+4*	0.65	0.44/0.44	1.50/1.51
2+3+4	0.37	0.99/0.16	0.35/1.58
2+3+4*	0.61	1.00/0.55	0.13/0.99
1+2+3+4	0.60	0.91/0.85	0.86/0.87

\*with an out-of-range sample removed

## Sensor 2

The prediction results using data from two, three, and four sets for training and validation for sensor 2 are given in Tables 3.22-3.24. No obvious out-of-range samples were found and the prediction results are generally better than sensor 1.

**Table 3.22 Performance of PLS prediction models for VWC established using various combinations of the sample data from sensor 2**

Sample sets used	CI	R <sup>2</sup> (training/validation)	RMSE (cm <sup>3</sup> /cm <sup>3</sup> ) (training/validation)
1+2	0.85	0.96/0.90	1.58/2.51
1+3	0.88	0.99/0.86	0.84/3.22
1+4	0.54	1.00/0.85	0.00/2.94
2+3	0.63	0.84/0.68	1.70/1.87
2+4	0.71	1.00/0.90	0.22/1.69
3+4	0.71	0.98/0.87	0.92/2.13
1+2+3	0.88	0.91/0.82	2.34/3.19
1+2+4	0.85	0.89/0.88	2.28/2.64
1+3+4	0.88	0.96/0.87	1.46/2.62
2+3+4	0.72	0.92/0.82	1.82/2.41
1+2+3+4	0.88	0.86/0.84	2.78/2.93

For NO<sub>3</sub>, set 2 had the larger range in (12.7 - 92.4 ppm). However, the range for set 3 (1.1- 6.2 ppm) does not match that of set 1 well. We removed the sample in set 2 that had the lowest NO<sub>3</sub> concentration, when we used sample sets 1+2, and the performance of the model was improved.

**Table 3.23 Performance of PLS prediction models for NO<sub>3</sub> established using various combinations of the sample data from sensor 2**

Sample sets used	CI	R <sup>2</sup> (training/validation)	RMSE(ppm) (training/validation)
1+2	0.90	0.95/0.53	5.50/19.5
1+2*	0.73	0.64/0.50	15.0/13.8
1+3	0.90	1.00/0.99	0.00/3.41
1+4	0.90	1.00/0.98	0.91/5.11
2+3	0.36	0.27/0.26	23.6/11.4
2+3*	0.69	0.60/0.59	9.54/8.45
2+4	0.36	0.90/0.89	8.88/4.38
3+4	0.61	0.65/0.63	0.94/0.71
1+2+3	0.92	1.00/0.64	0.03/18.8
1+2+4	0.93	0.94/0.72	7.04/16.2
1+3+4	0.90	1.00/0.97	0.00/5.97
2+3+4	0.29	0.99/0.79	1.28/4.36
1+2+3+4	0.92	0.94/0.92	7.11/8.19

\*with an out-of-range sample removed

For NH<sub>4</sub>, we found one out-of-range sample that had NH<sub>4</sub> concentration value of 12.4 ppm, whereas the range of remaining samples was 3.5-9.2 ppm (when sample sets 1+2 were used). Removal of this sample improved the prediction performance.

**Table 3.24 Performance of PLS prediction models for NH<sub>4</sub> established using various combinations of the sample data from sensor 2**

Sample sets used	CI	R <sup>2</sup> (training/validation)	RMSE(ppm) (training/validation)
1+2	0.42	1.00/0.27	0.02/2.20
1+2*	0.65	0.72/0.56	0.96/0.98
1+3	0.86	0.94/0.88	0.46/0.65
1+4	0.79	0.87/0.86	0.51/0.49
2+3	0.45	0.64/0.48	1.46/1.06
2+4	0.32	0.80/0.79	1.06/0.48
3+4	0.49	0.40/0.30	1.20/0.86
3+4*	0.75	0.74/0.65	0.61/0.64
1+2+3	0.63	1.00/0.79	0.02/1.25
1+2+4	0.43	0.83/0.51	0.69/1.63
1+2+4*	0.67	0.88/0.71	0.58/0.74
1+3+4	0.85	0.71/0.60	0.60/1.07
2+3+4	0.45	0.48/0.29	1.57/1.15
2+3+4*	0.95	0.77/0.47	0.59/0.99
1+2+3+4	0.63	0.79/0.74	0.95/0.93

\*with an out-of-range sample removed

### Sensor 3

For sensor 3, only one out-of-range sample was identified. Removal of this sample improved the performance of the sets 2+4 prediction model, as can be seen in the Table 3.25.



**Table 3.25 Performance of PLS prediction models for VWC established using various combinations of the sample data from sensor 3**

Sample sets used	CI	R <sup>2</sup> (training/validation)	RMSE (cm <sup>3</sup> /cm <sup>3</sup> ) (training/validation)
1+2	0.41	0.99/0.92	0.54/1.09
1+3	0.52	0.85/0.73	2.82/2.91
1+4	0.46	0.99/0.97	0.46/0.63
2+3	0.88	0.63/0.63	2.43/2.65
2+3*	0.75	0.99/0.85	0.49/1.35
2+4	0.82	1.00/0.57	0.03/5.03
2+4*	0.51	0.83/0.62	1.08/2.49
3+4	0.89	0.90/0.87	2.06/2.03
1+2+3	0.52	0.89/0.58	1.20/3.17
1+2+3*	0.85	0.99/0.87	0.36/1.58
1+2+4	0.49	0.93/0.81	1.27/1.84
1+3+4	0.60	0.78/0.77	3.07/2.85
2+3+4	0.83	0.59/0.48	3.49/3.66
1+2+3+4	0.60	0.77/0.66	2.89/3.01

\*with an out-of-range sample removed

For NO<sub>3</sub>, no obvious out-of-range samples were found and the prediction results are generally better than sensor 1 and sensor 2.

**Table 3.26 Performance of PLS prediction models for NO<sub>3</sub> established using various combinations of the sample data from sensor 3**

Sample sets used	CI	R <sup>2</sup> (training/validation)	RMSE(ppm) (training/validation)
1+2	0.65	0.74/0.71	8.13/10.3
1+3	0.88	1.00/0.96	0.00/6.41
1+4	0.88	1.00/0.96	1.45/5.91
2+3	0.76	0.89/0.89	5.14/4.67
2+4	0.76	1.00/0.99	1.04/1.53
3+4	0.66	0.99/0.98	0.27/0.34
1+2+3	0.88	0.91/0.76	6.71/12.07
1+2+4	0.88	1.00/0.77	0.00/12.05
1+3+4	0.88	0.90/0.85	7.83/10.65
2+3+4	0.76	0.97/0.51	2.81/9.08
1+2+3+4	0.88	1.00/0.68	0.73/12.79

One out-of-range sample was found for sensor 3 in predicting NH<sub>4</sub>, with the concentration value of 12.1 ppm, whereas the range of remaining samples was 2.9-10.1 ppm. A similar removal action was taken to improve the models.

**Table 3.27 Performance of PLS prediction models for NH<sub>4</sub> established using various combinations of the sample data from sensor 3**

Sample sets used	CI	R <sup>2</sup> (training/validation)	RMSE(ppm) (training/validation)
1+2	0.74	1.00/0.53	0.00/2.07
1+3	0.76	0.70/0.63	1.13/1.25
1+4	0.80	0.99/0.45	0.08/1.25
2+3	0.39	0.60/0.32	1.60/1.06
2+3*	0.88	0.70/0.66	0.79/0.75
2+4	0.47	0.55/0.48	1.70/1.16
2+4*	0.74	0.99/0.95	0.17/0.35
3+4	0.48	0.99/0.31	0.07/0.84
1+2+3	0.74	0.99/0.49	0.14/1.99
1+2+3*	0.76	0.77/0.70	1.11/1.01
1+2+4	0.74	0.39/0.37	1.68/2.11
1+2+4*	0.76	1.00/0.94	0.03/0.47
1+3+4	0.77	0.84/0.76	0.75/0.93
2+3+4	0.42	0.98/0.64	0.36/0.80
1+2+3+4	0.75	1.00/0.64	0.00/1.24
1+2+3+4*	0.91	0.76/0.76	0.97/1.01

\*with an out-of-range sample removed

#### Sensor 4

Good results were obtained for sensor 4 data in predicting VWC and NO<sub>3</sub>.

**Table 3.28 Performance of PLS prediction models for VWC established using various combinations of the sample data from sensor 4**

Sample sets used	CI	R <sup>2</sup> (training/validation)	RMSE (cm <sup>3</sup> /cm <sup>3</sup> ) (training/validation)
1+2	0.88	0.90/0.89	2.50/2.53
1+3	0.73	1.00/0.83	0.00/3.06
1+4	0.68	0.96/0.95	1.09/1.41
2+3	0.56	0.93/0.77	1.07/1.47
2+4	0.63	0.78/0.78	3.08/2.32
3+4	0.47	0.99/0.91	0.51/1.15
1+2+3	0.85	0.96/0.94	1.36/1.54
1+2+4	0.88	1.00/0.88	0.27/2.40
1+3+4	0.73	0.93/0.74	1.65/3.34
2+3+4	0.60	0.99/0.88	0.58/1.69
1+2+3+4	0.85	0.93/0.87	1.73/2.31

**Table 3.29 Performance of PLS prediction models for NO<sub>3</sub> established using various combinations of the sample data from sensor 4**

Sample sets used	CI	R <sup>2</sup> (training/validation)	RMSE(ppm) (training/validation)
1+2	0.81	0.83/0.80	11.8/15.3
1+3	0.83	1.00/0.95	2.30/10.7
1+4	0.83	0.98/0.94	4.73/11.4
2+3	0.49	1.00/0.85	0.87/7.47
2+4	0.48	0.95/0.75	7.89/10.1
3+4	0.43	0.99/0.96	0.41/0.47
1+2+3	0.88	1.00/0.95	0.00/8.89
1+2+4	0.88	1.00/0.95	1.09/8.94
1+3+4	0.83	0.90/0.90	11.08/13.5
2+3+4	0.45	1.00/0.83	1.39/6.87
1+2+3+4	0.88	1.00/0.93	0.06/10.2

For NH<sub>4</sub>, removal of an out-of-range sample helped improve most models. This sample had a NH<sub>4</sub> concentration value of 15.4 ppm. However, for the model that used all four sets of data, an additional out-of-range sample was removed, which had a value of 11.1 ppm.

**Table 3.30 Performance of PLS prediction models for NH<sub>4</sub> established using various combinations of the sample data from sensor 4**

Sample sets used	CI	R <sup>2</sup> (training/validation)	RMSE(ppm) (training/validation)
1+2	0.39	0.99/0.32	0.10/2.77
1+2*	0.52	0.79/0.64	0.83/0.87
1+3	0.99	0.91/0.90	0.58/0.78
1+4	0.41	0.97/0.56	0.25/1.29
2+3	0.51	0.65/0.56	2.03/1.45
2+3*	0.82	0.99/0.83	0.18/0.90
2+4	0.45	1.00/0.88	0.06/0.66
3+4	0.48	1.00/0.97	0.13/0.21
1+2+3	0.62	0.98/0.30	0.34/2.95
1+2+3*	0.79	0.99/0.53	0.02/1.31
1+2+4	0.56	0.96/0.50	0.36/2.35
1+3+4	0.99	0.97/0.73	0.33/1.14
2+3+4	0.49	0.54/0.40	2.04/2.43
2+3+4*	0.79	0.99/0.70	0.16/1.02
1+2+3+4	0.62	0.53/0.48	1.41/1.58
1+2+3+4*	0.99	0.44/0.42	1.84/1.66
1+2+3+4**	0.82	0.70/0.70	1.14/0.98

\*with one out-of-range sample removed

### 3.2.5.2 Models across multiple sensors

It was hoped that “universal” prediction models for multiple sensors could be established to measure soil properties. Existence of such models would greatly simplify the calibration procedures for the sensors.

For this study, all sets of data from two, three, and four sensors were used to establish the “universal” models.

### ***“Universal” models for measuring VWC***

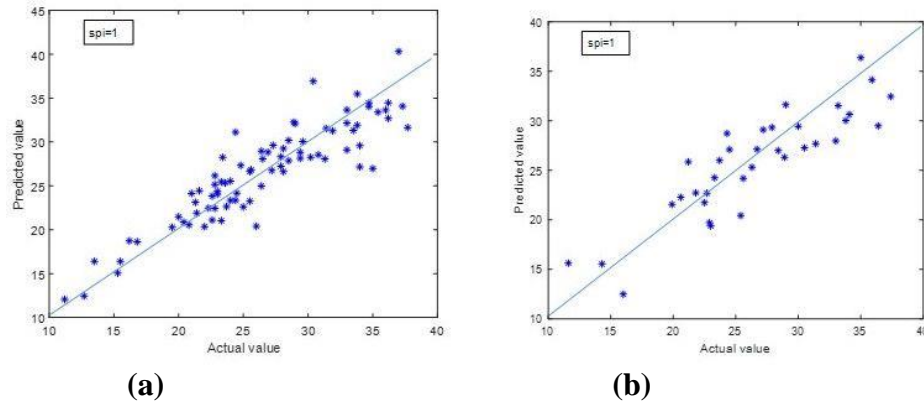
The performance of models established using all four sets of samples from two, three, and four sensors are shown in Table 3.31. It is interesting to observe that, with larger numbers of samples taken from multiple data sets for multiple sensors, out-of-range samples became a smaller problem. For example, when data from sensors 1 and 3 were used, we had lower  $R^2$  values and larger RMSE. However, because of the large number of samples (39 for training and 17 for validation), removing out-of-range samples did not seem to improve the performance of the model. This was probably because that the larger number of samples allowed more equal ranges of the training and validation data sets and more uniform distributions of the VWC values within the total range. Thus, the model became less sensitive to the out-of-range samples.

**Table 3.31 Performance of PLS prediction models for VWC established using various combinations of sensor data from all sets**

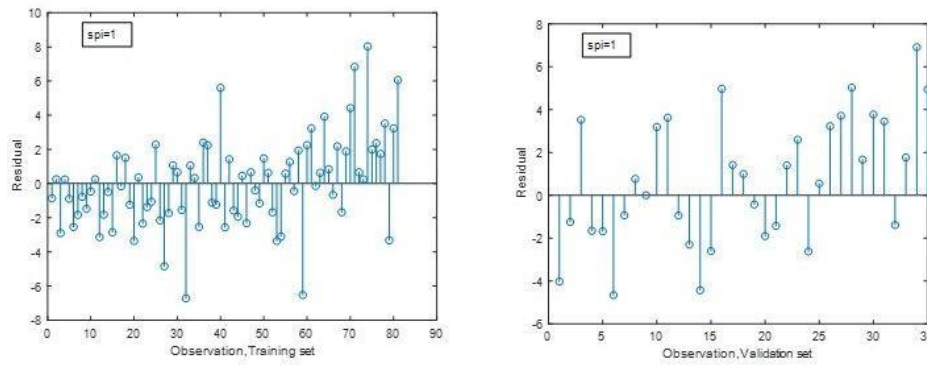
Sample sensors used	CI	$R^2$ (training/validation)	RMSE (cm <sup>3</sup> /cm <sup>3</sup> ) (training/validation)
1+2	0.94	0.91/0.81	1.75/2.75
1+3	0.90	0.55/0.53	3.14/3.70
1+3*	0.89	0.97/0.47	0.76/3.96
1+4	0.93	0.86/0.84	2.03/2.41
2+3	0.97	0.71/0.57	3.47/4.60
2+4	0.92	0.87/0.87	2.51/2.62
3+4	0.87	0.78/0.77	2.88/3.16
1+2+3	0.96	0.86/0.68	2.19/3.39
1+2+4	0.94	0.98/0.85	0.87/2.42
1+3+4	0.90	0.71/0.70	3.03/3.06
2+3+4	0.97	0.86/0.65	2.49/3.99
1+2+3+4	0.98	0.81/0.76	2.64/3.02

\*with an out-of-range sample removed

Figure 3.11 shows predicted vs actual VWC for the training and validation data sets by using all four data sets of all four sensors. The CI was as high as 0.98, the  $R^2$  values for the training and validation data sets were 0.81 and 0.76, and the RMSE were 2.64 and 3.02, respectively. Figure 3.12 shows the residuals of the prediction model.



**Figure 3.11 Performance of VWC prediction model using four sets of samples across four sensors (a) training data, (b) validation data**



**Figure 3.12 Residual values for the training and validation data for VWC by using four sets of samples across four sensors**

### ***“Universal” models for measuring $NO_3$***

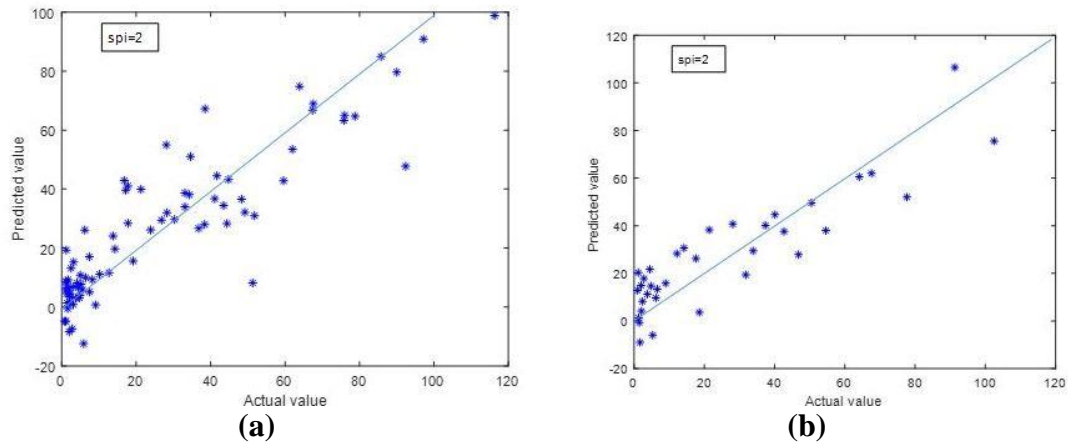
Performance of the prediction models established to measure  $NO_3$  using all sample sets across multiple sensors are shown in Table 3.32. It can be seen that the CI values were high for almost all the models, and most of the  $R^2$  values are satisfactory.

**Table 3.32 Performance of PLS prediction models for NO<sub>3</sub> established using various combinations of sensor data from all sets**

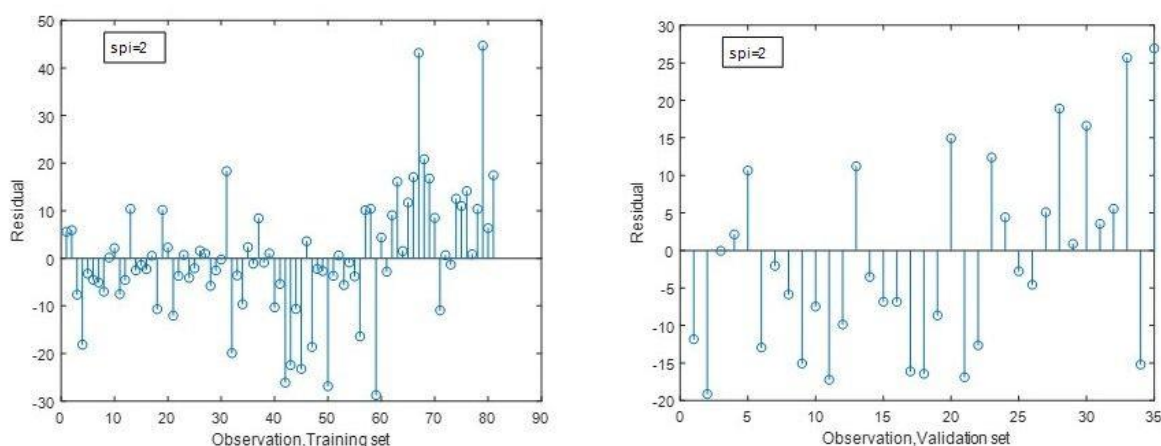
Sample sensors used	CI	R <sup>2</sup> (training/validation)	RMSE(ppm) (training/validation)
1+2	0.93	0.81/0.66	10.1/15.3
1+3	0.88	0.87/0.78	7.28/10.58
1+4	0.88	0.99/0.85	2.19/13.1
2+3	0.92	0.68/0.67	14.2/16.0
2+4	0.88	0.94/0.86	8.16/13.6
3+4	0.88	0.87/0.85	10.7/13.0
1+2+3	0.84	0.68/0.54	13.7/15.2
1+2+4	0.83	0.77/0.72	14.4/15.2
1+3+4	0.83	0.73/0.73	15.0/13.8
2+3+4	0.83	0.82/0.81	13.0/13.0
1+2+3+4	0.88	0.80/0.80	12.7/12.6

Figure 3.13 shows the predicted vs actual NO<sub>3</sub> for the training and validation data sets.

Figure 3.14 gives their residuals plots.



**Figure 3.13 Performance of NO<sub>3</sub> prediction model using four sets of samples across four sensors (a) training data, (b) validation data**



**Figure 3.14 Residual plots for training and validation data for NO<sub>3</sub> by using four sets of samples across four sensors**

### ***“Universal” models for measuring NH<sub>4</sub>”***

The performance of the models established to predict NH<sub>4</sub> using all sample sets across multiple sensor was generally low (Table 3.33 and Figures 3.15 and 3.16).

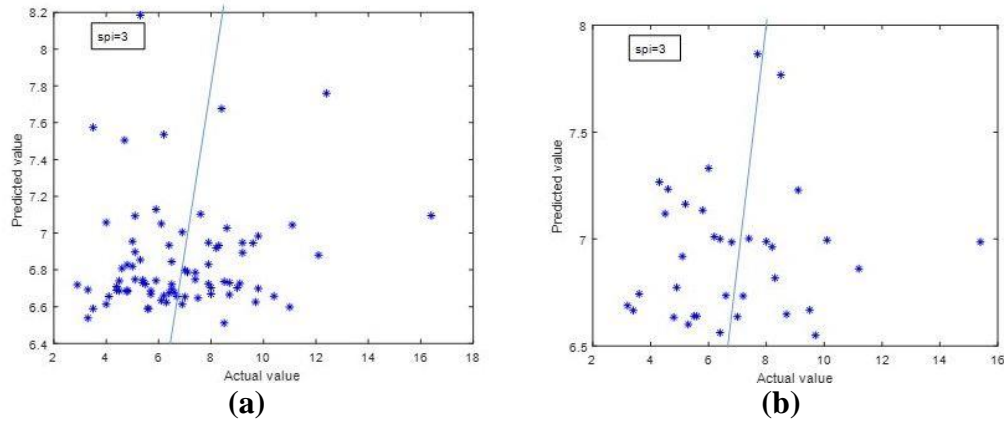
**Table 3.33 Performance of PLS prediction models for NH<sub>4</sub> established using various combinations of sensor data from all sets**

Sample sensors used	CI	R <sup>2</sup> (training/validation)	RMSE(ppm) (training/validation)
1+2	0.69	0.68/0.28	1.17/2.47
1+3	0.65	0.77/0.44	1.06/2.25
1+4	0.92	0.25/-0.1	2.07/3.12
1+4*	0.98	0.14/0.05	1.88/2.02
2+3	0.94	0.71/0.55	1.10/1.53
2+4	0.74	0.58/0.23	01.32/2.42
2+4*	0.83	1.00/0.26	0.01/1.67
3+4	0.71	0.80/0.12	0.97/2.65
1+2+3	0.66	0.82/0.27	1.00/1.85
1+2+4	0.69	0.10/0.05	239/2.24
1+2+4*	0.86	0.43/0.13	1.46/2.03
1+3+4	0.66	0.83/0.27	1.00/0.85
2+3+4	0.71	0.63/0.25	1.42/1.85
1+2+3+4	0.90	0.02/0.01	2.34/2.44
1+2+3+4*	0.84	0.35/0.33	1.71/1.64

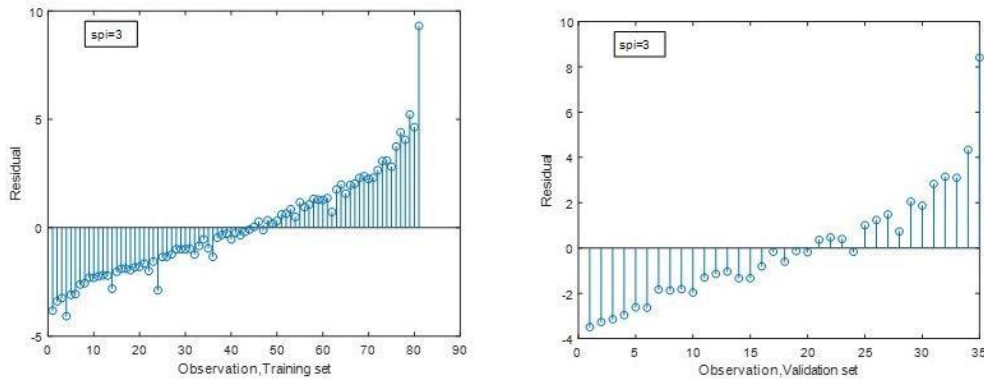
\*with an out-of-range sample removed



Figure 3.15 shows the predicted versus actual  $\text{NH}_4$  for the training and validation data sets. Figure 3.16 shows their residuals for the model using all four data sets across all four sensors.



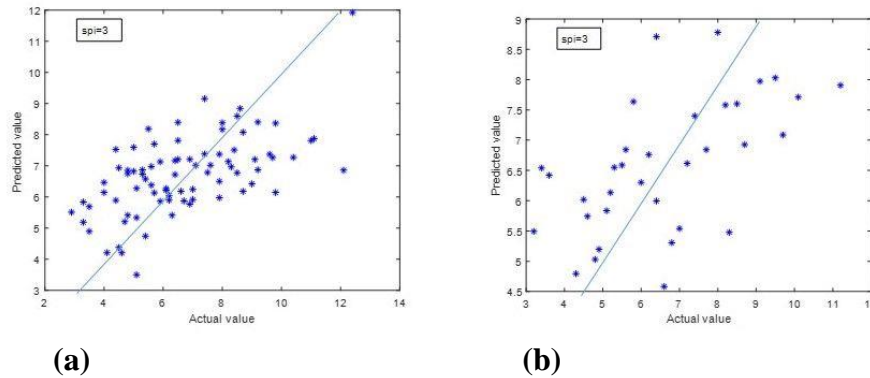
**Figure 3.15 Performance of  $\text{NH}_4$  prediction model using four sets of samples from four sensors (a) training data, (b) validation data**



**Figure 3.16 Performance of  $\text{NH}_4$  prediction model using four sets of samples from four sensors (a) training data, (b) validation data**

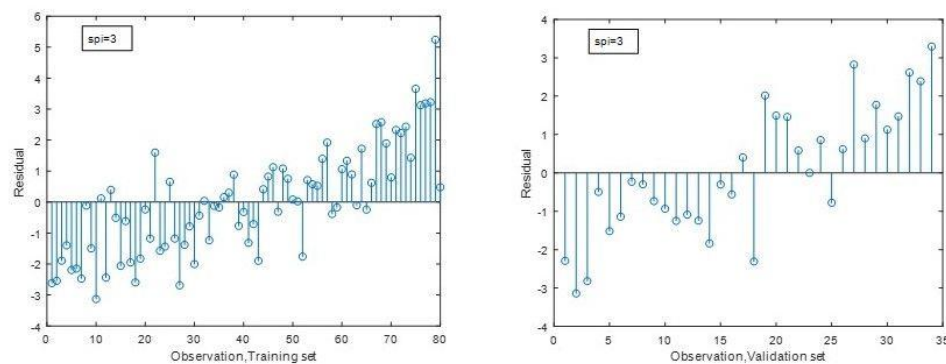
After studying the sample data, several out-of-range samples were identified. Figures 3.17 and 3.16 show the results of the predicted versus actual  $\text{NH}_4$  for the training and validation data sets and the residuals of the prediction model after these samples were removed. Still, the results are not satisfactory. This may indicate that, for measuring  $\text{NH}_4$ , the sensors will need to be individually calibrated. Establishment of “universal” models that can work across multiple

sensors may be difficult. On the other hand, the range of  $\text{NH}_4$  in all samples was very low (2.9-16.4 ppm, Appendix H). The samples may not have covered a sufficient range as a basis to establish effective prediction models.



**Figure 3.17 Performance of  $\text{NH}_4$  prediction model using four sets of samples for all sets in all sensors with samples removed (a) training data, (b) validation data**

Figure 3.18 shows residual values for training and validation data for  $\text{NH}_4$  by using all data for all sensors after removing some out-of-range samples.



**Figure 3.18 Residual plot for training and validation data for  $\text{NH}_4$  by using all data for all sensors after removing some out-of-range samples**

### 3.2.6 Selection of signature frequencies

The current design of the FR sensor uses 635 frequencies, ranging from near DC (50 Hz) to 120 MHz. The use of a large number of frequencies for measurement has many disadvantages: 1) it increases the cost of the sensor, 2) it slows down the measurement (the current design requires about 30 seconds to complete a measurement), 3) more importantly, there is no need for these many frequencies, especially when considering the collinearities that exist among the signals generated in these frequencies, 4) even more importantly, excessive numbers of frequencies may actually cause overfitting of the prediction models, hence reducing the measurement accuracy.

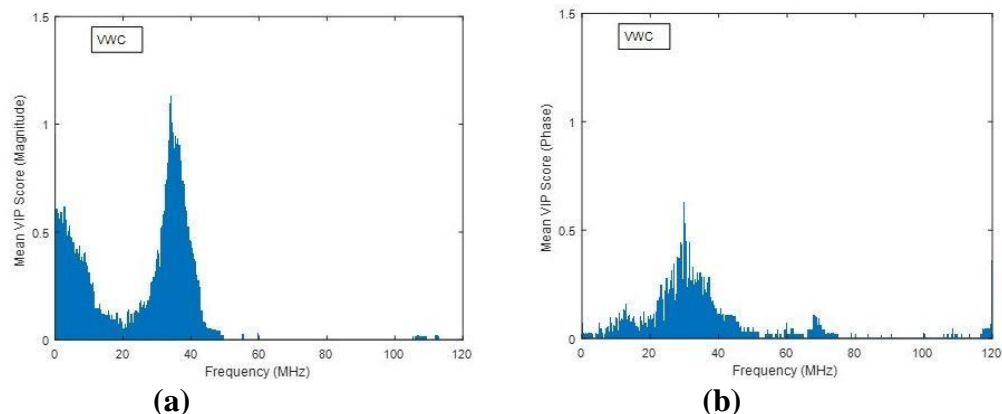
Based on these considerations, we concentrated our effort on identifying the frequencies that are specifically related to the response variable – the variable to be measured. As we have discussed in previous sections, we considered the frequencies with the largest VIP scores the signature frequencies for a specific response variable, and with the Matlab program we performed the PLS regression analysis, which allowed us to identify these signature frequencies.

Figures 3.19-3.21 show the frequencies with their VIP scores selected by the PLS regression procedure using the predictor and response loadings and scores in predicting VWC,  $\text{NO}_3$ , and  $\text{NH}_4$ , respectively. It can be seen that, for all three response variables, certain frequency components were identified from the gain spectra and phase spectra. The VIP scores shown in these figures are the average values of the VIP scores derived in 74 PLS regression runs – some across data sets within each sensor and some across sensors, which included all possible within- and across-sensor combinations. It is interesting to find that, for measuring VWC, the peak VIP score of 1.13 occurred at 34 MHz and another peak occurred at low frequencies near DC, whereas for  $\text{NO}_3$ , the peak VIP score of 1.44 was found to be at the near DC frequency (50 Hz). These findings agreed with our guesses. For  $\text{NO}_3$ , electric conduction is probably the activity that is most

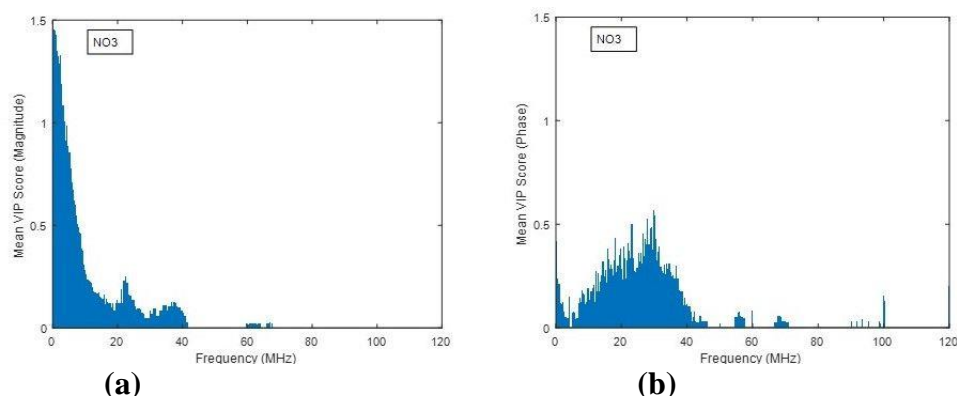
closely associated with the conductive ions. Water, however, varies the dielectric constant of the medium, and it also changes the mobility of the conductive anions and cations that are soluble in water.

The mean VIP scores for  $\text{NH}_4$  in the gain spectra were small. This probably explains why in both the laboratory and outdoor experiments, it was more difficult to measure  $\text{NH}_4$ . The VIP scores shown in the three phase spectra all peaked near 30MHz, which does not seem to be very helpful if our goal is to distinguish between the responses variables. However, the patterns of the VIP scores for VWC,  $\text{NO}_3$ , and  $\text{NH}_4$  in both the magnitude and phase spectra are different, which give hope for better predictions of the three response variables with careful selection of the signature frequencies. For now, we may use a threshold for the mean VIP scores, and define all frequency components that have an average VIP score in either the magnitude or phase spectra above the threshold as the signature frequency. This would greatly reduce the number of frequencies used for renditions, hence reducing the cost and measurement time, while yielding more accurate measurements. These assumptions will need to be proved through further studies.

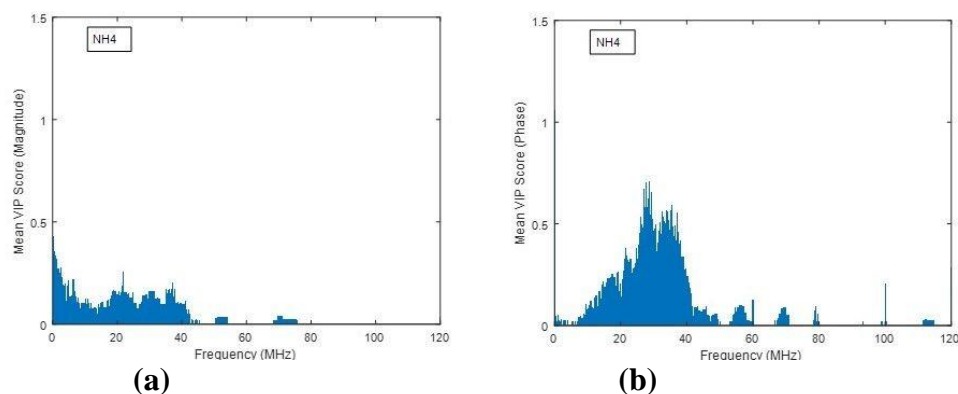
It is worth noticing that the number of predictor variables (frequencies) selected through the 74 runs of the PLS regression program are, in general, low. The average number of these frequencies selected for models predicting VWC,  $\text{NO}_3$ , and  $\text{NH}_4$  were 67, 69, and 48, respectively. At the same time, the optimum numbers of the PLS components ranged from 7 to 16.



**Figure 3.19 Average VIP scores derived from 74 PLS regression runs for measuring VWC: (a) Gain spectra, (b) Phase spectra (Two distinguished peaks (DC and 30 MHz) in the magnitude spectra)**



**Figure 3.20 Average VIP scores derived from 74 PLS regression runs for measuring  $\text{NO}_3^-$ : (a) Gain spectra, (b) Phase spectra (One distinguished peak (DC) in the magnitude spectra)**



**Figure 3.21 Average VIP scores derived from 74 PLS regression runs for measuring  $\text{NH}_4^+$ : (a) Gain spectra, (b) Phase spectra (One distinguished peak (27 MHz) in the phase spectra)**

## **Chapter 4 - Conclusions and Recommendations**

### **4.1 Conclusions**

In this study, we modified the probe of a frequency-response (FR)-based dielectric sensor that was previously developed in the BAE Instrumentation and Control Laboratory at Kansas State University, and we tested its ability to simultaneously measure multiple soil properties under both laboratory and outdoor conditions.

1. In the laboratory experiment, we tested the sensor in its ability to simultaneously measure soil volumetric water content (VWC), bulk density, and salinity. The R-squared values for these tests were, in general, above 0.85.
2. In the laboratory experiment, we tested the sensor in its ability to simultaneously measure soil volumetric water content (VWC), bulk density, and one of two nitrogen fertilizers: ammonium nitrate and ammonium sulfate. The R-squared values for these tests were, in general, above 0.8.
3. In the laboratory experiment, we tested the sensor in its ability to simultaneously measure soil volumetric water content (VWC), bulk density, and an organic nitrogen fertilizer – urea. The R-squared values for these tests were in general above 0.7.
4. In the laboratory experiment, we tested the sensor in its ability to simultaneously measure soil volumetric water content (VWC), bulk density, and two of three nitrogen fertilizers: ammonium nitrate, ammonium sulfate, and urea. The tests that included ammonium nitrate and ammonium sulfate gave R-squared values of higher than 0.6 for ammonium nitrate but were generally lower than 0.5 for ammonium sulfate. For the test that included urea and ammonium sulfate, the R-squared values for measuring VWC and urea were, in general, above 0.8 and 0.7, respectively, but that for measuring ammonium sulfate were low. On the other hand, for the test that included urea and ammonium nitrate, only VWC could be measured with R-squared values of higher than 0.8, while the fertilizers, either individually or combined, could not be measured correctly.

5. For the outdoor experiment, we planted corn plants in two specially made pots, and placed two FR sensors in each pot. The two sensors were buried at different depths, both in the root zone of the plants.
6. For the outdoor experiment, each FR sensor took 35 dielectric measurements within a period of 70 days, of which 29 proved to be valid. We also took soil samples at the same time to measure the VWC using a commercial sensor and sent the samples to the Soil Testing Laboratory at Kansas State University to determine the concentration of nitrate-N and ammonium-N in the soil using standard methods and used them as reference values.
7. We conducted a PLS (partial least squares) regression analysis to establish prediction models for VWC, nitrate-N, and ammonium-N using different combinations of data as the training and validation data sets. We also developed an improved approach to select the number of PLS components and the numbers of frequencies with high VIP (Variable Importance in Projection) scores (“signature frequencies”).
8. For the dielectric spectra obtained for each sensor across different data sets, we, in general, got R-squared values higher than 0.8 in predicting VWC, nitrate-N, and ammonium-N.
9. For the dielectric spectra obtained across different sensors, we, in general, got R-squared values higher than 0.7 in predicting VWC and nitrate-N. However, the R-squared values for predicting ammonium-N across sensors were low.
10. By studying the VIP scores provided by the PLS regression analyses, we identified frequency components in both the magnitude and phase dielectric spectra for three response variables – VWC, nitrate-N, and ammonium-N, and we identified special patterns in the VIP scores for VWC and nitrate-N. We believe that these findings lay the

foundation for more accurately identifying the signature frequencies for individual response variables, which may lead to a lower cost sensor that gives faster and more accurate measurements.

11. From the outdoor experiment, we learned lessons on how to design and install the sensor probe and its accessories in soil for long-term deployment.

## **4.2 Recommendations for further studies**

Recommendations include the following:

1. Improve the outdoor usability of the sensor by adding memory storage and a solar power supply.
2. Improve the design of the probe and accessories for tough agricultural applications.
3. Design a simplified calibration procedure to expand the applicability of the sensor.
4. Test the sensor in other fertilizer types, such as P and K.
5. Identify signature frequencies for other variables and test these frequencies during measurements.
6. Increase the frequency range for potential improvement in  $\text{NH}_4$  measurement.



## References

- Abrol, I.P., J. Yadav, and F. Massoud. 1988. Salt-affected soils and their management. FAO soils bulletin, Soil resources management and conservation service, FAO land and water development division: Rome, Italy, 39; 131. <http://www.fao.org/3/x5871e/x5871e00.htm>
- Adamchuk V.I., J.W. J. Hummel, M.T. Morgan, S.K. Upadhyaya. 2014. On-the-go soil sensors for precision agriculture. *Computers and Electronics in Agriculture*. 44: 71–91.
- Afsar. M., N. J. R. Birch, and R. N. Clarke. 1986. The measurement of the properties of materials, *Proc. IEEE* 74:183–199.
- Alabama Extension, soil, forage, and water testing laboratory. <https://ssl.acesag.auburn.edu/anr/soillab/soilquality.php>, accessed on 4/28/2019
- Al-Taie F. 1970. Salt-affected and water-logged soils of Iraq. Report to seminar on methods of amelioration of saline and water-logged soils, Baghdad, Iraq. In: Qureshi. A. S, W. Ahmed, Al-falhi. A. A. (2013) Optimum groundwater table depth and irrigation schedules for controlling soil salinity in central Iraq. *Irrig. and Drain*. 62: 414–424
- ASTM D 2216, Standard test methods for laboratory determination of moisture content of soil, ASTM International, West Conshohocken, PA, 2008.
- Boldaji N, M. Sharifi, A. Jamshidi, B. Younesi-Alamouti, and S. Minaee. 2011. A dielectric-based combined horizontal sensor for on-the-go measurement of soil water content and mechanical resistance. *Sens. Actuators a* 2011, 171, 131–137.
- Chighladze G. A. Kaleita, S. Birrell, and S. Logsdon. 2011. Estimating soil solution nitrate concentration from dielectric spectra using partial least squares analysis. *Soil Sci. Soc. Am. J.* 76:1536–1547, doi:10.2136/sssaj2011.0391.
- Committee of Agriculture and Water Resources Sector. 2009. Policy paper of agriculture and water resources sector. 2010–2014 five years national plan technical committee, ministry of planning and development cooperation, republic of Iraq; 25-7.
- Fine, A. K., H.M van Es, and R.R. Schindelbeck. 2017. Statistics, scoring functions, and regional analysis of a comprehensive soil health database. *Soil Sci. Soc. Am. J.* 81(3): 589-601.
- Food and agricultural organization of the UN (FAO). 2011. Crops. FAOSTAT: <http://faostat.fao.org/site/567/default.aspx#ancor>. 12.02.2012.
- Gaudin R., and J. Dupuy. 1999. Ammonical nutrition of transplanted rice fertilized with large urea granules. *Agron. J.*, 91(1), 33-36.  
doi:<http://dx.doi.org/10.2134/agronj1999.00021962009100010006x>.

- Gholizadeh, A., M. Amin, M. Anuar, A. R. Aimrun, and M. M. Saberioon. 2011. Temporal variability of SPAD chlorophyll meter readings and its relationship to total nitrogen in leaves within a Malaysian paddy field. *Aust. J. Basic Applied Sci.*, 5(5), 236-245.
- Haney R.L. 2014. Soil health. USDA-ARS.  
[http://www.nrcs.usda.gov/Internet/FSE\\_DOCUMENTS/nrcs144p2\\_043902.pdf](http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs144p2_043902.pdf) accessed 29 April 2019).
- Haney R. L., E.B. Haney, D.R. Smith, R.D. Harmel, and M.J. White. 2018. The soil health tool. Theory and initial broad-scale application. *Applied Soil Ecology*, 125: 162-168.
- Hindrikx, J.M.H., B.S. Das, D.L. Corwin, J.M. Wraith, and R.G. Kachanoski. 2002. Relationship between soil water solute concentration and apparent soil electrical conductivity. In *Methods of Soil Analysis*; Dane, J.H., Topp, G.C., Eds.; Soil Science Society of America: Madison, WI, USA, pp. 1275–1282.
- Hillhorst, M.A. 1998. Dielectric characterization of soil. Ph.D. dissertation, Wageningen Agricultural University, Wageningen, the Netherlands.
- Kelleners, T. J., D. A. Robinson, P. J. Shouse, J. E. Ayars, and T. H. Skaggs. 2005. Frequency dependence of the complex permittivity and its impact on dielectric sensor calibration in soils. *Soil Sci. Am. J.* 69(1): 67:76.
- Kim, H.J, K.A. Sudduth, and J.W. Hummel. 2009. Soil micronutrient sensing for precision agriculture. *J. Environmental Monitoring* 11, 1810-1824.
- Kupfer K., E. Trinks, N. Wagner, and C. Hubner, 2007. TDR measurements and simulations in high lossy bentonite materials, *Meas. Sci. Technol.* 18:1118–1136.
- Lee, K.H. 2005. A dielectric permittivity sensor for simultaneous measurement of multiple soil properties. A Ph.D. Dissertation. Kansas State University, USA.
- Lee, K.H., N. Zhang, W.B. Kuhn, and G.J. Kluitenberg. 2007<sub>1</sub>. A frequency-response permittivity sensor for simultaneous measurement of multiple soil properties: part I. The frequency-response method. *Transactions of the ASABE*, 50(6): 2315-2326
- Lee, K.H. and N. Zhang. 2007<sub>2</sub>. A frequency-response permittivity sensor for simultaneous measurement of multiple soil properties: Part II. Calibration-model tests. *Transactions of the ASABE*, 50(6): 2327-2336.
- Lunt, I.A., S.S Hubbard, and Y. Rubin. 2005. Soil moisture content estimation using ground-penetrating radar reflection data. *J. Hydrol.* 307:254–269.
- Marco, B. 2011. Measuring soil water content: A review. *HortTechnology* 21(3):293-300.
- McGowen, E. B., S. Sharma, S. Deng, H. Zhang, and J.G. Warren. 2018. An automated laboratory method for measuring CO<sub>2</sub> emissions from soils. *Agric. Environmental Letters* 3(1)

- Mengel, K., and M. Viro. 1978. The significance of plant energy status for the uptake and incorporation of  $\text{NH}_4\text{-N}$  by young rice plants. *Soil Sci. Plant Nutr.* 24(3):407-416. doi:<http://dx.doi.org/10.1080/00380768.1978.10433119>
- Minet, J., S. Lambot, G. Delaide, J.A. Huisman, H.A. Vereecken, and M.A. Vanclooster. 2010. Generalized frequency domain reflectometry modeling technique for soil electrical properties determination, *Vadose Zone J.* 9: 1063–1073.
- Muñoz-Carpena, R., A. Ritter, and D. Bosch. 2005a. Field methods for monitoring soil water status. In: Alvarez-Benedi, J., Muñoz-Carpena, R. (Eds.), *soil-water-solute process characterization*. CRC Press LLC, Boca Raton, pp. 167–195.
- Muñoz-Carpena, R., S. Shukla, and K. Morgan. 2004. Field devices for monitoring soil water content. *Extension bulletin*, vol. 343. Dept. of Agricultural and Biological Engineering, University of Florida.
- NRCS, <https://www.nrcs.usda.gov/wps/portal/nrcs/main/soils/health/>, accessed on 4/28/2019.
- Numerical Algorithms Group. 2007. Partial least squares. *NAGNews*, 64. Retrieved August 12, 2009 from <http://www.nag.com/IndustryArticles/partialleastsquares.pdf>.
- Ochieng, G.M., O.I. Ojo, and F.A. Otieno. 2013. Use of remote sensing and geographical information system (GIS) for salinity assessment of Vaal-Harts irrigation scheme, South Africa. *Environ. Syst. Res.* 2, 4 doi:10.1186/2193-2697-2-4. <http://www.environmentalsystemsresearch.com/content/2/1/4>
- Pandey, G., R. Kumar, R. Weber, 2013. Real time detection of soil moisture and nitrates using on-board in-situ impedance spectroscopy. *The 2013 IEEE international conference on systems, man, and cybernetics, manchester, 2013*, pp. 1081-1086. doi: 10.1109/SMC.2013.188.
- Pluta, S.E. and J.W. Hewitt, 2009. Non-destructive impedance spectroscopy measurement for soil characteristics. *The 2009 GeoHunan International Conference, Challenges and Recent Advances in Pavement Technologies and Transportation Geotechnics*. August 3-6, 2009, Changsha, Hunan, China. Pp. 144-149. [https://doi.org/10.1061/41041\(348\)21](https://doi.org/10.1061/41041(348)21)
- Robert, G, J. Evans, C. Kenneth, S. Bradley, and A. King. 2013 Adoption of site-specific variable rate sprinkler irrigation systems. *Irrig Sci* (2013) 31:871–887 DOI 10.1007/s00271-012-0365-x
- Robinson, D.A., C.S. Campbell, J.W. Hopmans, B.K. Hornbuckle, S.B. Jones, and R. Knight. 2008. Soil moisture measurement for ecological and hydrological moistureshed-scale observatories: a review, *Vadose Zone J.* 7 (1) (2008) 358–389.
- Robinson, D.A, S.B. Jones, J.M. Wraith, S.P. Friedman, and D. Or. 2003. A review of advances in dielectric and electrical conductivity measurement in soils using time domain reflectometry. *Vadose Zone Journal* 2:444-475.

- Saberioon M, M.S. Amin, A. Gholizadeh, and M.H. Ezri. 2014. A review of optical methods for assessing Nitrogen contents during rice growth. *Applied Engineering in Agriculture* Vol. 30(4): 657-669 © 2014 American Society of Agricultural and Biological Engineers ISSN 0883-8542 DOI 10.13031/aea.30.10478.
- Scholte, J.W., J.Q. Shang, and R.K Rowe. 2002. Improved complex permittivity measurement and data processing technique for soil-water systems. *Geotechnical Journal*, 25(2), 187-198 *Sci. Soc. Am. J.* 68:1515-1525.
- Sharpley, A.N., T.C. Daniel, J.T. Sims, and D.H. Pote. 1996. Determining environmentally sound soil phosphorus levels. *J. Soil Water Conserv.* 51:160–166.
- Shultz, S., 2009. Calibration of Permittivity Sensors to Measure Contaminants in Water and In Biodiesel Fuel. M.S. Thesis, 2009. Kansas State University, USA.
- Sinfield, J.V., D. Fagermana, and O. Colicb. 2010. Evaluation of sensing technologies for on-the-go detection of macro-nutrients in cultivated soils. *Computers and Electronics in Agriculture*: 70: 1–18.
- Skinner, A.J., and M.F. Lambert. 2011. An automatic soil pore-water salinity sensor based on a wetting-front detector” *IEEE Sensors Journal* 11 (1): 245.
- Soil Health Institute. 2019. National soil health measurements to accelerate agricultural transformation. <https://soilhealthinstitute.org/national-soil-health-measurements-accelerate-agricultural-transformation/>, accessed on 4/28/2019
- Soorya, E., M. Tejashree, and P. Suganya. 2013. Smart drip irrigation system using sensor networks. *International Journal of Scientific & Engineering Research*, Volume 4, Issue 5, 2039 ISSN 2229-5518 IJSER © 2013 <http://www.ijser.org>
- Sreenivas, K., L. Venkatratnam, and P.V. Narasimha Rao. 1995. Dielectric properties of salt-affected soils, *Int. J. Remote Sens.*, 16/4 641.
- Stewart, B.A., and D.R. Nielsen. 1990. Irrigation of agricultural crops. American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America.
- Stewart, R. D., J. Jian, A.J. Gyawali, W.E. Thomason, B.D., Badgley, M.S. Reiter, and M.S. Strickland, 2018. What we talk about when we talk about soil health. *Agricultural & Environmental Letters*, 3: 180033. doi:10.2134/acl2018.06.0033.
- Suweis, S., A. Rinaldo, D.Z. Van, E. Daly, A. Maritan, and A. Porporato. 2010. Stochastic modeling of soil salinity. *Geophys. Res. Lett.* 37, L07404.
- Takenaga, H. 1995. Science of rice plant: physiology. In the role of mineral nutrition on root growth of crop plants N.K. Fageria, A. Moreira (2011.Vol. 110, pp.251-331). *Advances in agronomy.* , 2011.

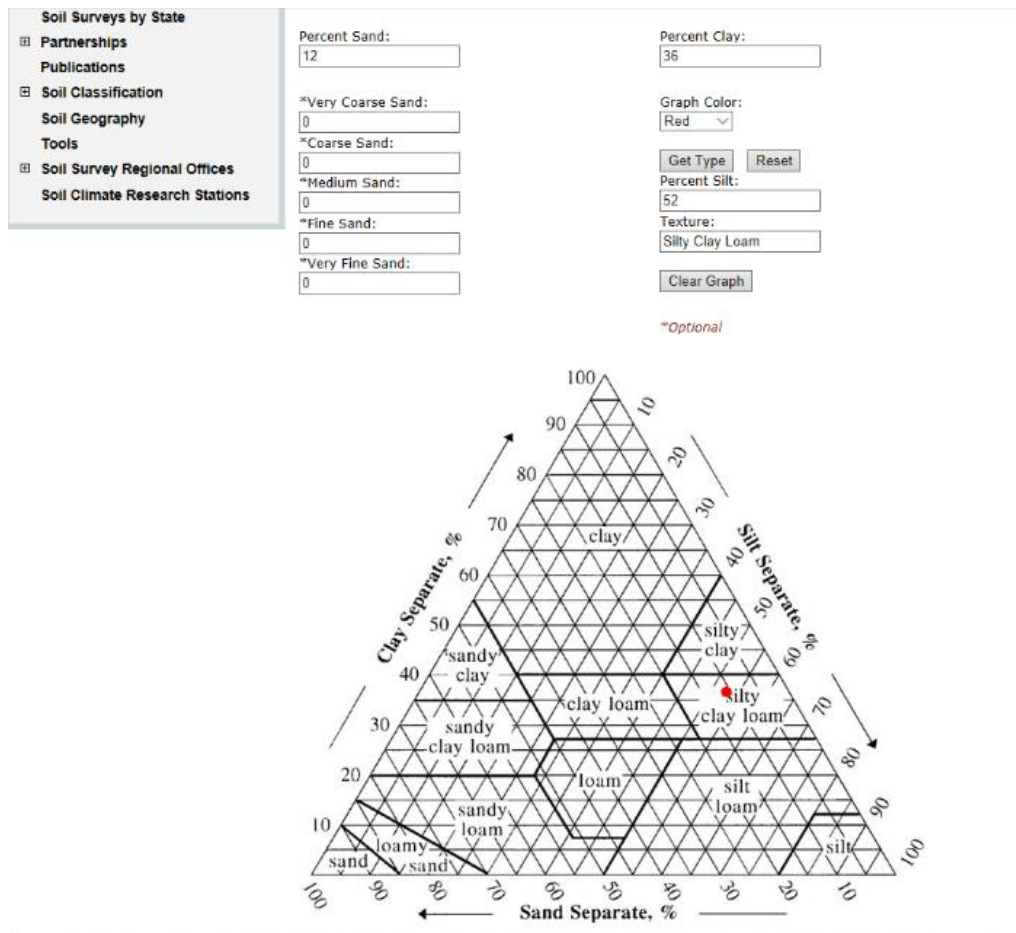
- Tang, N. 2009. A real-time control system for a frequency response-based permittivity sensor. M.S. Thesis. Kansas State University.
- Thomasson, J. A, R. Sui, M.S. Cox, and A. Alrajehy. 2001. Soil reflectance sensing for determining soil properties in precision agriculture. Transactions of the ASAE, v. 44, n. 6, p. 1445-1453.
- Topp, G.C., J. L. Davis, and A.P. Annan. 1980. Electromagnetic determination of soil water content: measurement in coaxial transmission lines. Water resource. Res. 16, 574–582.
- Topp, G.C., S. Zegelin, and I. White. 2000. Impacts of the real and imaginary components of relative permittivity on time domain reflectometry measurements in soils. Soil Sci. Soc. Am. J. 64:1244-1252.
- Utey, B.C., T. M. Wynn, N. Zhang, L. E. Teany, 2012. Evaluation of a permittivity sensor for continuous monitoring of suspended sediment concentration. Transactions of the ASABE. 54(4): 1299-1309
- Viscarra Rossel, R.A., and A.B McBratney. 1998a. Soil chemical analytical accuracy and costs: implications from precision agriculture. Australian Journal of Experimental Agriculture 38, 765 – 775.
- Wang, Y., T. Wig, J. Tang, and L. M. Hallberg. 2003. Dielectric properties of food relevant to RF and microwave pasteurization and sterilization, J. Food Eng. 57:257–268.
- Ware, B., 2012. Frequency response based permittivity sensors for measuring air contaminants, M.S. Thesis. Kansas State University, USA.
- Wise, B.M., J.M. Shaver, N.B. Gallagher, W. Windig, R. Bro, and R.S Koch. 2006. PLS\_Toolbox 4.0 Manual. Wenatchee: Eigenvector Research, Inc.
- Wikipedia, 2020. [https://en.wikipedia.org/wiki/Partial\\_least\\_squares\\_regression](https://en.wikipedia.org/wiki/Partial_least_squares_regression)
- Xi, X. and N. Zhang. 2011. Measuring ethanol/gasoline mixing ratio based on the dielectric properties. Sensor Letter. 9(3): 2011.
- Yadav, V., K. Ani, S. Sharan, A. K. Sinha, M.. K. Gupta, and R. A. Jangid. 2009. Measurement of dielectric behavior of fertilized soil at microwave frequency. Journal of Agricultural Science, 1/2 December (2009) 42.
- Yun, S., H. Qingrong, G. Huadong, L. Yuan, D. Qing, and H. Chunming. 2003. Effect of dielectric properties of moist salinized soils on backscattering coefficients extracted from RADARSAT image”, IEEE Trans. Geosci, and Remote Sens., 41/8 August (2003) 1879.
- Zazueta, F.S., and J. Xin. 1994. Soil moisture sensors. Bulletin, vol. 292. Florida Cooperative Extension Science, Institute of Food and Agricultural Sciences, University of Florida.

Zhang. N., G. Fan, K.H. Lee, G.J. Kluitenberg, and T.M. Loughin. 2004. Simultaneous measurement of soil water content and salinity using a frequency-response method. Soil Sci. Soc. Am. J. 68:1515-1525.

## Appendix A - Soil Texture

Consult the soil pyramid below to find the scientific classification of the soil used in the experiment. We got sand 12%, silt 52%, and clay 36%.

[https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/?cid=nrcs142p2\\_054167](https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/?cid=nrcs142p2_054167)



## Appendix B - Matlab Program for preliminary observation

```
clear all;
filename1 = uigetfile
[A] = textread(filename1, '', 'delimiter', ',');
filename2 = uigetfile
[B] = textread(filename2, '', 'delimiter', ',');
filename3 = uigetfile
[C] = textread(filename3, '', 'delimiter', ',');
filename4 = uigetfile
[D] = textread(filename4, '', 'delimiter', ',');
filename5 = uigetfile
[E] = textread(filename5, '', 'delimiter', ',');

Gain1 = A(2:636,3);
Phase1 = A(2:636,4);
Gain2 = B(2:636,3);
Phase2 = B(2:636,4);
Gain3 = C(2:636,3);
Phase3 = C(2:636,4);
Gain4 = D(2:636,3);
Phase4 = D(2:636,4);
Gain5 = E(2:636,3);
Phase5 = E(2:636,4);

for i = 1:20
    freq(i) = 50*(A(i+1,1)-50)+50;
end

for i = 21:40
    freq(i) = 10^(0.15*(A(i+1,1)-69)+3);
end

for i = 41:635
    freq(i) = 200000*(A(i+1,1)-89) + 1000000;
end

freq = freq';

figure
plot(freq, Gain1, 'b-', freq, Gain2, 'r-', freq, Gain3, 'g-', freq, Gain4, 'k-',
    freq, Gain5, 'm-');
xlabel('Frequency (Hz)');
ylabel('Gain (DV)');
title([filename1]);
legend(filename1, filename2, filename3, filename4, filename5);
grid on;

figure
plot(freq, Phase1, 'b-', freq, Phase2, 'r-', freq, Phase3, 'g-', freq, Phase4, 'k-',
    freq, Phase5, 'm-');
xlabel('Frequency (Hz)');
ylabel('Phase (DV)');
title([filename1]);
```



```
legend(filename1, filename2, filename3, filename4, filename5);  
grid on;
```

## **Appendix C - Outdoor test planter**

We used was a 14 in. x 27.5 in. slate rubber self-watering planter with a one-year guarantee for all-season durability - no cracking even in freezing temperatures. According to the product information, it was made from recycled materials with sleek 100% recycled rubber, an elegant look of slate. It was heavy-duty and durable for all seasons and was weighted to discourage tipping over. A third party laboratory test showed that it will not harm plants in extreme heat. It was stain resistant and easy to clean and had 1 drainage hole (with removable plug) that protected against over watering by allowing excess water to drain. The planters were bought from Home Depot in Manhattan, Kansas.



We made a window in the pot for taking samples and measurements using both our sensors and commercially available soil moisture sensors (Model No. 5TE, Decagon Devices, Pullman, WA). We put the 5TE sensors at the same depth that we put our sensors. On the opposite side was an outlet for the sensor wires. We designed the window and it had made by BAE technicians in the BAE workshop. The window had a firm lock to prevent soil and water from leaking. The planter could be placed on almost any ground surface, including grass, mulch, floors, and more.

<https://www.homedepot.com/p/Tierra-Verde-14-in-x-27-5-in-Slate-Rubber-Self-Watering-Planter-MT5100067CM/204672115>

## Appendix D - Datalogger code

```
'CR800 Series Datalogge
'10HS sensor
'wiring:
'white - SW-12
'red - SE CH1
'bare - gnd
'5TE sensor
'Wiring:
'White wire (power) -> SW12V
'Red wire (data) -> C1
'Bare wire (ground) -> G
'Declare Public Variables
Public batt_volt,tenHSmV,HS_VWC
Public SensorOut(3)
Public VWC
Alias SensorOut(1) = Eb
Alias SensorOut(2) = EC
Alias SensorOut(3) = Temp
Units Temp = Deg_C
Units EC = dS/m
Units VWC = m^3/m^3
'Define Data Tables
DataTable (Soildata,1,-1)
    DataInterval (0,30,Sec,10)
    Minimum (1,batt_volt,FP2,0,False)
    Sample (1,Eb,FP2)
    Sample (1,Temp,FP2)
    Sample (1,EC,FP2)
    Sample (1,VWC,FP2)
```

```

        Sample (1,HS_VWC,FP2)
EndTable

'Main Program
SequentialMode
BeginProg
    Scan (5,Sec,0,0)
        Battery (batt_volt)

        'Apply power to white wire of sensor through SW-12
        PortSet (9,1)
        'Delay for at least 250 mSec for sensor to enter SDI-12 mode.
        Delay (0,1,Sec)
        'Query sensor for 3 SDI-12 outputs. Default address for all Decagon Digital
sensors is 0.
        SDI12Recorder (SensorOut(),1,0,"M!",1.0,0)
        VoltSe (tenHSmV,1,mV2500,1,1,0,_60Hz,1.0,0)
        'Turn SW12V off
        PortSet (9,0)

        'Apply calibration to bulk dielectric reading (Eb)

$$\text{VWC} = 4.3\text{E-}6 * \text{Eb}^3 - 5.5\text{E-}4 * \text{Eb}^2 + 2.92\text{E-}2 * \text{Eb} - 5.3\text{E-}2$$

        'Apply calibration to 10SH senor

$$\text{HS\_VWC} = 2.97\text{e-}9 * \text{tenHSmV}^3 - 7.37\text{e-}6 * \text{tenHSmV}^2 + 6.69\text{e-}3 * \text{tenHSmV} - 1.92$$


        'Call Output Tables
        CallTable Soildata
    NextScan
EndProg

```

## Appendix E - Matlab program for the laboratory tests

```
% ncomp - Number of PLS components used.

% Predicting soil properties using PLS regression
% MathWorks Documentation on plsregress.
% https://www.mathworks.com/help/stats/plsregress.html
% Modified to run the soil nutrient data
%     Naiqian Zhang
%     April 19, 2019

clear all;
clc;

Reading_files_AMD2;
te_cons = cons;
b_cons = cons;

%predicted soil property index (spi)
% 1=VWC, 2=Ammonium sulfate, 4=bulk density
spi=2;

% mag_digital=mag_digital(1:fr,1:150*wci);
% phase_digital=phase_digital(1:fr,1:150*wci);

b_cons=te_cons;
cons=te_cons(:,spi);

%fst_col=50*(si-1)+1;

n_mag_digital=[];
n_phase_digital=[];
n_cons=[];
t_cons=[];

    n_mag_digital=mag_digital;
    n_phase_digital=phase_digital;
    n_cons=cons;
    t_cons=b_cons;

mag_dB=((((5000/1023)*n_mag_digital)-900)/30);
phase_deg=((n_phase_digital*(5000/1023))-1800)/10;

mag=mag_dB';
phs=phase_deg';
mag_phs=[mag phs];
```

```

phs_mag=[phs mag];

%Bulk EC at 1kHz
BEC=0.2457*exp(-0.1283.*mag(:,5));
BE_cons=[t_cons BEC];

%Select measured variabel
%%%%%%%%%%%%
% data_x=phs;
% data_x=mag;
data_x=phs_mag;
% data_x=mag_phs;

data_y=n_cons;
N_freq = 1270;
% N_freq = 635;
%%%%%%%%%%%%

[train_data,train_con,valid_data,valid_con]=divide_train_test(data_x,data_y);
sub_train_data=train_data;
sub_train_con=train_con;

sub_valid_data=valid_data;
sub_valid_con=valid_con;

% NZ 041819

% ncomp - Number of PLS components used.

MaxValR2 = 0;
for ncomp=1:8
    Call_PLScregress
    ValR2(ncomp) = rsquaredVAL
    if ValR2(ncomp) > MaxValR2
        MaxValR2 = ValR2(ncomp);
        MaxR2ncomp = ncomp;
    end
end

ncomp = MaxR2ncomp;

% Loading the data
[dummy,h] = sort(sub_train_con);
oldorder = get(gcf,'DefaultAxesColorOrder');
set(gcf,'DefaultAxesColorOrder',jet(60));
figure(1);
plot3(repmat(1:N_freq,[size(h),1]),repmat(sub_train_con(h),1,N_freq)',sub_train_data(h,:))';
set(gcf,'DefaultAxesColorOrder',oldorder);
xlabel('Wavelength Index'); ylabel('sub_train_con'); axis('tight');
grid on

```

```

% Use the plsregress function to fit a PLSR model with 20 PLS components and
one response.
X = sub_train_data;
y = sub_train_con;
[n,p] = size(X);
[Xloadings,Yloadings,Xscores,Yscores,betaPLS10,PLSPctVar] = plsregress(...
    X,y,20);

figure(2);
plot(1:20,cumsum(100*PLSPctVar(2,:)), '-bo');
xlabel('Number of PLS components');
ylabel('Percent Variance Explained in Y');

% Compute the fitted response values for the 7-component model.
[Xloadings,Yloadings,Xscores,Yscores,betaPLS] = plsregress(X,y,ncomp);
yfitPLS = [ones(n,1) X]*betaPLS;

figure(3);
plot(y,yfitPLS, 'bo');
xlabel('Observed Response');
ylabel('Fitted Response (Calibration set)');
legend({'PLSR'});

% R-squared value for the calibration data set
TSS = sum((y-mean(y)).^2);
RSS_PLS = sum((y-yfitPLS).^2);
rsquaredPLS = 1 - RSS_PLS/TSS

% Percent variance in X explained
figure(4);
plot(1:20,100*cumsum(PLSPctVar(1,:)), 'b-o');
xlabel('Number of Principal Components');
ylabel('Percent Variance Explained in X');
legend({'PLSR'}, 'location', 'SE');

% Choosing the Number of Components with 10-fold Cross-Validation by
% estimating MSE
[Xl,Yl,Xs,Ys,beta,pctVar,PLSmsep] = plsregress(X,y,10, 'CV',10);

% PCRmse = sum(crossval(@pcrsse,X,y,'KFold',10),1) / n;
figure(5);
plot(0:10,PLSmsep(2,:), 'b-o');
xlabel('Number of components');
ylabel('Estimated Mean Squared Prediction Error');
legend({'PLSR'}, 'location', 'NE');

% PLS weights
figure(6);
[Xl,Yl,Xs,Ys,beta,pctVar,mse,stats] = plsregress(X,y,ncomp);
plot(1:N_freq,stats.W, '-');
xlabel('Variable');
ylabel('PLS Weight');
legend({'1st Component' '2nd Component' '3rd Component'}, ...
    'location', 'SW');

```

```

% R-squared value for validation data set
[Xloadings,Yloadings,Xscores,Yscores,betaPLS] = plsregress(X,y,ncomp);
yfitPLS_val = [ones(n,1) sub_valid_data]*betaPLS;

TSS = sum((y-mean(y)).^2);
RSS_PLS_val = sum((sub_valid_con-yfitPLS_val).^2);
rsquaredVAL = 1 - RSS_PLS_val/TSS

figure(7);
plot(y,yfitPLS_val,'bo');
xlabel('Observed Response');
ylabel('Fitted Response (Validation set)');
legend({'PLSR'});

```



## Appendix F - Matlab program for data preparation and preprocessing for the outdoor tests

```
% Program that allows selections of data files, magnitude vs. phase
% and treatment of data from three runs
%
%           Naiqian Zhang, October 10, 2019
%
%Added data compaction and VIP sorting.
%   10/12/19, Naiqian Zhang
%
%
% Change files order dependant to thier order in Excel file
% Change Excel file name when use "Case=0 at switch replic"
%   Mohammed Hasan, October 27,2019
% 5 sets   November 4,2019
clc
close all
clear all
n_file = 116;

% User selected parameters
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% This program runs on only one spi, It divides the data sets in "cal_val"
% into training and validation sets based on the user selected ratio
% ("percent_cal". The model selection is based on minimum RMSE for the
% validation ste.
%
%           Naiqian Zhang, 070620
cal_val = [1:39];
spi = 1;
percent_cal = 0.7;

% N_test = 1, add testing data to exam the best PLS model
% N_test = 0, do not add testing

N_test = 0;
if N_test == 1
    testn = [33:39 61:67];
end

file_property = 'Soil properties M_NO3_NH4_5SETS_before 100719.xlsx';
% % % use 3 replication for each file
% file_property = 'Soil properties M_NO3_NH4_5SETS_3replications.xlsx';
property = xlsread(file_property);

prop1 = property(cal_val,spi);
[BB,II] = sort(prop1);
JJ = cal_val(II);
Jsize = size(JJ);
Jsize1 = Jsize(2);

MM=0;
```

```

NN=0;
for KK=1:Jsize1
    if KK == round(1/percent_cal*(MM+1))
        MM = MM+1;
        calib(MM) = JJ(1, KK);
    else
        NN = NN+1;
        valid(NN) = JJ(1, KK);
    end
end

if N_test == 1
    prop_test = property(testn, spi);
end

% calib = [ ];
% valid = [ ];

% calib = calib';
% valid = valid';

% divide_data
% 1 divide calibration and validation data into 70% and 30%
% 0 not divide
divide_data = 1;

% mag_phs: 1 for mag, 2 for phase, 3 for mag+phase
mag_phs =3;

% replic: 1 for separating three runs, 0 for averaging three runs
replic = 0;

% compress: 0 for not compress, 1 for compress
compress =0;

% n_compress: number of frequencies the original data is compressed to.
n_compress = 100;
if compress == 0
    n_compress = 635;
end
% n_select: the largest number of VIP frequencies selected.
n_VIPfreq = 200;
%%
% !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
% Note that n_VIPfreq+ncomp CANNOT exceed n_compress!
% !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

% n_average: number of neighboring frequencies used in averaging
% set n_average=1 if do not want to smooth the data.
n_average = 1;

%%
%%%%%%%%%%%%%% added by Mingqiang Han 05/24/2020%%%%%%%%%%%%%%
% algorithm_select
% 1 partial least square select variables, selected variables fitting

```

```

% PLS regression and support vector regression
% 0 elastic regression
algorithm_select = 1;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

s_train = size(calib);
n_train = s_train(2);
s_valid = size(valid);
n_valid = s_valid(2);
if N_test == 1
    s_testn = size(testn);
    n_testn = s_testn(2);
end

file_names(1,:) = 'no1-2019-07-27-1130.TXT';
file_names(2,:) = 'no1-2019-07-29-1930.TXT';
file_names(3,:) = 'no1-2019-07-30-1000.TXT';
file_names(4,:) = 'no1-2019-08-01-0900.TXT';
file_names(5,:) = 'no1-2019-08-02-1000.TXT';
file_names(6,:) = 'no1-2019-08-05-1000.TXT';
file_names(7,:) = 'no1-2019-08-09-1000.TXT';
file_names(8,:) = 'no1-2019-08-12-0900.TXT';

file_names(9,:) = 'no2-2019-07-27-1130.TXT';
file_names(10,:) = 'no2-2019-07-29-1930.TXT';
file_names(11,:) = 'no2-2019-07-30-1000.TXT';
file_names(12,:) = 'no2-2019-08-01-0900.TXT';
file_names(13,:) = 'no2-2019-08-02-1000.TXT';
file_names(14,:) = 'no2-2019-08-05-1000.TXT';
file_names(15,:) = 'no2-2019-08-09-1000.TXT';
file_names(16,:) = 'no2-2019-08-12-0900.TXT';

file_names(17,:) = 'no3-2019-07-27-1130.TXT';
file_names(18,:) = 'no3-2019-07-29-1930.TXT';
file_names(19,:) = 'no3-2019-07-30-1000.TXT';
file_names(20,:) = 'no3-2019-08-01-0900.TXT';
file_names(21,:) = 'no3-2019-08-02-1000.TXT';
file_names(22,:) = 'no3-2019-08-05-1000.TXT';
file_names(23,:) = 'no3-2019-08-09-1000.TXT';
file_names(24,:) = 'no3-2019-08-12-0900.TXT';

file_names(25,:) = 'no4-2019-07-27-1130.TXT';
file_names(26,:) = 'no4-2019-07-29-1930.TXT';
file_names(27,:) = 'no4-2019-07-30-1000.TXT';
file_names(28,:) = 'no4-2019-08-01-0900.TXT';
file_names(29,:) = 'no4-2019-08-02-1000.TXT';
file_names(30,:) = 'no4-2019-08-05-1000.TXT';
file_names(31,:) = 'no4-2019-08-09-1000.TXT';
file_names(32,:) = 'no4-2019-08-12-0900.TXT';

file_names(33,:) = 'no1-2019-08-14-1000.TXT';
file_names(34,:) = 'no1-2019-08-16-1700.TXT';
file_names(35,:) = 'no1-2019-08-20-0930.TXT';

```

```

file_names(36,:) = 'no1-2019-08-23-1000.TXT';
file_names(37,:) = 'no1-2019-08-25-1900.TXT';
file_names(38,:) = 'no1-2019-08-27-1030.TXT';
file_names(39,:) = 'no1-2019-08-29-0930.TXT';

file_names(40,:) = 'no2-2019-08-14-1000.TXT';
file_names(41,:) = 'no2-2019-08-16-1700.TXT';
file_names(42,:) = 'no2-2019-08-20-0930.TXT';
file_names(43,:) = 'no2-2019-08-23-1000.TXT';
file_names(44,:) = 'no2-2019-08-25-1900.TXT';
file_names(45,:) = 'no2-2019-08-27-1030.TXT';
file_names(46,:) = 'no2-2019-08-29-0930.TXT';

file_names(47,:) = 'no3-2019-08-14-1000.TXT';
file_names(48,:) = 'no3-2019-08-16-1700.TXT';
file_names(49,:) = 'no3-2019-08-20-0930.TXT';
file_names(50,:) = 'no3-2019-08-23-1000.TXT';
file_names(51,:) = 'no3-2019-08-25-1900.TXT';
file_names(52,:) = 'no3-2019-08-27-1030.TXT';
file_names(53,:) = 'no3-2019-08-29-0930.TXT';

file_names(54,:) = 'no4-2019-08-14-1000.TXT';
file_names(55,:) = 'no4-2019-08-16-1700.TXT';
file_names(56,:) = 'no4-2019-08-20-0930.TXT';
file_names(57,:) = 'no4-2019-08-23-1000.TXT';
file_names(58,:) = 'no4-2019-08-25-1900.TXT';
file_names(59,:) = 'no4-2019-08-27-1030.TXT';
file_names(60,:) = 'no4-2019-08-29-0930.TXT';

file_names(61,:) = 'no1-2019-09-02-1800.TXT';
file_names(62,:) = 'no1-2019-09-04-1600.TXT';
file_names(63,:) = 'no1-2019-09-06-0930.TXT';
file_names(64,:) = 'no1-2019-09-09-1000.TXT';
file_names(65,:) = 'no1-2019-09-11-1000.TXT';
file_names(66,:) = 'no1-2019-09-13-1200.TXT';
file_names(67,:) = 'no1-2019-09-14-1900.TXT';

file_names(68,:) = 'no2-2019-09-02-1800.TXT';
file_names(69,:) = 'no2-2019-09-04-1600.TXT';
file_names(70,:) = 'no2-2019-09-06-0930.TXT';
file_names(71,:) = 'no2-2019-09-09-1000.TXT';
file_names(72,:) = 'no2-2019-09-11-1000.TXT';
file_names(73,:) = 'no2-2019-09-13-1200.TXT';
file_names(74,:) = 'no2-2019-09-14-1900.TXT';

file_names(75,:) = 'no3-2019-09-02-1800.TXT';
file_names(76,:) = 'no3-2019-09-04-1600.TXT';
file_names(77,:) = 'no3-2019-09-06-0930.TXT';
file_names(78,:) = 'no3-2019-09-09-1000.TXT';
file_names(79,:) = 'no3-2019-09-11-1000.TXT';
file_names(80,:) = 'no3-2019-09-13-1200.TXT';
file_names(81,:) = 'no3-2019-09-14-1900.TXT';

file_names(82,:) = 'no4-2019-09-02-1800.TXT';
file_names(83,:) = 'no4-2019-09-04-1600.TXT';

```

```

file_names(84,:) = 'no4-2019-09-06-0930.TXT';
file_names(85,:) = 'no4-2019-09-09-1000.TXT';
file_names(86,:) = 'no4-2019-09-11-1000.TXT';
file_names(87,:) = 'no4-2019-09-13-1200.TXT';
file_names(88,:) = 'no4-2019-09-14-1900.TXT';

file_names(89,:) = 'no1-2019-09-17-1000.TXT';
file_names(90,:) = 'no1-2019-09-19-1000.TXT';
file_names(91,:) = 'no1-2019-09-22-1100.TXT';
file_names(92,:) = 'no1-2019-09-25-1000.TXT';
file_names(93,:) = 'no1-2019-09-27-1000.TXT';
file_names(94,:) = 'no1-2019-09-30-1000.TXT';
file_names(95,:) = 'no1-2019-10-02-1030.TXT';

file_names(96,:) = 'no2-2019-09-17-1000.TXT';
file_names(97,:) = 'no2-2019-09-19-1000.TXT';
file_names(98,:) = 'no2-2019-09-22-1100.TXT';
file_names(99,:) = 'no2-2019-09-25-1000.TXT';
file_names(100,:) = 'no2-2019-09-27-1000.TXT';
file_names(101,:) = 'no2-2019-09-30-1000.TXT';
file_names(102,:) = 'no2-2019-10-02-1030.TXT';

file_names(103,:) = 'no3-2019-09-17-1000.TXT';
file_names(104,:) = 'no3-2019-09-19-1000.TXT';
file_names(105,:) = 'no3-2019-09-22-1100.TXT';
file_names(106,:) = 'no3-2019-09-25-1000.TXT';
file_names(107,:) = 'no3-2019-09-27-1000.TXT';
file_names(108,:) = 'no3-2019-09-30-1000.TXT';
file_names(109,:) = 'no3-2019-10-02-1030.TXT';

file_names(110,:) = 'no4-2019-09-17-1000.TXT';
file_names(111,:) = 'no4-2019-09-19-1000.TXT';
file_names(112,:) = 'no4-2019-09-22-1100.TXT';
file_names(113,:) = 'no4-2019-09-25-1000.TXT';
file_names(114,:) = 'no4-2019-09-27-1000.TXT';
file_names(115,:) = 'no4-2019-09-30-1000.TXT';
file_names(116,:) = 'no4-2019-10-02-1030.TXT';

% Starting 10/7/2019, there were wiring problems and most data files became
% invalid as indicated by "%".
% no1-2019-10-07-1030, no3-2019-10-07-1030, and no4-2019-10-07-103
% are the only valid data after 10/02/19.
% We decide not to any data file on or after 100719. - 021920
%%file_names(117,:) = 'no1-2019-10-07-1030.TXT';
% file_names(118,:) = 'no1-2019-10-12-1530.TXT';
% file_names(119,:) = 'no1-2019-10-17-1100.TXT';
% file_names(120,:) = 'no1-2019-10-20-1200.TXT';
% file_names(121,:) = 'no1-2019-10-23-1500.TXT';
% file_names(122,:) = 'no1-2019-10-28-1500.TXT';
%
% file_names(123,:) = 'no2-2019-10-07-1030.TXT';
% file_names(124,:) = 'no2-2019-10-12-1530.TXT';
% file_names(125,:) = 'no2-2019-10-17-1100.TXT';
% file_names(126,:) = 'no2-2019-10-20-1200.TXT';
% file_names(127,:) = 'no2-2019-10-23-1500.TXT';
% file_names(128,:) = 'no2-2019-10-28-1500.TXT';

```

```

%%
%% file_names(129,:) = 'no3-2019-10-07-1030.TXT';
% file_names(130,:) = 'no3-2019-10-12-1530.TXT';
% file_names(131,:) = 'no3-2019-10-17-1100.TXT';
% file_names(132,:) = 'no3-2019-10-20-1200.TXT';
% file_names(133,:) = 'no3-2019-10-23-1500.TXT';
% file_names(134,:) = 'no3-2019-10-28-1500.TXT';
%%
%%file_names(135,:) = 'no4-2019-10-07-1030.TXT';
% file_names(136,:) = 'no4-2019-10-12-1530.TXT';
% file_names(137,:) = 'no4-2019-10-17-1100.TXT';
% file_names(138,:) = 'no4-2019-10-20-1200.TXT';
% file_names(139,:) = 'no4-2019-10-23-1500.TXT';
% file_names(140,:) = 'no4-2019-10-28-1500.TXT';
%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% edit by Mingqiang Han 05/24/2020%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% divide calibration and validation data into 70% and 30%
if divide_data == 1
    data_Index = [calib valid];
    total_num = length(calib)+length(valid);
    n_train = round(total_num*0.7);
    randIndex = randperm(total_num);
    calib = data_Index(randIndex(1:n_train));
    valid = data_Index(randIndex(n_train+1:end));
end
n_valid = length(valid);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

for k = 1:n_file
    text_file=textread(file_names(k,:),'', 'delimiter','');
    len = size(text_file);
    % change the variable name "length" to "len" since "length" is a built-in
    % function name Mingqiang Han 05/24/2020
    % length = size(text_file);
    j = 0;
    for i = 1:len(1)
        if text_file(i,1) ~= 1111 & text_file(i,1) ~= 5555 & text_file(i,1)
~= 9999
            j = j+1;
            data(j,:) = text_file(i,:);
        end
    end

    switch replic
    case 1
        for l = 1:3
            for m = 1:635
                mag_digital(m, (k-1)*3+1) = data((l-1)*635+m, 3);
                phs_digital(m, (k-1)*3+1) = data((l-1)*635+m, 4);
            end
            prop((k-1)*3+1,:) = property(k,:);
        end
    end
end

```

```

        N_train = 3*n_train;
        N_valid = 3*n_valid;
        if N_test ==1
            N_testn = 3*n_testn;
        end
    case 0
        for m = 1:635
            mag_digital(m,k) =
                (data(m,3)+data(m+635,3)+data(m+1270,3))/3.;
            phs_digital(m,k) =
                (data(m,4)+data(m+635,4)+data(m+1270,4))/3.;
        end
        prop = property;
        N_train = n_train;
        N_valid = n_valid;
        if N_test ==1
            N_testn = n_testn;
        end
    end
end

freq = data(1:635, 2)*10000;
for i = 1:27
    freq(i) = 50 + 368.52*(i-1);
end

mag_train_x = zeros(635,N_train);
mag_train_y = zeros(N_train);
mag_valid_x = zeros(635,N_valid);
mag_valid_y = zeros(N_valid);
if N_test == 1
    mag_testn_x = zeros(635,N_testn);
    mag_testn_y = zeros(N_testn);
end

phs_train_x = zeros(635,N_train);
phs_train_y = zeros(N_train);
phs_valid_x = zeros(635,N_valid);
phs_valid_y = zeros(N_valid);
if N_test == 1
    phs_testn_x = zeros(635,N_testn);
    phs_testn_y = zeros(N_testn);
end

switch replic
case 1
    for i = 1:n_train
        for j = 1:3
            mag_train_x(:,(i-1)*3+j) = mag_digital(:,(calib(i)-1)*3+j);
            phs_train_x(:,(i-1)*3+j) = phs_digital(:,(calib(i)-1)*3+j);
            prop_train((i-1)*3+j,:) = prop((calib(i)-1)*3+j,:);
        end
    end
    for i = 1:n_valid
        for j = 1:3
            mag_valid_x(:,(i-1)*3+j) = mag_digital(:,(valid(i)-1)*3+j);

```

```

        phs_valid_x(:, (i-1)*3+j) = phs_digital(:, (valid(i)-1)*3+j);
        prop_valid((i-1)*3+j,:) = prop((valid(i)-1)*3+j,:);
    end
end
if N-testn == 1
    for i = 1:n_testn
        for j = 1:3
            mag_testn_x(:, (i-1)*3+j) = mag_digital(:, (testn(i)-
1)*3+j);
            phs_testn_x(:, (i-1)*3+j) = phs_digital(:, (testn(i)-
1)*3+j);
            prop_testn((i-1)*3+j,:) = prop((testn(i)-1)*3+j,:);
        end
    end
end
case 0
    for i = 1:n_train
        mag_train_x(:,i) = mag_digital(:,calib(i));
        phs_train_x(:,i) = phs_digital(:,calib(i));
        prop_train(i,:) = prop(calib(i),:);
    end
    for i = 1:n_valid
        mag_valid_x(:,i) = mag_digital(:,valid(i));
        phs_valid_x(:,i) = phs_digital(:,valid(i));
        prop_valid(i,:) = prop(valid(i),:);
    end
    if N_test == 1
        for i = 1:n_testn
            mag_testn_x(:,i) = mag_digital(:,testn(i));
            phs_testn_x(:,i) = phs_digital(:,testn(i));
            prop_testn(i,:) = prop(testn(i),:);
        end
    end
end
end

switch mag_phs
case 1
    train_x = mag_train_x';
    valid_x = mag_valid_x';
    if N_test == 1
        testn_x = mag_testn_x';
    end
    fr = freq;
case 2
    train_x = phs_train_x';
    valid_x = phs_valid_x';
    if N_test == 1
        testn_x = phs_testn_x';
    end
    fr = freq;
case 3
    train_x = [mag_train_x; phs_train_x]';
    valid_x = [mag_valid_x; phs_valid_x]';
    if N_test == 1
        testn_x = [mag_testn_x; phs_testn_x]';
    end
    fr = [freq; freq];
end

```



```

        n_compress = n_compress * 2;
end

N_freq = size(fr);

if compress == 1
    Atrain_x = movmean(train_x,n_average, 2);
    Avalid_x = movmean(valid_x,n_average, 2);
    Ctrain_x = zeros(N_train, n_compress);
    Cvalid_x = zeros(N_valid, n_compress);
    % Cfreq = zeros(n_compress);
    fr_step = (N_freq-1)/(n_compress-1);
    for i = 1:n_compress
        for j = 1:N_train
            Ctrain_x(j,i) = Atrain_x(j,1+round((i-1)*fr_step(1)));
        end
        for j = 1:N_valid
            Cvalid_x(:,i) = Avalid_x(:,1+round((i-1)*fr_step(1)));
        end
        %Cfreq(i) = freq(1+round((i-1)*fr_step(1)));
        %freq = Cfreq;
    end
    train_x = Ctrain_x;
    valid_x = Cvalid_x;
    %fr = n_compress;
    %N_freq = n_compress;
end

[divide_data,mag_phs,replic,n_compress,n_VIPfreq,n_average, N_test]
for sp = spi:spi
    train_y = prop_train(:,sp);
    valid_y = prop_valid(:,sp);
    if N_test == 1
        testn_y = prop_testn(:,sp);
    end
    %algorithm selection option added by Mingqiang Han 05/24/2020
    switch algorithm_select
        case 0
            Lasso_selection_062120;
        case 1
            %PLS_Method_1;
            % PLS_Method_2;
            PLS_Method_3_071020;
            %PLS_SVM_062620_NZ_MH
            % regression_model;

    end
end

% save report8 calib valid divide_data mag_phs replic n_compress n_VIPfreq
n_average ...
%     t_min t_max v_min v_max CI ...
%     Method spi MaxR2VIPn_comp MaxR2VIPncomp MaxR2VIPnVIP MaxR2TraVIPR2
MaxR2TraRMSE1 MaxR2ValVIPR2 MaxR2ValRMSE ...
%     Signature_1_mag Signature_1_phs Signature_2_mag Signature_2_phs
Signature_3_mag Signature_3_phs ...

```

```
%      Weight_1_mag Weight_1_phs Weight_2_mag Weight_2_phs Weight_3_mag
Weight_3_phs -ascii
```

## Appendix G - Matlab program for PLS regression for the outdoor

### tests

```
%function VIP_separate(min_VIP, spi, calib, valid, freq, prop, mag_phs,
mag_digital, phs_digital)
%function VIP_separate(min_VIP, spi, calib, valid, freq, prop, mag_phs,
mag_digital, phs_digital)
% Combined program for searching "signature frequencies" using "Variable
Importance
% Projection" method to run partial least squares analysis.
%         Naiqian Zhang
%         September 27, 2019

% Added data compaction and VIP sorting.
%         10/12/19, Naiqian Zhang

%
% Change files order dependant to their order in Excel file
% Change Excel file name when use "Case=0 at switch replic"
%
%         Mohammed Hasan, October 27, 2019
% Comment out the useless code
%         Mingqiang Han, 03/22/2020
%
% finding VIP frequencies using each PLS component number
%         Mingqiang Han, 06/26/2020

% clc;
%clear all

%=====
% Method 3: (7/2/2020)
%
% 1. Run PLS with ALL freq and number of components increasing from 1 to
N_Train-1.
% 2. For each number of components, find the contributions of the freq and
%     rank the freq (VIP)
%     then run PLS with the VIP numbers increasing from 1 to n_VIPfreq and
number of component
%     increasing from 1 to N_train-1, find the highest validation R2 while the
training R2 is higher
%     than validation R2.
%
% Number of times the VIP ranking is done - N_Train-1.
%
% Note: The number of components at which the VIP are found and ranked is
n_comp
%         The number of components at which the highest R2 is found is ncomp.
%         So the highest R2 is found at a number of components that may not be
the number of components at which the VIP were selected.
```

```

n_component = N_train-1;

MaxR2VIPncomp = 0;
MaxR2VIPnVIP = 0;
MaxR2TraVIPR2 = -1000;
MaxR2ValVIPR2 = -1000;
MinRMSEValVIP = 1000;

t_min(spi) = min(train_y);
t_max(spi) = max(train_y);
v_min(spi) = min(valid_y);
v_max(spi) = max(valid_y);
rt = t_max(spi)-t_min(spi);
rv = v_max(spi)-v_min(spi);

if N_test == 1
    tn_min(spi) = min(testn_y);
    tn_max(spi) = max(testn_y);
end

if v_min(spi)>=t_max(spi) || t_min(spi)>=v_max(spi)
    CI(spi) = 0;
else
    if v_min(spi)>=t_min(spi)
        cv_min = v_min(spi);
    else
        cv_min = t_min(spi);
    end
    if v_max(spi)>=t_max(spi)
        cv_max = t_max(spi);
    else
        cv_max = v_max(spi);
    end
end
cv = cv_max-cv_min;
CI(spi) = cv^2/(rt*rv);

[spi t_min(spi) t_max(spi) v_min(spi) v_max(spi) CI(spi)]
if N_test ==1
    [tn_min(spi) tn_max(spi)]
end

for n_comp=1:n_component
    %%
    %%%%%%%%%%%%%find VIP frequencies for each PLS component%%%%%%%%%%%%
    num_var(n_comp) = n_comp;

    % Select the VIP frequencies for each component.
    [Xloadings,Yloadings,Xscores,Yscores,betaPLS,PLSPctVar,MSE,stats] =
    plsregress(...
        train_x,train_y,n_comp);
    % Calculate normalized PLS weights
    W0 = bsxfun(@rdivide,stats.W,sqrt(sum(stats.W.^2,1)));

```

```

% Calculate the product of summed squares of XS and YL
sumSq = sum(Xscores.^2,1).*sum(Yloadings.^2,1);
% Calculate VIP scores for NCOMP components
vipScores = sqrt(size(Xloadings,1) * sum(bsxfun(@times,sumSq,W0.^2),2) ./
sum(sumSq,2));

% Calculate VIP scores cumulatively for NCOMP components
vipScoresCum = sqrt(size(Xloadings,1) * sum(bsxfun(@times,sumSq,W0.^2),2)
./ cumsum(sumSq,2));
[A, VIP_freq] = sort(vipScoresCum(:, n_comp), 'descend');
%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% search for max R2 for each PLS component%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Added PLS regressions at combinations of
% VIP frequencaaaaaay (1 ~ n_VIPfreq) and PLS component (1 ~ N_train-1)
% Draw 3D surf for both Training and Validation Search for the maximum R2
% validation
%
%      6/23/2020 Naiqian Zhang
%
for n_VIP = 1:n_VIPfreq
    for ncomp=1:n_VIP-1
        if(ncomp <= N_train-1)
            trainVIP_x = train_x(:,VIP_freq(1:n_VIP));
            validVIP_x = valid_x(:,VIP_freq(1:n_VIP));
            trainVIP_y = train_y;

[Xloadings,Yloadings,Xscores,Yscores,betaPLS,PLSPctVar,MSE,stats] =
plsregress(...
            trainVIP_x,trainVIP_y,ncomp);
            trainVIP_yfit = [ones(size(trainVIP_x,1),1)
trainVIP_x]*betaPLS;

            SigmatrainYerror(n_VIP,ncomp) = sum((trainVIP_y-
trainVIP_yfit).^2)/N_train;

            TSS = sum((trainVIP_y-mean(trainVIP_y)).^2);
            RSS_PLS = sum((trainVIP_y-trainVIP_yfit).^2);
            R2_trainVIP(n_VIP,ncomp)= 1 - RSS_PLS/TSS;

            validVIP_yfit = [ones(size(validVIP_x,1),1)
validVIP_x]*betaPLS;

            SigmavalidYerror(n_VIP,ncomp) = sum((valid_y-
validVIP_yfit).^2)/N_valid;
            meanvalidy = mean(valid_y);
            TSS = sum((valid_y-meanvalidy).^2);
            RSS_PLS = sum((valid_y-validVIP_yfit).^2);
            R2_validVIP(n_VIP,ncomp) = 1 - RSS_PLS/TSS;

            if N_test == 1
                testnVIP_x = testn_x(:,VIP_freq(1:n_VIP));
                testnVIP_y = testn_y;

```

```

        testnVIP_yfit = [ones(size(testnVIP_x,1),1)
testnVIP_x]*betaPLS;

        meantestnVIP_y = mean(testnVIP_y);
        TSS = sum((testnVIP_y-meantestnVIP_y).^2);
        RSS_PLS = sum((testnVIP_y-testnVIP_yfit).^2);
        R2_testn(n_VIP,ncomp) = 1 - RSS_PLS/TSS;
        SigmatestnYerror(n_VIP,ncomp) = sum((testnVIP_y-
testnVIP_yfit).^2)/N_testn;
        end

        if (R2_validVIP(n_VIP,ncomp) > MaxR2ValVIPR2) &&
(R2_trainVIP(n_VIP,ncomp) >= R2_validVIP(n_VIP,ncomp))
            %if (SigmavalidYerror(n_VIP,ncomp)< MinRMSEValVIP) &&
(R2_trainVIP(n_VIP,ncomp) >= R2_validVIP(n_VIP,ncomp))
                %[R2_validVIP1(n_VIP,ncomp) MaxValVIPR2
R2_trainVIP1(n_VIP,ncomp)]
                MaxR2trainVIP_y = trainVIP_y;
                MaxR2trainVIP_yfit = trainVIP_yfit;
                MaxR2valid_y = valid_y;
                MaxR2validVIP_yfit = validVIP_yfit;
                if N_test == 1
                    MaxR2testn_y = testnVIP_y;
                    MaxR2testn_yfit = testnVIP_yfit;
                end
                MaxR2VIP_freq = VIP_freq;
                MaxR2A = A;

                MaxR2TraVIPR2 = R2_trainVIP(n_VIP,ncomp);
                MaxR2ValVIPR2 = R2_validVIP(n_VIP,ncomp);
                MinRMSEValVIP = SigmavalidYerror(n_VIP,ncomp);
                % MaxR2TraRMSE = sqrt(MSE(2,ncomp));
                MaxR2TraRMSE1 = sqrt(SigmatrainYerror(n_VIP,ncomp));
                MaxR2ValRMSE = sqrt(SigmavalidYerror(n_VIP,ncomp));
                MaxR2VIPn_comp = n_comp;
                MaxR2VIPncomp = ncomp;
                MaxR2VIPnVIP = n_VIP;
                if N_test == 1
                    MaxR2testn_y = testnVIP_y;
                    MaxR2testn_yfit = testnVIP_yfit;
                    MaxR2TesR2 = R2_testn(n_VIP,ncomp);
                    MaxR2TesRMSE = sqrt(SigmatestnYerror(n_VIP,ncomp));
                end
                %[spi MaxR2VIPn_comp MaxR2VIPncomp MaxR2VIPnVIP
MaxTraVIPR2 MaxValVIPR2]
            end
        end
    end
end
end

figure
residuals = MaxR2trainVIP_y - MaxR2trainVIP_yfit;
stem(residuals)
xlabel('Observation,Training set');

```

```

ylabel('Residual');
txt = ['spi=', num2str(spi)];
annotation('textbox',[0.2,0.8, 0.1, 0.1],'String', txt);

figure
residuals = MaxR2valid_y - MaxR2validVIP_yfit;
stem(residuals)
xlabel('Observation,Validation set');
ylabel('Residual');
txt = ['spi=', num2str(spi)];
annotation('textbox',[0.2,0.8, 0.1, 0.1],'String', txt);

if N_test ==1
    figure
    residuals = MaxR2testn_y - MaxR2testn_yfit;
    stem(residuals)
    xlabel('Observation,Testing set');
    ylabel('Residual');
    txt = ['spi=', num2str(spi)];
    annotation('textbox',[0.2,0.8, 0.1, 0.1],'String', txt);
end

switch spi
case 1
    x_labelT = 'Actual VWC (cm3/cm3) - Training set';
    x_labelV = 'Actual VWC (cm3/cm3) - Validation set';
    y_label = 'Predicted VWC (cm3/cm3)';
    txt_box= 'VWC';
case 2
    x_labelT = 'Actual NO3 concentration (ppm) - Training set';
    x_labelV = 'Actual NO3 concentration (ppm) - Validation set';
    y_label = 'Predicted NO3 concentration (ppm)';
    txt_box = 'NO3';
case 3
    x_labelT = 'Actual NH4 concentration (ppm) - Training set';
    x_labelV = 'Actual NH4 concentration (ppm) - Validation set';
    y_label = 'Predicted NH4 concentration (ppm)';
    txt_box = 'NH4';
end

figure
plot(MaxR2trainVIP_y,MaxR2trainVIP_yfit,'b*')
xlabel(x_labelT);
ylabel(y_label);
txt = [txt_box];
annotation('textbox',[0.2,0.8, 0.1, 0.1],'String', txt);

figure
plot(MaxR2valid_y,MaxR2validVIP_yfit,'b*')
xlabel(x_labelV);
ylabel(y_label);
txt = [txt_box];
annotation('textbox',[0.2,0.8, 0.1, 0.1],'String', txt);

if N_test == 1
    figure

```

```

    plot(MaxR2testn_y,MaxR2testn_yfit,'b*')
    xlabel('Actual value (testing set)');
    ylabel('Predicted value');
    txt = ['spi=', num2str(spi)];
    annotation('textbox',[0.2,0.8, 0.1, 0.1],'String', txt);
end

Method = 3;

[Method spi MaxR2VIPn_comp MaxR2VIPncomp MaxR2VIPnVIP MaxR2TraVIPR2
MaxR2TraRMSE1 MaxR2ValVIPR2 MaxR2ValRMSE]

if N_test==1
    [MaxR2TesR2 MaxR2TesRMSE]
end

n_mag=0;
n_phs=0;
X_mag=[];
Y_mag=[];
X_phs=[];
Y_phs=[];

for ifreq = 1:MaxR2VIPnVIP
    if(MaxR2VIP_freq(ifreq)<= 635)
        n_mag = n_mag+1;
        X_mag(n_mag) = data(MaxR2VIP_freq(ifreq),2)/100;
        if (MaxR2VIP_freq(ifreq)<=27)
            X_mag(n_mag) = (50+368.52*(MaxR2VIP_freq(ifreq)-1))/1000000;
        end
        Y_mag(n_mag) = MaxR2A(ifreq);
    else
        n_phs = n_phs+1;
        X_phs(n_phs) = data(MaxR2VIP_freq(ifreq)-635,2)/100;
        if (MaxR2VIP_freq(ifreq)<=635+27)
            X_phs(n_phs) = (50+368.52*(MaxR2VIP_freq(ifreq)-635-1))/1000000;
        end
        Y_phs(n_phs) = MaxR2A(ifreq);
    end
end

if(size(X_mag) ~= 0)
    figure
    bar(X_mag,Y_mag)
    %set(gca,'XScale','log');
    %xlim([0 635]);
    axis([0 120 0 5])
    xlabel('Frequency (MHz)')
    ylabel('VIP importance (Magnitude)')
    txt = ['spi=', num2str(spi)];
    annotation('textbox',[0.2,0.8, 0.1, 0.1],'String', txt);
end

if(size(X_phs) ~= 0)
    figure

```



```

bar(X_phs,Y_phs)
%set(gca,'XScale','log');
%xlim([0 635]);
axis([0 120 0 5])
xlabel('Frequency (MHz)')
ylabel('VIP importance (Phase)')
txt = ['spi=', num2str(spi)];
annotation('textbox',[0.2,0.8, 0.1, 0.1],'String', txt);
end

switch spi
case 1
file_VIP = 'VIP_cumulative1.xlsx';
case 2
file_VIP = 'VIP_cumulative2.xlsx';
case 3
file_VIP = 'VIP_cumulative3.xlsx';
end

VIP_score = xlsread(file_VIP);
for i = 1:1270
New_VIP_score(i,1) = 0;
end

if mag_phs == 2
for i = 1:MaxR2VIPnVIP
New_VIP_score(MaxR2VIP_freq(i)+635,1) = MaxR2A(i)';
end
else
for i = 1:MaxR2VIPnVIP
New_VIP_score(MaxR2VIP_freq(i),1) = MaxR2A(i)';
end
end

VIP_score = [VIP_score New_VIP_score];
xlswrite(file_VIP, VIP_score);

for i = 1:1270
Mean_VIP(i) = mean(VIP_score(i,:));
end
Mean_VIP = Mean_VIP';

n_mag = 0;
n_phs = 0;
X_mag=[];
Y_mag=[];
X_phs=[];
Y_phs=[];

for i = 1:1270
if i<= 635
n_mag = n_mag+1;
X_mag(n_mag) = data(i,2)/100;
if (i<=27)
X_mag(n_mag) = (50+368.52*(i-1))/1000000;

```

```

        end
        Y_mag(n_mag) = Mean_VIP(i);
    else
        n_phs = n_phs+1;
        X_phs(n_phs) = data(i-635,2)/100;
        if i<=635+27
            X_phs(n_phs) = (50+368.52*(i-635-1))/1000000;
        end
        Y_phs(n_phs) = Mean_VIP(i);
    end
end
end

```

```

figure
plot(X_mag,Y_mag)
%set(gca,'XScale','log');
%xlim([0 635]);
axis([0 120 0 3])
xlabel('Frequency (MHz)')
ylabel('Mean VIP importance (Magnitude)')
txt = ['spi=', num2str(spi)];
annotation('textbox',[0.2,0.8, 0.1, 0.1],'String', txt);

```

```

figure
plot(X_phs,Y_phs)
%set(gca,'XScale','log');
%xlim([0 635]);
axis([0 120 0 3])
xlabel('Frequency (MHz)')
ylabel('Mean VIP importance (Phase)')
txt = ['spi=', num2str(spi)];
annotation('textbox',[0.2,0.8, 0.1, 0.1],'String', txt);

```

```

switch spi
case 1
    Signature_1_mag = X_mag;
    Signature_1_phs = X_phs;
    Weight_1_mag = Y_mag;
    Weight_1_phs = Y_phs;
case 2
    Signature_2_mag = X_mag;
    Signature_2_phs = X_phs;
    Weight_2_mag = Y_mag;
    Weight_2_phs = Y_phs;
case 3
    Signature_3_mag = X_mag;
    Signature_3_phs = X_phs;
    Weight_3_mag = Y_mag;
    Weight_3_phs = Y_phs;
end

```

```

% figure

```

```

%
% plot(num_var, MaxTraVIPR2, num_var,  MaxValVIPR2)
% xlabel('number of components')
% ylabel('max PLS R^2 for each component')
% legend('train data','valid data')

% [spi MaxR2ncomp MaxCalR2 MaxValR2]

% Step1_SPI_Magphs_ncomp_MaxCalR2_MaxValR2 = [spi mag_phs MaxR2ncomp MaxCalR2
MaxValR2]

% figure
% plot(num_var,R2_train,num_var,R2_valid)
% xlabel('number of components')
% ylabel('PLS R^2')
% legend('train data','valid data')

% figure
% plot(num_var,R2_train1,num_var,R2_valid1)
% xlabel('number of components')
% ylabel('PLS R^2')
% legend('train data','valid data')

```

## Appendix H - VWC, NO<sub>3</sub>, and NH<sub>4</sub> values for outdoor test samples

### VWC Values for all samples for outdoor tests

Sensor 1				Sensor 2				Sensor 3				Sensor 4			
Set1	Set 2	Set3	Set4	Set1	Set 2	Set3	Set4	Set1	Set 2	Set3	Set4	Set1	Set 2	Set3	Set4
12.7	21	30.5	21.3	11.6	29.6	35.9	22.5	11.2	26.3	35	19.9	13.5	29	37.4	23
21.4	27.9	28.1	24.8	15.3	31.4	37	22.8	21.6	28.5	36	22	16.2	33	33.8	16.8
22.6	28.5	21.2	22.8	14.3	36.2	37.7	25	21.8	24.4	35	22.7	16	36.4	31.9	22.3
22.6	31.4	27.2	23.3	19.5	33.2	33	26	22.8	33	33	23.3	20.4	36.2	30	24.4
28.9	26.4	25.6	23	28.4	34.7	27.3	25.4	30.8	31.3	34	20.8	26.4	34.7	30.4	23.4
27.9	29.4	24	22.9	25.5	35.4	23.6	20	26.9	26.7	24	20.6	30.2	34.1	28.1	23.7
25.6	24.3	24.5	23.3	37.3	33.8	27.9	15.5	29.4	26.5	34	23.7	33.8	33.5	23	24.5
25.5								27.2				29			
12.7-28.9	21-31.4	21.2-30.5	21.3-24.8	11.6-28.4	29.6-36.2	23.6-37.7	15.5-26	11.2-30.8	24.4-33	24-36	19.9-23.7	13.5-33.8	29-36.4	23-37.4	16.8-24.5
12.7 – 30.5				11.6 – 37.7				11.2-36.0				13.5 – 37.4			

### NO<sub>3</sub> Values for all samples for outdoor tests

Sensor 1				Sensor 2				Sensor 3				Sensor 4			
Set1	Set 2	Set3	Set4	Set1	Set 2	Set3	Set4	Set1	Set 2	Set3	Set4	Set1	Set 2	Set3	Set4
17.8	40.1	6.6	9.1	19.2	92.4	6.2	3.8	17.6	42.7	7.5	4.1	16.8	102.5	12.2	2.4
44.8	21.3	2.2	4.7	33.9	13.8	3.1	2.5	67.6	33.1	4.5	2.8	43.5	75.8	5.6	2.1
54.6	17.2	1.3	5	36.8	12.7	1.1	4.5	76	33.1	0.8	3.2	44.4	50.5	1	2
59.6	48.3	9	1.8	62	49.2	5.2	3.2	64.1	38.4	10.2	1.7	78.8	21.4	2.8	1.2
14.2	26.8	6.2	6.4	77.7	34.6	1.7	2	38.5	30.3	3.8	1.3	91.3	41.7	5.9	1.5
37.4	18.6	1.3	4.9	63.9	14.3	1.5	2.1	67.6	31.8	1.3	2.2	90	46.8	5.3	1.6
17.8	23.9	1.2	51.3	85.8	28.1	1.4	2.1	51.8	28.3	4.8	8.2	116.3	34.3	1.1	7.4
41.1								28.1				97.2			
14.2-59.6	17.2-48.3	1.2-9	1.8-51.3	19.2-85.8	12.7-92.4	1.1-6.2	2-4.5	17.6-67.6	28.3-42.7	0.8-10.2	1.3-8.2	16.8-116.3	21.4-102.5	1-12.2	1.2-7.4
1.2 – 59.6				1.1 – 92.4				0.8 – 67.6				1.00 – 116.3			

### NH<sub>4</sub> Values for all samples for outdoor tests

Sensor 1				Sensor 2				Sensor 3				Sensor 4			
Set1	Set 2	Set3	Set4	Set1	Set 2	Set3	Set4	Set1	Set 2	Set3	Set4	Set1	Set 2	Set3	Set4
9.7	3.4	5.9	5.6	8.5	6	4.7	5.9	9.7	3.2	5.3	5.3	11	4.5	5.2	5.5
5.6	7.9	4.5	9	5.7	7.7	4.3	7	6.4	3.5	4.6	7.5	6.9	6.2	4	5.7
6.5	7.9	6.1	5.1	6.5	5.1	5.1	9.1	8	6.6	7	6.2	6.4	6.2	6.6	8.3
9.5	8.2	5.6	5.1	8.7	7.6	4.4	6.1	8.7	6.8	4.6	7.1	6.5	9.2	5	7.4
8	16.4	6.3	11.2	8.5	3.5	5.4	6.9	10.1	2.9	6.4	8.7	7.4	9.8	11.1	6.7
8.2	4.8	5.5	6.4	8.3	5	7.2	7	9.8	4.8	4.8	5.8	8	15.4	4.8	5.4
8.5	5.3	3.3	4.4	9.1	12.4	3.6	4.9	7.9	12.1	3.3	4.5	8.6	9.6	4	4.1
10.4				9.2				7.4				8.4			
5.6- 10.4	3.4- 16.4	3.3- 6.3	4.4- 11.2	5.7- 9.2	3.5- 12.4	3.6- 7.2	4.9- 9.1	6.4- 10.1	2.9- 12.1	3.3- 6.4	4.5- 8.7	6.4- 11	4.5- 15.4	4- 11.1	4.1- 8.3
3.3 – 16.4				3.5 – 12.4				2.9 – 12.1				4.00 -15.8			