ESTIMATION OF SOIL MOISTURE INDICES USING DIFFUSE REFLECTANCE SPECTROSCOPY

by

Sarathjith M.C.

Department of Irrigation and Drainage Engineering
KELAPPAJI COLLEGE OF AGRICULTURAL ENGINEERING AND TECHNOLOGY
TAVANUR - 679 573, MALAPPURAM
KERALA, INDIA
2019

ESTIMATION OF SOIL MOISTURE INDICES USING DIFFUSE REFLECTANCE SPECTROSCOPY

by

Sarathjith M C

(2017-18-003)

THESIS

Submitted in partial fulfillment of the

requirement for the degree of

Master of Technology

In

Agricultural Engineering

Faculty of Agricultural Engineering and Technology

Kerala Agricultural University

Department of Irrigation and Drainage Engineering

KELAPPAJI COLLEGE OF AGRICULTURAL ENGINEERING AND TECHNOLOGY

TAVANUR - 679 573, MALAPPURAM

KERALA, INDIA

2019

DECLARATION

I hereby declare that this thesis entitled **"Estimation of soil moisture indices using diffuse reflectance spectroscopy"** is a bonafide record of research work done by me during the course of research and that the thesis has not previously formed the basis for the award of any degree, diploma, associateship, fellowship or other similar title of any other University or Society.

Place: Tavanur — Sarathjith M C

Date : — (2017-18-003)

CERTIFICATE

Certified that this thesis, entitled **"Estimation of soil moisture indices using diffuse reflectance spectroscopy"** is a record of research work done independently by **Mr. Sarathjith M C (2017-18-003)** under my guidance and supervision and that it has not previously formed the basis for the award of any degree, fellowship or associateship to him.

Place: Tavanur
Date :

Dr. Anu Varughese
(Major Advisor, Advisory Committee)
Assistant Professor (Agrl. Engg.)
Dept. of IDE
KCAET, Tavanur

CERTIFICATE

We the undersigned members of the advisory committee of **Mr. Sarathjith.M.C (2017-18-003)** a candidate for the degree of **Master of Technology in Agricultural Engineering** agree that this thesis entitled **"Estimation of soil moisture indices using diffuse reflectance spectroscopy"** may be submitted by **Mr. Sarathjith.M.C (2017-18-003),** in partial fulfillment of the requirement for the degree.

Dr. Anu Varughese
(Chairman, Advisory Committee)
Assistant Professor (Agrl. Engg.)
Dept. of IDE
KCAET, Tavanur

Dr. Sasikala D
(Member, Advisory Committee)
Professor and Head
Dept. of IDE
KCAET, Tavanur

Er. Vishnu B
(Member, Advisory Committee)
Associate Professor (Agrl. Engg.)
RARS, Kumarakom

Mrs. Josephina Paul
(Member, Advisory Committee)
Assistant Professor
Department of SAC
KCAET, Tavanur

EXTERNAL EXAMINER
(Name and Address)

ACKNOWLEDGEMENT

With legitimate pride and responsibility, I extend my profound sense of gratitude to the Chairman of my Advisory Committee, Dr. Anu Varughese, Assistant Professor, Department of Irrigation and Drainage Engineering (Dept. of IDE), Kelappaji College of Agricultural Engineering & Technology (KCAET), Tavanur for her prompt response, motivation, constructive criticisms, constant review of work, inspiring guidance and moral support extended since the onset my research.

I express my sincere appreciation to Dr. Sathian K.K. (The Dean, KCAET, Tavanur) for his valuable suggestions and providing all the necessary facilities to pursue this research. I am immensely thankful to all the members of Advisory Committee; Dr. Sasikala D. (Professor and Head, Dept. of IDE, KCAET, Tavanur), Er. Vishnu B. (Associate Professor, Regional Agricultural Research Station, Kumarakom) and Mrs. Josephina Paul (Assistant Professor, Department of Supportive and Allied Courses, KCAET, Tavanur) for their creative suggestions and moral support provided during my research.

I am grateful to all my classmates namely, Er. Rank Prasang Harji, Er. Anjana S.R., Er. Adarsh, S.S., and Er. Akhila Shiney and other post graduate students of KCAET, Tavanur for their sensible help and assistance during the conduct of this research.

I fail on my part if I do not mention my better half Mrs. Pritty S Babu for her constant encouragement and support. It is my privilege to express my sincere gratitude to my beloved parents for their affection, prayers and moral support throughout my research. Above all, I humbly acknowledge the grace and blessings of the supreme power that capacitates to fulfill this well nurtured dream.

Sarathjith M.C.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SYMBOLS AND ABBREVATIONS

%	percentage
&	ampersand
+	addition
<	less than
≤	less than or equal to
=	equal to
>	greater than
A	absorbance
AIC	Akaike information criterion
α	level of significance
AMI	adjacency value of mutual information
atm	atmospheric pressure
β	regression coefficient
bicor	biweightmidcorrelation vector
CARS	competitive adaptive reweighted sampling
cm	centimeter
CovProc	covariance procedure
DRS	diffuse reflectance spectroscopy
DT	de-trending
EDF	exponential decreasing function
et al.	and others
etc	et cetera
F	*F*-statistic
FC	field capacity
FD	first derivative
Fig.	figure
g	gram

H_0	null hypothesis
H_2SO_4	sulfuric acid
LV	latent variables
m^2	square meter
MI	mutual information
MIR	mid infrared
mm	millimeter
MPA	multi-purpose analyzer
MSC	multiplicative scatter correction
MSE	mean squared error
NDRI	normalized difference reflectance index
NINSO	near and shortwave infrared domain for soil
L	moisture content estimation from linear regression
NINSO	near and shortwave infrared domain for soil
N	moisture content estimation from non-linear regression
NIR	near infrared
nm	nanometer
NSMI	normalized soil moisture index
OM	organic matter
p	*p*-statistic
PCA	principal component analysis
PCR	principal component regression
PLSR	partial least squares regression
R	reflectance
r	Pearson correlation coefficient
R^2	coefficient of determination
rcoplot	residual case order plot

RMSE	root mean squared error
RPD	residual prediction deviation
SD	second derivative
SNV	standard normal variate
SqRes	squared residual
StN	signal-to-noise vector
USDA	United States Department of Agriculture
VIP	variable importance for projection
viz.	videlicet
WHC	water holding capacity
WISOIL	water index SOIL
WP	wilting point

INTRODUCTION

CHAPTER 1

INTRODUCTION

Rapid and accurate estimation of water holding capacity (WHC) of soil is a key aspect of scientific irrigation scheduling, which aims to supply right quantity of water at right time to crop plants to maximize production with minimal environmental impacts (Campbell and Campbell, 1982). The WHC, being the total amount of water that a soil hold is defined by field capacity (FC) and wilting point (WP) as the upper and lower limits (Gupta *et al.*, 2016; Salter and Williams, 1965a). Thus, information on FC and WP forms a critical input to irrigation scheduling. An optimized irrigation schedule aided by continuous monitoring of FC and WP partakes multiple benefits; improves water conservation (Gupta *et al.*, 2016), reduce costs of pumping, reduce competition for water and reduce environmental effects, among others. Over decades, several approaches have been devised for the determination of these soil moisture indices (also referred as soil moisture constants) directly and indirectly. They includes, field test plot (Salter and Williams 1965b), sunflower method (Taylor and Ashcroft, 1972), desiccator method (Lehane and Staple, 1951), pressure plate apparatus (Richards and Fireman, 1943), pedotransfer functions (Givi *et al.*, 2004; Tietje and Tapkenhinrichs, 1993; Gupta and Larson, 1979) and by means of saturation percentage of soil (Grewal *et al.*, 1990; Dahiya *et al.*, 1988). The soil moisture constants being dynamic in nature (Kirkham, 2014), their near-real time assessment via aforesaid conventional methods remains a challenging task. Moreover, most of the conventional methods are laborious and time consuming and hence not advisable for the estimation of the soil moisture constants at different spatial and temporal scales.

With advancements in past few decades, remote sensing technique appear to have the potential to address the aforesaid challenges. Specifically, hyperspectral sensors operating in the visible, near-infrared and shortwave-infrared wavelength domain (400-2500 nm) has demonstrated their ability to characterize different soil attributes using measurements in laboratory (Kinoshita *et al.*, 2012; Viscarra Rossel *et al.*, 2006b), field (Lagacherie *et al.*, 2008), airborne (Nouri *et al.*, 2017; Vaudour

et al., 2016) and space borne (Nowkandeh *et al.*, 2018). In hyperspectral remote sensing, spectral signature of target (soil) collected at very fine spectral bands is used in conjunction with established retrieval algorithms to estimate the attribute of interest. Thus, the major pre-requisite in hyperspectral remote sensing is to develop retrieval algorithms (also referred as calibration functions) to translate spectral signature to attribute values. Generally, such calibration functions are initially built and evaluated using spectral signature measured under laboratory conditions by means of spectroradiometers (Castaldi *et al.*, 2016). The approach is often referred as diffuse reflectance spectroscopy (DRS) due to the 'diffuse' nature of spectral signature. In the past few decades, the DRS has been renowned as a prominent tool for soil analysis and digital soil mapping with its inherent advantages; accurate, simple, rapid, cost-effective (Pittaki-Chrysodonta *et al.*, 2018), non-destructive, non-invasive (Ben-Dor *et al.*, 2009), alternative to conventional techniques (Brown *et al.*, 2006a), estimation of multiple attributes simultaneously and amenable to different modes of remote sensing (Viscarra Rossel *et al.*, 2006b) after making necessary corrections for atmospheric and other interferences.

Over years, several studies have demonstrated the ability of DRS to assess different soil attributes. Some selected examples from recent literature includes the assessment of soil organic carbon (Li *et al.*, 2015; Singh *et al.*, 2013), nutrient contents (Abdi *et al.*, 2012; Mouazen *et al.*, 2007), electrical conductivity (Farifteh *et al.*, 2010; Shrestha 2006), cation exchange capacity (Bilgili *et al.*, 2010; Fox and Metla 2005), soil mineralogy (Vendrame *et al.*, 2012; Clark 1999), carbonates (Lagacherie *et al.*, 2008), moisture content (Fabre *et al.*, 2015), texture (Gholizadeh *et al.*, 2016; Lacerda *et al.*, 2016; Bilgili *et al.*, 2010), parameters of aggregate size distribution (Sarathjith *et al.*, 2014), hydraulic properties (Santra *et al.*, 2009), among others. More details on DRS based soil assessment can be find in the review of Ben-Dor *et al.* (2009) and Stenberg (2010). Although efforts were made to assess the parameters of water retention function (Pittaki-Chrysodonta *et al.*, 2018; Babaeian *et al.*, 2015; Santra *et al.*, 2009) using DRS, very limited studies have attempted the estimation of FC and WP directly from spectral signature (Kinoshita

et al., 2012; Viscarra Rossel and Webster 2012; Janik *et al.*, 2009) and hence demand further investigation.

The DRS rely on statistical algorithms that relate spectral signature and soil attributes. Such methods may broadly be classified as spectral indices based approach, full-spectrum based feature projection approach and variable (spectral feature) selection based approach. The spectral indices based approach use either single band relative reflectance (Weidong *et al.*, 2002) or multi-band features (suitably combined to form a ratio or normalized difference index) as predictors in a simple linear regression framework to estimate soil attributes. The single band relative reflectance depends on the accuracy of the reference spectrum measurement and hence not appropriate for samples with high spatial variability. In contrast, multi-band features are relatively stable across soil types (Haubrock *et al.*, 2008). One major difficulty associated with spectral index based approach is to find appropriate spectral features to be combined with regard to the heterogeneity of samples in the database. However, the approach has not been tested for the estimation of FC and WP. In the second approach (feature projection) comprise of transforming the multi-collinear spectral data in the entire spectral range into uncorrelated variables (usually by an orthogonal transformation). These uncorrelated variables may be used to establish the desired linkage in a multivariate regression framework. Some of the selected feature projection algorithms used in DRS of soils includes, principal component regression (PCR) (Chang *et al.*, 2001), partial least squares regression (PLSR) (Abdi *et al.*, 2012; Mouazen *et al.*, 2010), boosted regression trees (Brown *et al.*, 2006a) support vector machines (Sarathjith *et al.*, 2016b; Genot *et al.*, 2011), multivariate adaptive regression splines (Nawar and Mouazen 2017; Shepherd and Walsh 2002) and artificial neural networks (Viscarra Rossel and Behrens 2010; Daniel *et al.*, 2003). Among them, the PLSR appeared to be the most versatile and popular technique in DRS of soils due to its ability to account for multicollinearity issue, better interpretability and efficient computation (Stenberg *et al.*, 2010; Viscarra Rossel *et al.*, 2006b). The third approach namely, feature selection or spectral variable selection is an efficient way

to diminish complexity and improve robustness of DRS models (Xiaobo *et al.*, 2010) with no compromise for prediction accuracy (Fernández Pierna *et al.*, 2009). Several variable selection approaches exists of which competitive adaptive reweighted sampling (CARS) method (Li *et al.*, 2009) was found to be satisfactory in soil dataset (Vohland *et al.*, 2014) subject to the limitation associated with the use of random numbers in variable selection. Very recently, Sarathjith *et al.* (2016a) suggested an ordered predictor selection (OPS) approach which employed an exponential decreasing function (EDF) to overcome the limitation in CARS approach. The OPS approach was tested in soil datasets and found successful in developing parsimonious DRS models with similar or improved accuracy. However, the OPS approach has not been evaluated for the estimation of soil moisture constants and hence warrant further investigation. To the best of my knowledge, no studies have been attempted the comparative performance of the aforesaid approaches in the DRS based estimation of soil moisture constants. Hence, this study was undertaken with the following objectives.

1. To evaluate the use of spectral indices to characterize moisture content at FC and WP

2. To develop DRS models for the estimation of soil moisture content at FC and WP

3. To evaluate spectral variable selection on the performance of DRS models of soil moisture content at FC and WP

REVIEW OF LITERATURE

CHAPTER 2

REVIEW OF LITERATURE

2.1 Soil Moisture Indices

Soil water content is a key factor for scientific interventions related to irrigation scheduling, agronomy and hydrology. Moreover it has a significant role in defining the land productivity via the ability of soil to hold and release water upon crop transpiration demand (Ritchie 1981). In general, water in soil is categorized as gravitational, capillary and hygroscopic water based on its occurrence in soil matrix. Gravitational water typically occurs in the macro-pores and has insignificant contribution to water uptake by plants. It moves through the soil under the force of gravity and rapidly drains down to the water table. Capillary water is held in the soil micro-pores due to surface tension and forms the plant available water. Under soil drying conditions, capillary water becomes gravitational water due to increase in soil pore size (lower surface tension). Hygroscopic water is a thin layer of water tightly bound on soil particles (due to adhesion) and not occur in pores. It has no contribute to plant available water. Under given conditions, both capillary and hygroscopic water is considered to be in equilibrium with soil. Equilibrium points namely, maximum capillary capacity and hygroscopic coefficient are defined at which soil has maximum amount of capillary and hygroscopic water, respectively. The soil water content at equilibrium points is referred as soil moisture constant (http://ecoursesonline.iasri.res.in/mod/page/view.php?id=14178, last accessed on June 14, 2019). The aforesaid soil water classification and related soil moisture constants flaws to define the soil water uptake by plants under field conditions. Because, part of the water at lower and upper limits of capillary water are not plant available. To overcome the shortcomings, two additional soil moisture constants have been defined namely, FC and WP.

The FC is defined as the capacity of soil to hold water against the force of gravity. It is a state at which only capillary/micro-pores of soil are filled with water. The FC is the upper limit of plant available water (Salter and Williams, 1965a) at which plant start to use water in the soil for normal functions (Veihmeyer and Henderickson, 1949). Several factors affect FC viz. previous soil water history, soil texture, structure, type of clay, organic matter (OM) and temperature (Kirkham, 2014). Soil water content at FC is not an equilibrium value as there is no cessation of water movement through soils (Veihmeyer and Henderickson, 1949). The dynamic nature of soil water is affected by the constant influence of either or both water addition (precipitation and irrigation) and removal (drainage and evapotranspiration) processes. Thus, a range of soil water content values are associated with FC (Kirkham, 2014) with matric potentials range between 0.10 atm (sandy soil) and 0.33 atm (clay soil). However, in practice, soil water content at 0.33 atm (one-third bar) is widely used for the estimation of FC.

The WP, also referred as permanent wilting point (Kirkham, 2014) indicates the lower limit of plant available water (Salter and Williams, 1965a) below which plant wilt. It is the state of soil at which it is incapable of water supply to the plant. Indicator plants such as sunflower is often used to determine the WP. To avoid practical difficulties of the approach, the WP may be approximated from FC value by dividing with a factor ranging between 2.0 and 2.4 for soils with low and high silt contents, respectively (Israelsen and Hansen, 1962). Similar to FC, the WP is also not an equilibrium value and considered to vary between matric suction from 7 to 40 atm as influenced by the crop, consumptive use, soil texture and salt content. However, soil moisture tension at 15 atm is generally regarded for the estimation WP.

Several conventional methods exists for the determination of FC and WP which may be broadly classified as direct and indirect methods as briefly discussed below. The direct methods includes field test plot (to determine FC), sunflower method, desiccator method (to determine WP) and pressure plate apparatus (to determine both FC and WP). The methods namely, pedotransfer functions,

parameterization of soil moisture characteristic curve, saturation percentage and spectral reflectance based approach are indirect ways of estimating the soil moisture constants.

2.1.1 Field Test Plot

This method is used for the field determination of FC. In this method, a test plot (usually 2.5 m^2) on a bare field is flooded until the desired soil layers gets saturated. The plot can be covered using plastic sheets or mulches to evade evaporation. The moisture content of the samples down the profile are measured at specified intervals (12-24 hours) until the values of two sequential samples are almost equal. The lowest influx value is regarded as the FC. This procedure depend on soil texture and structure (Salter and Williams, 1965b).

2.1.2 Sunflower Method

This method WP determination make use of indicator plants such as sunflower (*Helianthus annuus*) grown in containers with soil (about 500g) and hence referred as sunflower method. In this method, the sunflower plant is watered adequately only up to the emergence of third set of true leaves during which evaporation from soil is limited using wax or by sealing the container. Thereafter, the plant continue in the low evaporative environment without additional water supply until it wilt. The ability of the plant to recover is examined by transferring it to a humid and dark chamber. If the plant remain wilted and does not recover overnight, the soil moisture content is considered to be at WP (Taylor and Ashcroft, 1972).

2.1.3 Desiccator Method

In desiccator method (Lehane and Staple, 1951), vapor equilibrium between soil sample and known concentrations of H_2SO_4 in the desiccator is examined. In this method, duplicates of air-dried and sieved (2 mm) soil sample and standard soil sample of known WP taken in open, wide-mouth and shallow weighing bottles are

kept in desiccator containing H_2SO_4 (1.25% by weight). The desiccator is evacuated and kept in a dark room for 14 days during which the pressure is monitored at regular interval. Once the equilibrium is reached, the weighing bottles are subjected to gravimetric estimation of moisture content. The WP of the soil sample is then related to that of standard soil by considering the ratio of WP to moisture content are equal between them.

2.1.4 Pressure Plate Apparatus

The pressure plate apparatus (Richards and Fireman, 1943) can be used to determine both FC and WP in the laboratory. In this method, soil samples are saturated and a suction (moisture tension) is applied using a pressure plate. As a result, a part of the soil water is removed while the other is retained in the sample depending on the applied soil moisture tension. Once the water outflow is ceased, the soil sample is subjected to gravimetric moisture content estimation. Similarly, soil water content variation at different matric suctions can be recorded (water retention curve). The moisture content of soil at matric suction of 0.33 and 15 atm are most commonly used to represent FC and WP (Richards and Weaver, 1943), respectively.

2.1.5 Pedotransfer Functions

Pedotransfer functions relate soil water retention (hydraulic data) with other soil attributes (Tietje and Tapkenhinrichs, 1993). Thus, soil proxy data (soil survey data, physical, structural or compositional attributes) can be suitably translated into soil hydraulic characteristics (Schaap *et al.*, 2001). They can be categorized into three groups (Cornelis *et al.*, 2001). The first group estimate soil water content at specific matric potentials, the second group estimate the parameters of analytical equation (eg. Brooks and Corey model, van Genuchten equation) characterizing soil moisture relationship and the third group consists of physical-conceptual models. In group 1 & 2, the multiple linear regression (Gupta and Larson, 1979) and neural network (Pachepsky *et al.*, 1996) are the most commonly used techniques to build

pedotransfer functions while the last group rely on scaling and fractal mathematics (Tyler and Wheatcraft 1989). The empirical nature, specific data requirement and modest accuracy are the main reasons that confront the use of pedotransfer functions (Givi *et al.*, 2004).

2.1.6 Saturation Percentage

The saturation percentage is the ratio of amount of water required to saturate a soil to its dry weight. The use of saturation percentage as a predictor variable is a simple approach to estimate of FC and WP. Both linear (Karkanis, 1983) and log-linear (Dahiya *et al.*, 1988) relationship between predictor (saturation percentage) and response (FC or WP) variables have been reported for undisturbed soil samples. In undisturbed soil, the relationship with saturation percentage was noted to be linear for both FC (R^2=0.94) and WP (R^2=0.91) (Grewal *et al.*, 1990).

As discussed above, several methods have been used to determine soil constants. However, the selection of most appropriate method depends on several factors including time, labor, data availability, spatial coverage, accuracy, cost etc. Although accurate, the field methods to determine FC (field test plot) and WP (sunflower method) are labor intensive and time consuming and not appropriate for its application over different spatial scales. The pressure plate apparatus allows laboratory scale determination of both disturbed and undisturbed soil samples. Moreover, it allows determination of both FC and WP in one sample setting. However, the method is laborious and time-consuming (Salter, 1967). The pedotransfer function approach has gained much relevance as it enable the prediction of soil moisture contents given the proxy soil data such as soil survey data, structural, physical and compositional attributes. But, generation of surrogate data itself is time consuming (Givi *et al.*, 2004). The method of parameterization of soil water characterization curve have similar limitations as that of pressure plate apparatus as the water retention characteristics of soil (generated by pressure plate apparatus) is a primary input. Thus, an alternative method is needed to overcome the aforesaid shortcomings of the conventional techniques. Remote sensing in the

solar domain (400-2500 nm) has the potential to be a comprehensive solution to these issues with its ability for non-destructive and rapid measurements with good spatial and temporal coverage. The fundamental requirement for remote sensing approach is the generation of calibration functions to translate spectral reflectance values into reasonable estimates of soil moisture constants. Such calibration functions can be derived using DRS. The following sections covers several aspects of DRS including fundamentals, data analysis, calibration function development and evaluation.

2.2 Basics of Infrared Reflectance Spectroscopy

The discovery of infrared radiation via Herschel's experiment and the two associated conclusions (water absorbs radiation and the absorption is wavelength dependent) paved the way for infrared reflectance spectroscopy. The term reflectance spectroscopy refers to a technique of measurement, analysis and inference of interaction of infrared radiation of the electromagnetic spectrum (750–100000 nm) with the target of interest. As the target interacts with infrared radiation, its internal energy increases and consequently lead to vibrational transitions at molecular level (Stuart, 2004). This result in absorption of the incoming radiation depending on the composition of the target. Two criteria are to be met for the absorption of infrared radiation to happen, 1) dipole moment of the molecule changes due to vibrational transitions; 2) frequency of both the vibrational mode and incoming radiation matches (Bokobza, 2002; Johnston and Aochi, 1996). The dipole moment indicates the difference in absolute charge of atoms in a molecule with respect to the distance between them. The induced dipole moment change defines the degree of infrared absorption; higher the change in dipole moment result in stronger absorption and vice versa. Hence, no absorption of infrared radiation takes place in case of homo-nuclear molecules as no dipole moment change is induced due to vibration mode. The vibrational modes are resultant of either stretching or bending vibrations. The former alters the interatomic bond distance symmetrically or asymmetrically in a continuous

manner. The latter cause changes in the bond angle by rocking, scissoring, twisting and wagging. The induced dipole moment change in the molecule is higher in case of stretching than that of bending. Also, asymmetric vibrations are stronger than symmetric vibrations within stretching category (Stuart, 2004). Those vibrations that induce dipole moment change in the target are regarded as active vibrations and the subjected molecules or functional groups are referred as 'spectrally active' in the infrared range.

2.3 Infrared Wavelengths for Soil Analysis

Infrared-active vibrations suitable for soil analyses typically occurs in the near-infrared (NIR, 700–2500 nm) and mid-infrared (MIR, 2500–25000 nm) wavelength range. The NIR and MIR spectroscopy differ in the nature of energy interaction. The MIR region is characterized by the fundamental absorptions while the NIR region embodies the overtones and combinations of fundamental vibrations in the MIR frequencies (Williams and Norris, 1987). The energy required for fundamental vibration is low and the absorptions are generally sharp and frequency specific. In contrast, energy required for overtones is high which result in broad (less specific) and weak absorption features in the NIR spectrum (Brown, 2007).

Several studies have compared the performance of NIR and MIR spectroscopy in soil analysis (Soriano-Disla *et al.*, 2014; Gholizadeh *et al.*, 2013; Reeves III 2010; Viscarra Rossel *et al.*, 2006b). In many cases, the MIR outperformed NIR despite opposite result reported in the literature (Soriano-Disla *et al.*, 2014). The reason for inferior performance of NIR might be attributed to less specificity of wavelengths (Viscarra Rossel *et al.*, 2006b) and diffusion of light associated with soil physical structure and moisture content (Bellon-Maurel and McBratney 2011; Williams and Norris, 1987). However, NIR spectroscopy is more convenient in practice than MIR as the former enable non-destructive and non-invasive analysis both in-situ and laboratory conditions. The other benefits of NIR includes, rapid measurement, less sample preparation, estimation of multiple attributes from single spectrum (Viscarra Rossel *et al.*, 2006b). Moreover, NIR

wavelengths are amenable to different mode of remote sensing. All these advantages collectively resulted in wide acceptance of NIR spectroscopy as a prominent tool for soil analysis and digital soil mapping. In this study, the utility of NIR spectroscopy to estimate soil moisture constants is examined and hence the following discussion is limited to the same. As the spectral signature of soil obtained from non-destructive measurement is 'diffuse' in nature (due to its inherent composition and physical structure), NIR spectroscopy is commonly known as diffuse reflectance spectroscopy (DRS).

2.4 Spectral Signature of Soil

The spectral signature is a unique characteristic of soil (Ben-Dor *et al.*, 2009). It may be regarded as an integrated response of type of parent rock, extend of weathering, physical, chemical, biological, mineralogy and structure and hence differs across soils. The NIR spectral signature of soil is defined by the overtones and combinations of fundamental vibrations in the MIR wavenumbers (Santra *et al.*, 2015; Ladoni *et al.*, 2010; Ben-Dor *et al.*, 1999) associated with the covalent bonds of C–H, N–H and O–H functional groups (Workman and Shenk 2004; Malley and Martin, 2003). A typical soil spectrum consists of three noticeable absorptions around 1400, 1900 and 2200 nm. The first two absorptions (1400 and 1900 nm) are designated as water absorption peaks as they are resultant of first overtone of hydroxyl (O–H) stretching vibrations and combination of the same with H–O–H bending modes (Clark,1999). The O–H stretch in combination with metal–OH bending vibrations (related to clay mineral) cause characteristic absorptions within 2200–2300 nm (Stenberg, 2010; Chabrillat *et al.*, 2002). Apart from water absorption features, the spectral signature of soil is mainly influenced by iron oxides in the near-infrared (870–1000 nm) and carbonates in short wave infrared region, specifically in 1850–1870 nm, 1970–2000 nm and 2120–2160 nm (weak combination bands) and in 2300–2350 nm due to overtone (Chang and Laird, 2002; Clark, 1999; Clark *et al.*, 1990).

All the aforementioned spectral features are associated with absorption of incident electromagnetic energy due to the presence of spectrally active soil constituents or chemical chromophores. It mostly consists of soil organic matter, iron oxides, moisture and clay content. There are also other factors known as physical chromophores that affect the whole spectrum (Hill *et al.*, 2010) rather than absorption at specific wavelengths; such as azimuth angle of the source, incident angle, intensity of radiation, viewing angle, particle size and sample geometry (Ben-Dor, 2011). Thus, any spectrum is a combined response of the interaction of physical and chemical chromophores with the energy of incident electromagnetic radiation.

2.5 Spectral Data Analysis

Calibration function development involved several step-by-step procedure including mathematical treatments on raw spectral signature (pre-processing), data partitioning into calibration and validation subsets and regression modeling as briefly discussed below.

2.5.1 Data Pre-processing

Spectral reflectance (R) or absorbance (A) (Equation 2.1) of soil can be regarded as an integration of information about its constituents (absorptions) and scattering of electromagnetic radiation at the irregular soil surface. The scattering component may result in extraneous spectral variations; non-linearity and baseline shifts (Rinnan *et al.*, 2009) and hence its negative effects on the reflectance signal has to be substantially diminished. Spectral pre-processing techniques are usually employed to serve the purpose with a view to improve the performance of calibration functions (Barnes *et al.*, 1989).

$$A(R) = \ln\left(\frac{1}{R}\right) \qquad [2.1]$$

Rinnan *et al.* (2009) has classified all the pre-processing methods under two broad categories; derivatives and scatter correction methods. The former category comprised of first derivative (*FD*) (Equation 2.2) and second derivative (*SD*) (Equation 2.3) techniques while standard normal variate (*SNV*), de-trending (*DT*) and multiplicative scatter correction (*MSC*) constituted the latter category. The derivative methods remove background effects on the spectra and also enhance spectral features. The *SD* eliminates both baseline shift and linear trend in the spectra while *FD* accounts for only baseline shift.

$$FD(R) = \frac{R_{i+1} - R_i}{\lambda_{i+1} - \lambda_i} \qquad [2.2]$$

$$SD(R) = \frac{FD_{i+1} - FD_i}{0.5\left(\lambda_{i+2} - \lambda_i\right)} \qquad [2.3]$$

The *SNV* and *DT* (Barnes *et al.*, 1989) remove particle size and scattering effects and thereby address curvilinearity and baseline shift issues in the spectra (Buddenbaum and Steffens, 2012). The *SNV* (Equation 2.4) computes the ratio of mean (μ_R) centered reflectance to its standard deviation (σ_R) while *DT* fits a second order polynomial to the spectrum transformed by *SNV* and the difference (due to scattering) at each wavelength is corrected.

$$SNV(R) = \frac{R - \mu_R}{\sigma_R} \qquad [2.4]$$

The *MSC* (Martens *et al.*, 1983) accounts for baseline shift in the reflectance spectra. Least square method (Equation 2.5) is used to fit each spectrum (R) and a reference spectrum (R_{ref}) and *MSC* is computed using scattering (a), offset (b) and soil constituent (e) information as given in Equation 2.6. Usually, the average of all spectrum is preferred for R_{ref}.

$$R = a + bR_{ref} + e \qquad [2.5]$$

$$MSC = \frac{R - a}{b} \quad [2.6]$$

2.5.2 Data Partition

One basic requirement for reliable judgement on the performance calibration functions is that it has to be trained (calibration) and tested (validation) in similar datasets. Otherwise, it may result it either over-fitting or over-fitting of calibration functions depending on the distribution of samples and mislead evaluation of calibration functions (Rajer-Kanduč *et al.,* 2003). In this regard, data partitioning/subsetting methods are critical in DRS studies.

Several methods are available for data partitioning among which random selection is commonly used (Morellos *et al.,* 2016; Islam *et al.,* 2006; Ludwig *et al.,* 2002). The approach is simple and independent of both spectra and attribute values. In addition, both the samples selected and whole dataset is expected to have similar distribution. But the approach is incapable to account the extrapolation issue (Rajer-Kanduč *et al.,* 2003) and do not always ensure similarity between the subsets although efforts were taken to address the same (Vasques *et al.,* 2009a). In other approach referred as 'sorting algorithm' make use of attribute values for subset selection. In this approach, attribute values are sorted and samples at pre-defined intervals are chosen for validation while the remaining samples constitute the calibration subset (Sarathjith *et al.,* 2016a; Viscarra Rossel and Lark, 2009; Martin *et al.,* 2002). This approach demand pre-hand information of soil attribute values and not suitable if there are small number of samples in the dataset with extreme attribute values. There are other methods that rely only on spectral data. For instance, the method implemented by Chang *et al.* (2001) examine the similarity of *FD* spectra in terms of Euclidean distance for subletting. In contrast, the Kennard–Stone method (Kennard and Stone, 1969) examine the spectral dissimilarity between samples. Data partitioning method based on both predictor and response variables together are also in practice; sample set partitioning based on joint *x–y*

distances proposed by Galvão *et al.* (2005) is an example. A comparison between different data partitioning algorithm has been rarely attempted in soil DRS studies and hence warrant further analysis.

2.5.3 Data Modelling Algorithms

With the advent of hyperspectral sensors, spectral signature of target can be generated at a very high spectral resolution as in case of DRS. Although such signature provide detailed spectral information about the target, it result in the 'curse of dimensionality'. Moreover, hyperspectral data are generally high redundant, multi-collinear and often subject to Hughes phenomenon. All these factors collectively influence data handling and modelling. Hence, the selection of an appropriate algorithm has gained much relevance in hyperspectral data modelling. Usually, the performance of such algorithms are expressed in terms of coefficient of determination (R^2) and root mean squared error (RMSE) of the observed and model predicted values. Different algorithms were evaluated to account for these challenges, among them the most widely accepted and commonly used approaches include the use of a) spectral indices, b) feature projection and c) feature selection.

2.5.3.1 Spectral Indices

In general, spectral index represent the combination of spectral reflectance values at two or more wavelengths although in some cases individual wavelengths are also used as spectral index (Weidong *et al.*, 2002). It helps to enhance spectral feature (relative to selected wavelengths) relevant to target attribute. The spectral index based calibration functions are build using it as predictor in simple linear or nonlinear regression with target attribute. The approach is computationally fast and simple.

Spectral indices typically used in reflectance studies of soil are relative reflectance, ratio index, derivative index, difference index and normalized

difference index. Relative reflectance index is the ratio of spectral reflectance of wet soil to that in its dry state (Weidong *et al.,* 2002). The ratio index represents the ratio reflectance values at two discrete wavelengths. The water index SOIL (WISOIL) based on wavelengths 1300 and 1450 nm (Fabre *et al.,* 2015; Bryant *et al.,* 2003) is an example of ratio index. The derivative index approximate the finite difference between two successive wavelengths instead of arithmetic difference in case of difference index (Weidong *et al.,* 2003). The normalized difference index is the ratio of difference to the sum of reflectance values at two different wavelengths. The normalized soil moisture index (NSMI) based on 1800 and 2119 nm (Haubrock *et al.,* 2008) is an example of this type. Weidong *et al.* (2003) compared the performance of relative reflectance, derivative and difference approaches and noted that first derivative absorbance index (R^2=0.88; RMSE=0.064) outperform others (R^2=0.63–0.88; RMSE=0.064–0.083) in the estimation of soil moisture content. Recently, Fabre *et al.* (2015) compared WISOIL, NSMI and two other new normalized difference indices developed in their study, namely normalized index of near and shortwave infrared domain for soil moisture content estimation from linear (NINSOL) and non-linear regression (NINSON). The NINSOL use 2076 and 2230 nm wavelengths while NINSON is based on wavelengths at 2122 and 2230 nm. They noted better results for NINSOL (R^2=0.87; RMSE=4.4) than others R^2=0.74–0.85; RMSE=4.8–6.2). It may be noted that all the aforesaid indices were used in conjunction with spectra of wet soil to estimate soil moisture content. However, their suitability to assess soil moisture constants from spectra of dry soils has been rarely investigated.

2.5.3.2 Feature Projection

Feature projection (also referred as feature extraction) transform high dimensional spectral data to low dimensional space by an appropriate transformation (linear or non-linear). The most popular feature projection technique in DRS is principal component analysis (PCA). In PCA, the multi-collinear spectral variables are subjected to an orthogonal transformation into a set uncorrelated

principal components or scores. Singular value decomposition of spectral data or Eigen value decomposition of the covariance of spectral data may be used for orthogonal transformation. Those scores that describe the maximum variance in the data can be subsequently used as predictors in regression; the approach is commonly known as PCR. The PCR has been widely used in DRS studies of soil (Mouazen *et al.*, 2010; Chang *et al.*, 2001). The PLSR proposed by Wold *et al.* (2001) is another approach similar to PCR. In PCR, only spectra data (predictor) is used to build scores while PLSR consider both spectra (predictor) and attribute (response) data. The PLSR is an integration of dimension reduction and regression and the selection of successive scores depend on the maximum covariance between the spectra and attribute values (Viscarra Rossel and Behrens, 2010). Hence, PLSR is expected to give better results than PCR as noted in some studies (Mouazen *et al.*, 2010; Vasques *et al.*, 2008). In addition to the aforesaid prominent methods, other techniques such as regression tree (Brown *et al.*, 2006b), committee trees (Vasques *et al.*, 2009b), multivariate adaptive regression splines (Shepherd and Walsh, 2002), support vector machines (Sarathjith *et al.*, 2016b; Genot *et al.*, 2011) and artificial neural networks (Viscarra Rossel and Behrens, 2010; Daniel *et al.*, 2003) were also tested in DRS studies of soil.

Although, several methods exists for feature projection and regression analysis, the PLSR is the most widely and frequently used algorithm in practice due to its characteristics; fast computation, statistically efficient, execute variable selection automatically and enable classification (Boulesteix and Strimmer, 2007). However, the PLSR seem to be inferior to multivariate adaptive regression splines and artificial neural networks in the estimation of pH, organic carbon and clay content of soil (Viscarra Rossel and Behrens, 2010). Similar result was reported in the assessment of infiltration rate of water into soil (Goldshleger *et al.*, 2012). The reason for the inferior performance of PLSR in these studies may be due to its inability to characterize the non-linear relationship that existed between spectra and soil attribute. However, in practice, the difference in performance of PLSR and

other non-linear approaches can be endured with its inherent advantages mentioned above.

2.5.3.3 Feature Selection

The spectral signature in the operational range of DRS is usually weak and complex due to the overlapping of overtones and combination of fundamental absorption bands mainly associated with N–H, O–H and C–H functional groups. This may impose redundant and irrelevant information in the spectra and thereby confound the relevant information related to any attribute under concern. Thus, the use of such spectra would result in calibration functions of inferior performance. In contrast, selection of appropriate number of useful wavelengths is capable to yield simple, accurate and robust calibration functions (Xiaobo *et al.*, 2010; Nadler and Coifman, 2005). In this context, feature selection (also referred as variable selection) has gained significance in DRS.

Several methods of feature selection has been discussed in DRS literature. They include non-linear methods such as simulated annealing (Kirkpatrick *et al.*, 1983), successive projections algorithm (Araújo *et al.*, 2001), wavelet transformation (Ge and Thomasson, 2006), genetic algorithm (Leardi *et al.*, 1992), PLSR based methods, among others. The PLSR based variable selection methods are superior to others in terms of computational efficiency and simplicity. The PLSR methods are broadly categorized as filter, wrapper and embedded methods (Mehmood *et al.*, 2012). The filter methods simply select variables based on the rank assigned to output vectors of PLSR. The wrapper methods perform model fitting and feature selection at every iteration. Embedded methods constitute those methods incorporated to PLSR algorithm. As both the wrapper and embedded methods are complex and computationally slow, filter methods may be the most suitable option for feature selection. Nonetheless, their reliability depends on desired threshold value for ranking variables. To avoid use of the threshold value, Li *et al.* (2009) has proposed CARS approach which involved Monte Carlo scheme and random selection of variables. Due to the randomized selection, the approach

fail to give unique result. This shortcoming can be addressed by OPS approach (Teófilo *et al.*, 2009) which employed variable indicators as a measure of ranking variables. Recently, Sarathjith *et al.* (2016) successfully implemented the OPS approach in soil datasets by incorporating an exponential decreasing function to select the number of variables. This approach was found to be simple and enable parsimonious variable selection. Table 2.1 lists the regression statistics of feature projection (using PLSR) and selection (OPS or CARS) methods for selected soil attributes commonly estimated via DRS. This table may be used to appraise the consistent improvement in the performance of feature selection compared to projection approach.

Table 2.1 Regression statistics in the validation of feature projection and selection methods in the estimation of selected soil attributes

Attribute	Feature projection (PLSR)		Feature selection (OPS or CARS)		Reference
	R^2	RMSE	R^2	RMSE	
Organic carbon	0.74	0.33	*0.60*[*]	*0.41*	[1]
	0.57	0.24	0.61	0.23	[2]
Total carbon	0.51	0.77	0.64	1.72	[3]
	0.79	0.43	0.78	0.45	[4]
pH	0.78	0.21	0.81	0.19	[2]
	0.72	0.46	*0.81*	*0.37*	[1]
	0.66	0.57	0.66	0.57	[4]
Clay content	0.01	9.81	0.05	10.82	[3]
	0.47	0.22	0.48	0.22	[2]
	0.77	7.83	0.77	7.80	[4]
Sand content	0.06	6.97	0.10	6.40	[3]
	0.55	5.61	0.58	5.45	[2]
	0.69	9.11	0.69	9.40	[4]

[*]Values in italics represents CARS approach; R^2: coefficient of determination; RMSE: root mean square error; PLSR: partial least square regression; OPS: ordered predictor selection; CARS: competitive adaptive reweighted sampling; [1] Vohland *et al.* (2014); [2] Sarathjith *et al.* (2016a); [3] Raj *et al.* (2018); [4] Wartini *et al.* (2019).

2.5.4 Evaluation of Calibration Functions

The performance of DRS based calibration functions are usually expressed in terms of R^2 (Equation 2.7) and RMSE (Equation 2.8) between observed (Y) and predicted ($\hat{Y}$) values. The R^2 describes the variance of observed values explained by the calibration function. It has no unit and range between 0 and 1. The RMSE is a measure of error and hence bear same unit as the attribute values. Both these statistics are reliant on the attribute range (Bellon-Maurel and McBratney, 2011).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \qquad [2.7]$$

$$RMSE = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2} \qquad [2.8]$$

where, $\bar{Y}$ indicates the average of Y and n represents the number of soil samples. Another statistic namely, residual prediction deviation (RPD) is commonly used in DRS to account for the range effect on RMSE. It is defined as the ratio of standard deviation of observed values in validation to the respective RMSE (Equation 2.9).

$$RPD = \frac{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}{\frac{1}{n}\sqrt{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}} \qquad [2.9]$$

It may be noted that no fixed standard is followed in DRS literature to appraise the performance of calibration functions. Some of the accuracy classification scheme typically used in DRS of soil are listed in Table 2.2, among which that suggested by Chang *et al.* (2001) has been most frequently used in soil DRS studies.

Table 2.2 Selected criteria to evaluate the accuracy of calibration functions in diffuse reflectance spectroscopy of soil

Reference	Criteria				Performance of calibration function
	R^2	RPD	RER	SEC/SD	
Chang *et al.* (2001)	< 0.50	< 1.40	-	-	Poor
	0.50 – 0.80	1.40 – 2.00	-	-	Moderate
	0.80 – 1.00	> 2.00	-	-	Accurate
Dunn *et al.* (2002)	-	< 1.60	-	-	Poor
	-	1.60 – 2.00	-	-	Acceptable
	-	> 2.00	-	-	Excellent
Coûteaux *et al.* (2003)	-	-	-	0.20 – 0.50	Caution
	-	-	-	≤ 0.20	Good
Malley *et al.* (2004)	0.70 – 0.80	1.75 – 2.25	8.00 – 10.00	-	Moderately useful
	0.80 – 0.90	2.25 – 3.00	10.00 – 15.00	-	Moderately successful
	0.90 – 0.95	3.00 – 4.00	15.00 – 20.00	-	Successful
	> 0.95	> 4.00	> 20.00	-	Excellent
Saeys *et al.* (2005)	-	< 1.50	-	-	Not usable
	-	1.50 – 2.00	-	-	Distinguish high and low values
	-	2.00 – 2.50	-	-	Approximate predictions
	-	2.50 – 3.00	-	-	Good
	-	> 3.00	-	-	Excellent
Viscarra Rossel *et al.* (2006a)	-	< 1.00	-	-	Very poor
	-	1.00 – 1.4	-	-	Poor
	-	1.4 – 1.8	-	-	Fair
	-	1.8 – 2	-	-	Good
	-	2 – 2.5	-	-	Very good
	-	> 2.5	-	-	Excellent

R^2: coefficient of determination; RPD: residual prediction deviation; RER: ratio error range; SEC: standard error of calibration; SD: standard deviation

2.5.5 Estimation of Soil Moisture Constants using Diffuse Reflectance Spectroscopy

Although numerous studies have been conducted to estimate basic soil properties and nutrient contents via DRS, very limited studies investigated its ability to assess soil moisture constants. Recently, Viscarra Rossel and Webster (2012) examined the DRS performance in the estimation of FC and WP (among other selected soil properties) using an Australian database consisting of about 21,500 archived soil samples. They used a model tree-based approach to separate data into homogeneous groups. An ordinary least-squares regression was implemented to relate soil attribute and spectral signature within each group. They observed a moderate level of prediction for both FC (RPD=1.68) and WP (RPD=1.95). Similar result was also observed by both Kinoshita *et al.* (2012) and Babaeian *et al.* (2015). The former study employed PLSR to assess FC (RPD=1.81) and WP (RPD=1.97) of Ultisols in western Kenya. Babaeian *et al.* (2015) made use of spectra transfer functions to estimate soil moisture content at different matric potentials varying from 0 to -15000 cm. The implemented approach was found to yield relatively better results at low and intermediate soil moisture contents (R^2>0.50; RMSE< 0.018 cm^3 cm^{-3}) compared to that towards saturation. The R^2 noted in their study for the estimation of FC (soil moisture content at -330 cm) and WP (soil moisture content at -15000 cm) was 0.52 and 0.63, respectively. Instead of focusing on specific soil moisture constants, some studies linked spectral reflectance and parameters of water retention functions (Pittaki-Chrysodonta *et al.*, 2018; Santra *et al.*, 2009). Apart from these studies, very limited attempts were made to relate soil moisture constants directly with the spectral reflectance of dry soil and thus warrant further investigation.

MATERIALS AND METHODS

CHAPTER 3

MATERIALS AND METHODS

3.1 Soil Database

This study was conducted using an available soil database (*N*=302) consisted of values of moisture content at -0.33 bar (FC) and -15 bar (WP) matric suction measured using pressure plate apparatus, texture (Gee and Bauder, 1986), organic matter (Walkley and Black, 1934), spectral absorbance (12489–3594 cm^{-1} or 801–2782 nm), among others. It may be noted that the texture and organic matter values were not available for 11 and 17 number of samples, respectively while the values of soil moisture indices and spectra were available for all the soil samples. The spectral data acquisition of air dried soil samples (< 2 mm) taken in glass petri dishes were performed using a Fourier transform based multipurpose analyzer (MPA) (Bruker Optik GmbH, Germany) at a resolution of 16 cm^{-1} using 'integrating sphere' as the measurement channel. The OPUS software associated with the instrument was used for spectra acquisition. For each sample, spectral measurements were made at each quadrant of the petri dish and thus yielded four replicated spectra. Each spectrum was set to be an average of 64 internal spectra. The average of replicated spectra was used for the subsequent analyses after transforming into reflectance units. Several preliminary analyses identified that the spectral data in 1100–2500 nm (Vendrame *et al.*, 2012) yielded better results and hence this wavelength range was used for all the analyses in this study.

3.2 Data Analysis

Data analysis in this study comprises of outlier removal, data partitioning, data preprocessing, regression modeling and variable selection approaches implemented to develop calibration functions which link soil spectra and the attributes of interest. All the analyses were performed using MATLAB (R2017a, Mathworks) software.

The calibration function development scheme implemented in the study is depicted in Fig. 3.1. Initially, the outliers in the soil database were identified and removed. Let us consider (*X,Y)* and (*x,y*) as the spectral and attribute data in ordered pair of all samples before and after outlier removal, respectively. After outlier removal, the dataset was divided into calibration (x_c, y_c) and validation (x_v, y_v) subsets for which the spectra and attribute data were represented using suffixes *c* and *v,* respectively. The calibration functions linking soil spectra and attribute values were developed by a threefold approach as defined by the objectives of the study. In the first approach, indices generated from the spectral data (x_c^{si}, x_v^{si}) were linked with the attribute of interest by simple linear regression while the second and third approaches used the full-spectrum (x_c^{fs}, x_v^{fs}) and selected spectral variables (x_c^{vs}, x_v^{vs}) to establish the linkage using PLSR. A detailed description of the data analysis procedure is given below.

3.2.1 Outlier Removal

Outlier detection and removal was based on the residuals (difference between observed and predicted values) of the PCR relationship between soil spectra (predictor variable) and attribute of interest (response variable). In this approach, principal components of soil spectra were initially estimated and later related with the attribute by means of multiple linear regression. Then, the residuals with error bars (corresponding to 95% confidence interval) of each observation were plotted using *rcoplot* function in MATLAB. The observation for which the error bar not intersecting the zero residual line was regarded as an outlier and subsequently removed from the dataset (Sarathjith *et al.*, 2014).

3.2.2 Data Partitioning

The dataset was divided into calibration and validation subsets in the ratio 3:1 using 'sorting' algorithm (Sarathjith *et al.*, 2016a; Viscarra Rossel and Lark, 2009) to train and test regression models, respectively. In this approach, soil samples were set in ascending order of attribute values and every third sample starting from

second were used for validation and all the remaining samples constituted the calibration subset. A two sample Student's *t*-test for equal means and Levene's *F*-test for equal variances were performed at 5% level of significance to examine the similarity in the distribution of attribute values between calibration and validation subsets. The descriptive statistics in terms of measures of central tendency and measures of dispersion of soil attributes were computed separately for calibration and validation subsets. The frequency distribution and the quartiles of soil attributes were examined graphically by means of histograms and box-whisker plots.

3.2.3 Development of Spectral Indices based Calibration Functions

Spectral indices combine spectral reflectance values at two or more wavelengths to enhance spectral feature related to the attribute of interest. While there exists numerous spectral indices, this study examined the most commonly used index namely, normalized difference reflectance index (NDRI) obtained as the ratio of difference in reflectance values at two different wavelengths to the value obtained by their addition (Equation 3.1). As the spectra generated using MPA has uneven number of data points within a specific wavelength interval and also many of them were not at integer wavelengths. Thus, the spectra were subjected to Piecewise Cubic Hermite Interpolation (Alamar *et al.*, 2007) to 1 nm sampling interval at discrete wavelengths in the desired wavelength range prior to NDRI calculation. The NDRI values were generated all the pairwise combination (Haubrock *et al.*, 2008) of reflectance at two interpolated wavelengths (R_i and R_j) in 1100–2500 nm range.

$$NDRI(i, j) = \frac{R_i - R_j}{R_i + R_j} \qquad [3.1]$$

Calibration functions relating the generated spectral index values and the soil attributes were developed using simple linear regression. They were trained using the calibration subset and tested on validation subset. The performance of the developed calibration functions were evaluated in terms of R^2, RMSE and RPD.

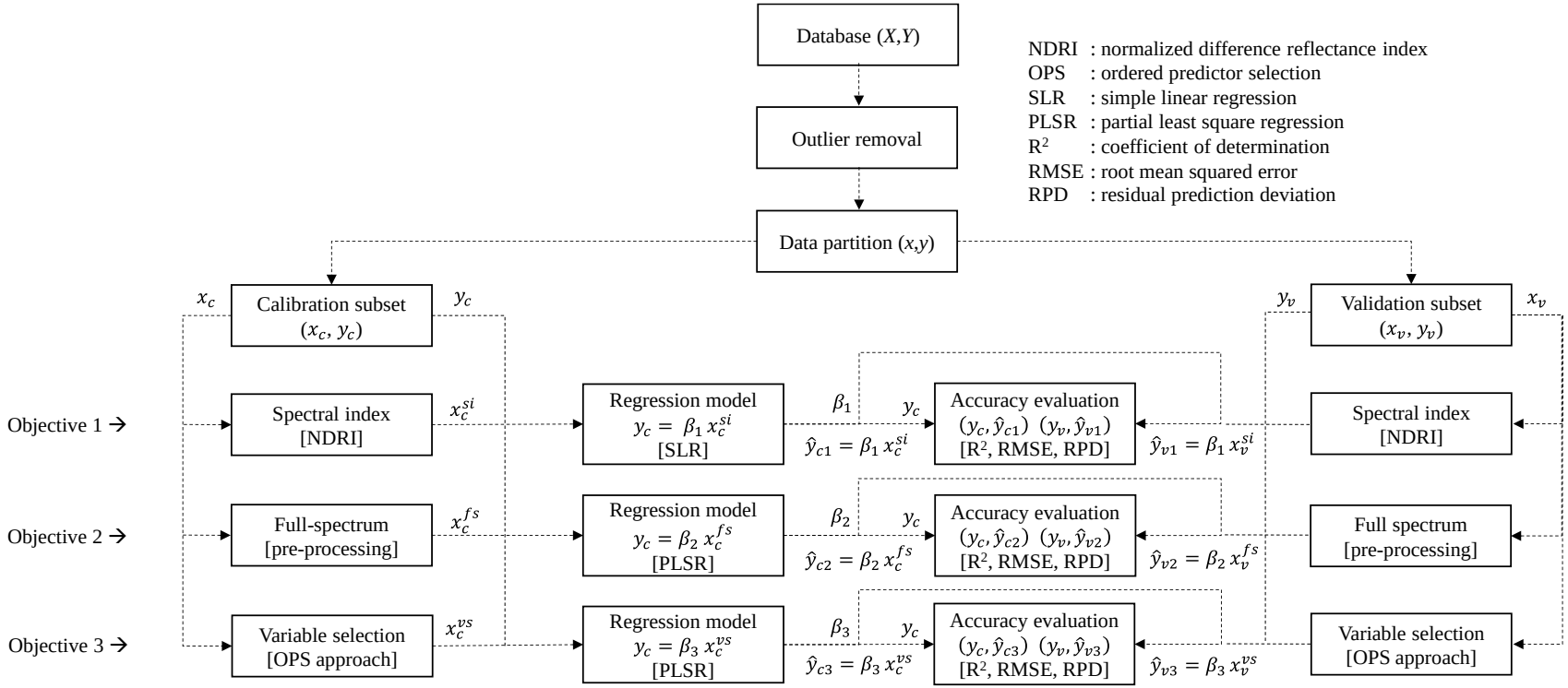

Fig. 3.1 Calibration function development scheme

3.2.4 Development of Full-spectrum based Calibration Functions

Spectral pre-processing aimed to remove physical phenomena (scattering effects and other undesired variations) in the spectra (Rinnan *et al.*, 2009) and hence regarded as an essential pre-requisite for the development of calibration functions. In this study, different spectral pre-processing techniques generally used in spectroscopic studies were evaluated which included, scatter correction methods *viz*. *SNV* (Equation 2.6), *MSC* (Equation 2.7), *DT* (Equation 2.8) and derivatives namely, *FD* (Equation 2.9) and *SD* (Equation 2.10). Other spectral transformations used in this study included the pairwise combinations of scatter correction methods with derivatives. Both *R* and *A* spectra were subjected to the aforesaid transformations. In addition, the untransformed reflectance and absorbance were also included in the analysis. Thus, a total of 24 spectra transformations (12 based on reflectance + 12 based on absorbance) were examined in the study.

Calibration functions that link pre-processed spectra (1100–2500 nm) with the soil attribute (FC or WP) were established using the calibration subset via PLSR algorithm (Wold *et al.*, 2001) and their performance was evaluated with the validation subset. In PLSR, both the predictor (spectra) and response variables (soil attribute) were considered to build uncorrelated latent variables (*LV*) from possibly correlated spectral variables by orthogonal transformation. The *LV* (also known as scores, factors or components) were then used as predictor variables in regression. However, the use of more or less number of *LV* than a defined optimum would either over-fit or under-fit PLSR model. To avoid such behaviors, the optimum number of *LV* was selected based on leave-one-out cross-validation approach (Viscarra Rossel 2007) in this study. In this strategy, PLSR models were built for specific number of *LV* using calibration dataset with one sample left for testing. The procedure continued until each sample in the calibration dataset became a test sample exactly once. At each iteration, the square of difference between observed and predicted value of test sample (squared-error) was recorded. The squared-error when divided by the number of iterations yielded the mean-squared-error (MSE)

corresponding to number of *LV* used. Similarly, the MSE for different number of *LV* varying from 1 to 20 were recorded. Then, minimum MSE criterion was followed to select optimum number of *LV* to develop PLSR based calibration functions.

The regression statistics, namely R^2 (Equation 2.10) and RMSE (Equation 2.11) in the calibration and validation of PLSR models were computed. In addition, RPD (Equation 2.13) and Akaike Information Criterion (*AIC*) (Viscarra Rossel, 2008; Akaike, 1973) of validation were examined. Among the PLSR based calibration functions developed using different spectra pre-processing techniques, the best one was chosen based on minimum value of *AIC* of validation (Equation 3.2).

$$AIC = n \times \ln(RMSE) + 2 \times LV \qquad [3.2]$$

where, *n* indicates the sample size while *LV* and *RMSE* represents the number of latent variables and root mean squared error in the validation of a PLSR model, respectively. The main reason to use *AIC* as a selection criterion lies on its ability to consider together both the accuracy (in terms of *RMSE*) and complexity (in terms of *LV*) of calibration functions.

Identification of most relevant functional groups and their absorption modes are important to elucidate the prediction performance of calibration functions. For the purpose, initially, the most significant wavelengths were identified based on a variable indicator namely, *β-VIP*; product of absolute values of regression coefficient (*β*) (Vasques *et al.*, 2009a) and variable importance for projection (*VIP*) (Viscarra Rossel 2008) after normalization. The *β-VIP* was also used by Sarathjith *et al.* (2014a) for feature identification. The higher and lower values of *β-VIP* represents the most significant and least significant wavelengths in prediction, respectively. Then, the most relevant wavelengths so identified were assigned to the spectrally active functional groups and their absorption modes as compiled from Bayer *et al.* (2012), Ben-Dor *et al.* (1997) and Viscarra Rossel and Behrens (2010).

3.2.5 Development of Variable Selection based Calibration Functions

The OPS approach suggested by Sarathjith *et al.* (2016) was implemented in this study. Variable indicators being the descriptors of predictor-response relationship form the basis of OPS approach. The approach involved the arrangement of variable indicator values in the decreasing order of their absolute magnitude after normalization. Further, wavelengths with low absolute magnitude were forcefully excluded iteratively using an EDF as computed in Equation 3.3 in which $i = 1,2,3,\ldots,m$ denotes the iteration number, m and p represents an integer set to 50 (Li *et al.*, 2009) and total number of spectral variables (NSV), respectively. The EDF strategy facilitated a rapid selection of spectral variables during initial iterations and very refined selection thereafter (Li *et al.*, 2009).

$$r_i = a \times \exp(-k \times (i+1)) \qquad [3.3]$$

$$k = \frac{\ln(0.5 \times p)}{m-1} \qquad [3.4]$$

$$a = (0.5 \times p)^{1/(m-1)} \qquad [3.5]$$

At each iteration a calibration function was generated (hereinafter referred as 'subset model') with NSV equal to the product of r_i and p (Li *et al.*, 2009). This resulted in m number of subset models with different NSV. The PLSR was used as the regression algorithm to relate subset of spectral variables and soil attribute. The optimum number of latent variables to be used in subset models was selected based on the leave-one-out cross-validation approach similar to that mentioned in the Section 3.2.4. The number of latent variables corresponding to the first local minimum of MSE values was regarded as the optimum. For each attribute, the maximum number of latent variables to be used in the cross-validation was limited to its optimum number found for its full-spectrum counterpart (Vohland *et al.*, 2014). The same calibration and validation samples identified as described in the Section 3.2.2 were used for the evaluation of subset models. The regression statistics of subset models were computed. The selection of the optimum subset

model was based on the minimum RMSE of validation and the corresponding NSV was chosen as the optimum. For each soil attribute, the aforesaid procedure was executed using different variable indicators which may be classified as PLSR-dependent and PLSR-independent categories. The PLSR-dependent variable indicators comprised of β (Equation 3.9), *VIP* (Equation 3.10) and squared residual (*SqRes)* (Equation 3.11) while, Pearson correlation coefficient (*r*) (Equation 3.12), biweightmidcorrelation vector (*bicor*) (Equation 3.13), adjacency values of mutual information (*AMI*) (Equation 3.20), signal-to-noise vector (*StN*) (Equation 3.21) and covariance procedures vector (*CovProc*) (Equation 3.22) constituted the PLSR-independent category. All the variable indicators were normalized and their absolute magnitude values were used. In general, the absolute magnitude value of variable indicator (except *SqRes*) at each wavelength describe the significance of linkage between respective spectral feature and soil attribute. In case of *SqRes*, the elemental values with low absolute magnitude are more important (Teófilo *et al.*, 2009) to describe spectra-attribute linkage. Hence, the reciprocal of *SqRes* (hereinafter referred as *SqRes*) was used as a variable indicator (Sarathjith *et al.*, 2016a) in this study. As different variable indicators can be combined (Teófilo *et al.*, 2009), pair-wise combinations (element wise product value) of aforesaid indicators were also treated as variable indicators. Thus, the utility of 36 variable indicators altogether (8 individual + 28 combinations) in conjunction with OPS approach were tested in this study.

For the computation of variable indicators, let X ($I \times J$) be the spectral data matrix (I observations and J spectral variables) and Y ($I \times 1$) be the soil attribute values. Initially, the PLSR model found the scores of X (T) as

$$T = XW \qquad [3.6]$$

where, W ($J \times K$) is the weight matrix with K number of factors (latent variables). Then, the matrix of T ($I \times K$) values were used as predictors to reconstruct X (Equation 3.7) and estimated Y (Equation 3.8) using loading matrices P ($J \times K$) and C ($1 \times K$), respectively.

$$X = TP^t + E = X^r + E \quad [3.7]$$

$$Y = TC^t + F \quad [3.8]$$

The X^r ($I \times J$) represents the reconstructed X matrix while E ($I \times J$) and F ($I \times 1$) indicates residuals and 't' denotes transpose of a matrix. Substituting T in Equation 3.8 yielded the standard form of multivariate regression model to estimate Y from X using $\beta = WC^t$ as the regression coefficient vector with $J \times 1$ dimension (Equation 3.9).

$$Y = XWC^t + F = X\beta + F \quad [3.9]$$

The VIP_j represents the importance of the j^{th} predictor variable ($j = 1,2,3,\ldots,J$) based on a model with k factors, w_{kj} denotes the loading weight of the j^{th} variable in the k^{th} PLSR factor, SSY_k and SSY_t represent the explained and total sum of squares of Y, respectively.

$$VIP_j(k) = J \times \sum_k w_{kj}^2 \left(\frac{SSY_k}{SSY_t} \right) \quad [3.10]$$

Let x_j and x^r_j represents j^{th} spectral variable ($j = 1, 2, 3,\ldots,J$) in X and X^r, respectively, x_{ij} denotes i^{th} element (i=1,2,3,…,I) in x_j while $\bar{x}_j$, x_j^m, x_j^{mad} and $\bar{Y}$, Y^m, Y^{mad} indicates the mean, median and median absolute deviations of x_j and Y, respectively. Then, *SqRes* was computed as,

$$SqRes_j = \frac{1}{(x_j - x_j^r)^t (x_j - x_j^r)} \quad [3.11]$$

The r indicates a linear measure of dependency between X and Y based \on their mean values (Equation 3.12). It vary in the range between -1 (strong

negative correlation) and +1 (strong positive correlation). The linear relationship between *X* and *Y* reduces as *r* tends to zero.

$$r(x_j,Y)=\frac{\sum_{i=1}^{I}(x_{ij}-\bar{x}_j)(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{I}(x_{ij}-\bar{x}_j)^2}\sqrt{\sum_{i=1}^{I}(Y_i-\bar{Y})^2}} \qquad [3.12]$$

The *bicor* is a correlation measure based on median. It was reported to be more robust to outliers than *r* (Wilcox 2005).

$$bicor(x_j,Y)=\frac{\sum_{i=1}^{I}(x_{ij}-x_j^m)w_i^x(Y_i-Y^m)w_i^Y}{\sqrt{\sum_{i=1}^{I}\left[(x_{ij}-x_j^m)w_j^x\right]^2}\sqrt{\sum_{i=1}^{I}\left[(Y_i-Y^m)w_i^Y\right]^2}} \qquad [3.13]$$

where, w_i^x symbolizes the weight (between 0 and 1) given for x_{ij} by assigning the indicator $L(1-|u_i|)$ a value of 1 if $L(1-|u_i|)>0$ or 0 otherwise. The weight decreases as x_{ij} moves away from the median and becomes 0 when it exceeds $9x_j^{mad}$ (Wilcox 2005).

$$w_i^x=(1-u_i^{\,2})L(1-|u_i|) \qquad [3.14]$$

$$u_i=\frac{x_{ij}-x_j^m}{9x_j^{mad}} \qquad [3.15]$$

The mutual information (*MI*) enumerate the information content that one random variable holds about the other and capable to account for the nonlinear relationships between variables (Battiti, 1994)

$$MI(x_j,Y)=H(x_j)+H(Y)-H(x_j,Y) \qquad [3.16]$$

$$H(x_j)=-\sum_{x\in X}p(x)\ln p(x) \qquad [3.17]$$

$$H(Y) = -\sum_{y \in Y} p(y) \ln p(y) \quad [3.18]$$

$$H(x_j, Y) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \ln p(x, y) \quad [3.19]$$

where, $H(x_j)$ and $H(Y)$ represent the marginal entropy of x_j and Y, respectively while $H(x_j,Y)$ indicates their joint entropy, $p(x)$ and $p(y)$ are the probability mass functions of x_j and Y, respectively, and $p(x,y)$ represents their joint distribution. The *MI*-toolbox of MATLAB developed by Hanchuan Peng (http://www.mathworks.in/matlabcentral/fileexchange/14888-mutual-information-computation, last accessed on May 18, 2019) was used for *MI* calculation.. Then, the *AMI* was computed as suggested by Song *et al.* (2012) as,

$$AMI(x_j, Y) = \frac{2MI(x_j, Y)}{H(x_j) + H(Y)} \quad [3.20]$$

The slope ($\hat{b}_j$) and residual (e_j) parameters of simple linear regression between Y and x_j was used to compute *StN*.

$$StN_j = \frac{\hat{b}_j}{e_j{}^t e_j} \quad [3.21]$$

The *CovProc* was computed as

$$CovProc = diag(X^t Y Y^t X) \quad [3.22]$$

3.2.5 Accuracy Evaluation of Calibration Functions

All the calibration functions generated in this study were evaluated in terms of R^2 (Equation 2.8), RMSE (Equation 2.9), and the RPD of validation (Equation 2.10). The accuracy criteria based on RPD of validation as suggested by Chang *et al.* (2001) was used in this study to classify the calibration functions into accurate (RPD > 2), moderate (1.4 < RPD < 2) and poor (RPD < 1.4) classes.

RESULTS AND DISCUSSION

CHAPTER 4

RESULTS AND DISCUSSION

The DRS is a promising tool for soil analysis and digital soil mapping. Several spectral libraries at varying spatial scales were already developed with a view to develop 'global' calibration functions of soil attributes. Almost all the spectral libraries are based on the spectral signature of dry soil and are satisfactory in establishing linkage with several soil attributes. However, limited studies have examined its utility to explain water retention behavior from dry spectral reflectance of soil and thus warrant further investigation. In this chapter, the results of three data modeling approaches and their comparison in the estimation of soil moisture constants (FC and WP) from dry spectral reflectance of about 302 soil samples are presented.

4.1. Exploratory Analysis of Soil Attributes

As part of exploratory analyses, the descriptive statistics, histograms, box plots and correlation structure of soil attributes were examined.

4.1.1 Descriptive Statistics, Histograms and Box Plots

The dataset consisted of soil samples covering all the 12 textural classes as per the United States Department of Agriculture (USDA) textural classification system (Fig. A1 of Appendix A). Among them, sandy loam was the dominant class contributing 35.1% of total soils (N = 291) in the dataset (with textural values available) followed by loam (19.2%), loamy sand (18.6%) and silty loam (15.8%) classes. The histograms and boxplots of soil texture, organic matter and pH of soil samples examined in this study are shown in Fig. A2 of Appendix A. The range of OM content was found to be 0.02–3.39% with an average value 0.43%. The pH range (4.25–8.03) revealed that the dataset consisted of soil samples with reaction classes varying between very strongly acidic and moderately alkaline as defined by the USDA soil pH classification scheme.

The range and average value of FC in the dataset was observed as 4.24–48.23% and 17.64%, respectively while their values were noted to be 1.12–24.65% and 7.52% in case of WP. Both the FC (skewness = 0.78; kurtosis = 3.12) and WP (skewness = 0.99; kurtosis = 3.11) appeared to have a skewed distribution as depicted in their histogram and boxplots (Fig. 4.1).

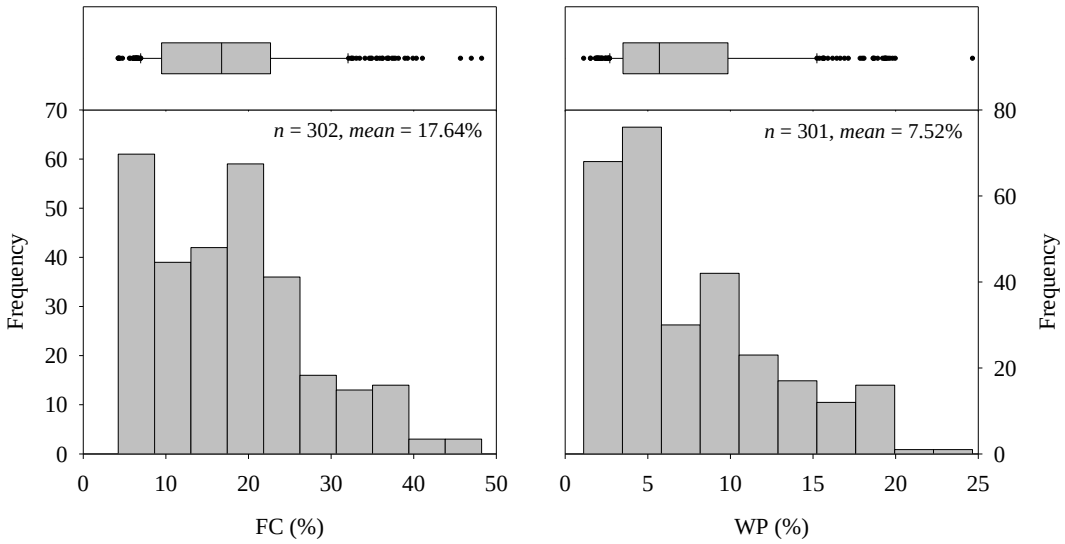

Fig. 4.1 Histogram and box plot of field capacity (FC) and wilting point (WP) of soil samples

4.1.2 Correlation Structure

Table 4.1 lists the Pearson correlation coefficients as a measure of linear dependency among soil attributes. The soil moisture constants (FC and WP) appeared to have excellent correlation among themselves. They have no significant correlation with OM and only FC was significantly correlated with pH. The FC and WP exhibited a strong correlation with sand (negative) and clay (positive) contents as indicative of the major influence of texture on water retention behavior of soils. The aforesaid correlation underlined the fact that the finer the texture, the higher is the soil moisture constant and vice versa (Kirkham 2014).

Table 4.1 Pearson correlation coefficient between soil attributes

Attribute	FC	WP	pH	OM	Sand	Clay
FC	1.00					
WP	0.89**	1.00				
pH	0.24**	0.16	1.00			
OM	0.12	0.06	-0.08	1.00		
Sand	-0.35**	-0.21*	-0.29**	0.02	1.00	
Clay	0.53**	0.43**	0.12	-0.03	-0.59**	1.00

*Significant at $p < 0.001$
**Significant at $p < 0.0001$

4.2 Soil Spectral Signature

The spectral reflectance of soil (Fig. 4.2) consisted of three prominent absorption peaks centered on 1400, 1900 and 2200 nm which are linked with clay minerals (Wetterlind and Stenberg 2010). The spectral features around 1400 and 1900 nm can be assigned to the hydroxyl (O–H) group associated with water while metal-hydroxyl stretching characterize the absorption around 2200 nm (Bricklemyer and Brown, 2010; Chang and Laird, 2002; He *et al.*, 2005; Post and Noble, 1993).

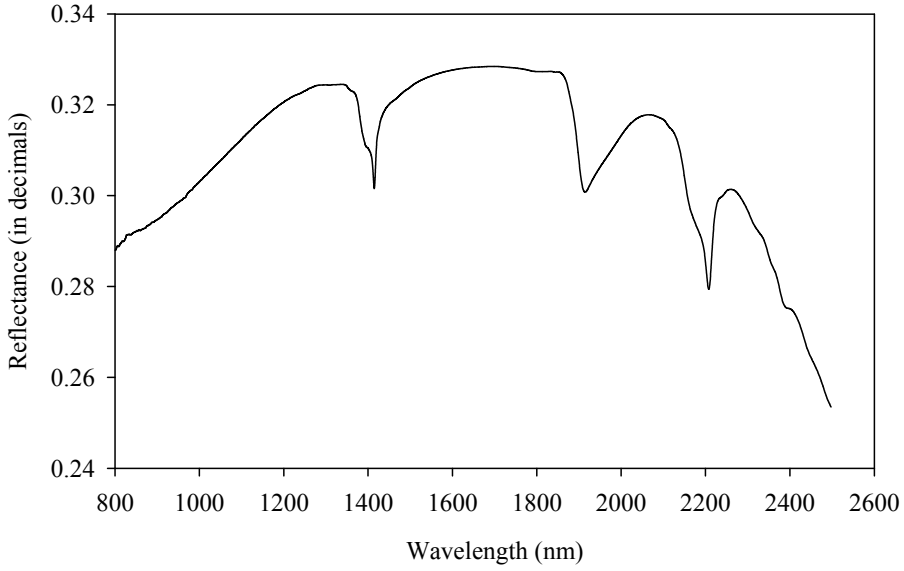

Fig. 4.2 Average spectral reflectance of soil

Figure 4.3 shows the number of data points in the spectrum for every 100 nm interval starting from 800 nm for the entire operational range of the instrument (MPA). The spectra generated by MPA has more number of data points at shorter wavelengths and gradually decrease towards longer wavelengths. The inclusion of larger number of data points with very fine sampling interval in shorter wavelengths appeared to incur poor calibration performance during preliminary data analysis. Hence, spectral data points in 800–1099 nm wavelength range were not considered for data analysis and modelling involved in this study.

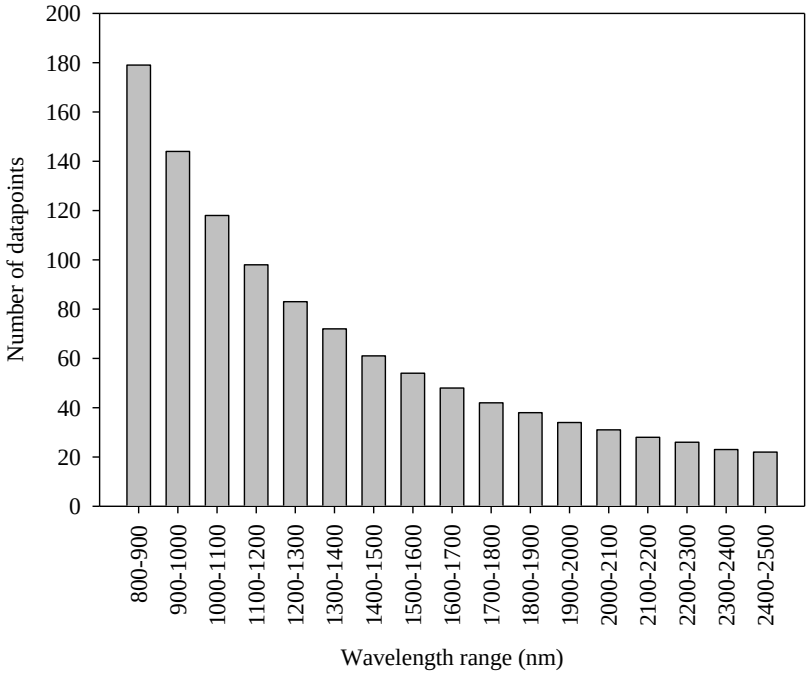

Fig. 4.3 Number of wavelengths in soil reflectance spectrum

4.3 Data Analysis

4.3.1 Outlier Detection and Removal

For a more reliable data analysis and calibration function development, the dataset was inspected for outliers using a combination of PCR and *rcoplot*. This method considered both the spectral and attribute information for outlier detection. In this, the principal components of the soil spectra were computed and the selected

ones (based on their explained variance) were used as predictor variables in a multiple linear regression framework to estimate specific soil attribute values. Then, the residuals obtained using PCR approach were examined with their error bars (corresponding to 95% confidence interval) by means of *rcoplot* for the detection and removal of outliers (Fig. 4.4). Accordingly, number of outliers detected in the analysis of FC and WP were 9 and 11, respectively. The outlier samples so detected for each attribute were excluded from subsequent analyses.

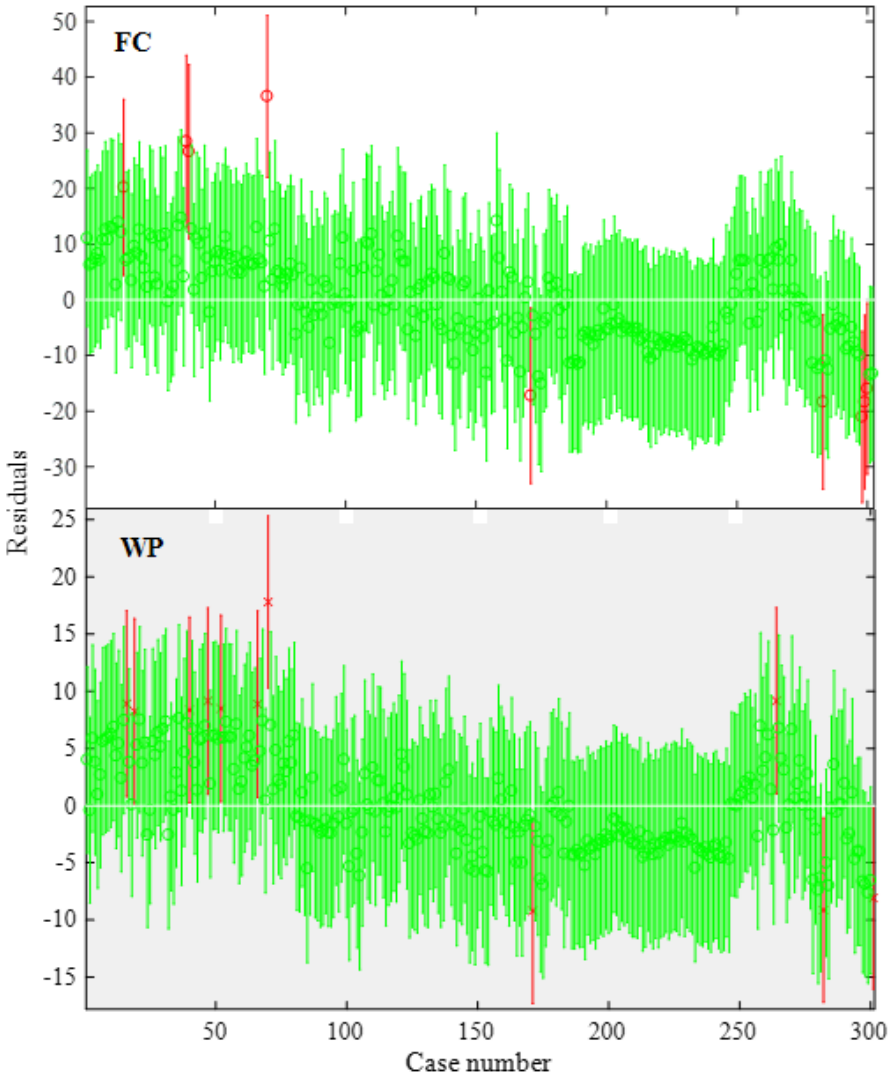

Fig. 4.4 Residual case order plot of residuals of field capacity (FC) and wilting point (WP) for detection of outliers (cross markers)

4.3.2 Data Partitioning

The descriptive statistics of FC and WP in the calibration and validation subsets were examined separately (Table 4.2) as a check for the ability of data partition algorithm used to yield similar subsets. It was noted that the data partition algorithm was capable to ensure the range of attribute values of validation subset within that of calibration. Moreover, the mean and coefficient of variation values were found to be similar across calibration and validation subsets. The similarity of the subsets was further endorsed by the results of two sample Student's *t*-test for equal means and Levene's *F*-test for equal variances at 5% significance level for FC ($H_0 = 0$; $p = 0.87$) and WP ($H_0 = 0$; $p = 0.75$).

Table 4.2 Descriptive statistics of soil attributes in calibration and validation subsets

Attribute	Calibration				Validation			
	n	**Range**	**Mean**	**CV**	***n***	**Range**	**Mean**	**CV**
FC (%)	220	4.24 – 48.23	17.55	54.31	73	4.24 – 45.67	17.47	54.66
WP (%)	218	1.12 – 24.65	7.48	67.62	72	1.53 – 19.98	7.49	67.60

n: number of samples
CV: coefficient of variation in percentage

4.3.3 Development of Calibration Functions

The study has examined the utility of three different approaches typically used in spectral data modeling of soil. The first approach used normalized difference spectral indices generated from reflectance spectra as predictor variable in simple linear regression with soil attributes. In the second approach, the full-spectrum was used in conjunction with PLSR to develop the desired calibration functions. The third approach evaluated the potential of OPS method of spectral variable selection on the performance of PLSR model. The results of these implemented approaches are discussed below.

4.3.3.1 Spectral Index based Calibration Functions

In this approach, NDRI values were generated for all pairwise combinations of wavelengths and performed simple linear regression to establish linkage with soil attribute of interest. The map of R^2 values in the calibration (upper triangle) and validation of simple linear regression of NDRI with FC and WP are shown in Fig. 4.5 and Fig. 4.6, respectively. It may be noted that the map appear to be visually symmetric with respect to the diagonal. The reason for the similarity may be attributed to similar performance of simple linear regression model in both calibration and validation.

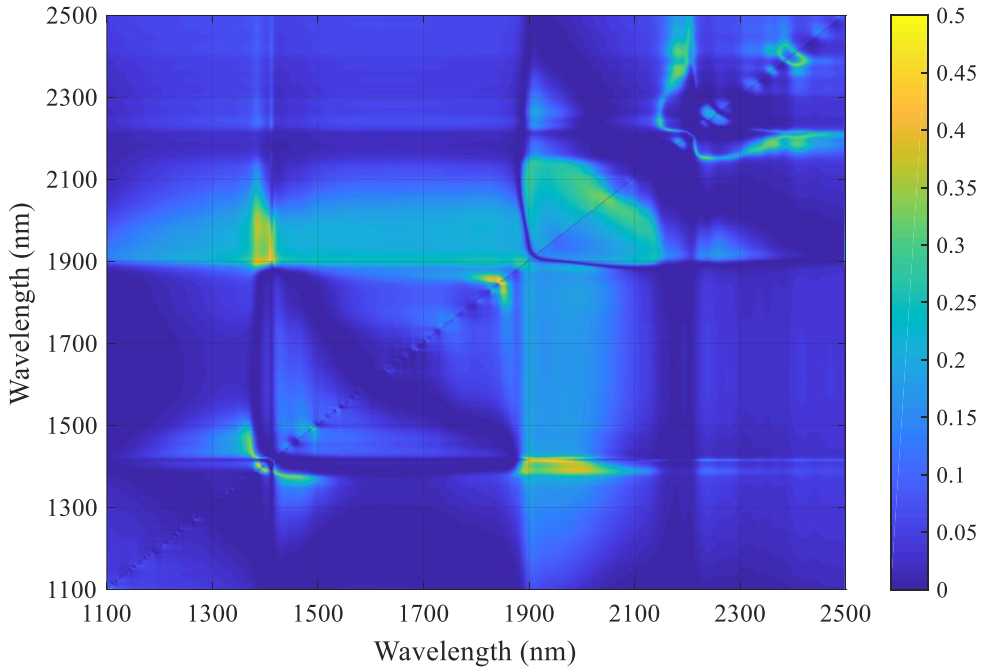

Fig. 4.5 Coefficient of determination of calibration (upper triangle) and validation (lower triangle) of simple linear regression between normalized difference reflectance indices and field capacity

In both FC and WP cases, the NDRI values generated using combination of wavelengths near or around the water absorption features (1400 and 1900 nm) were found to yield relatively better performance (R^2>0.45) values in both calibration and validation. In addition, the combination of wavelengths in 1900–2100 nm range

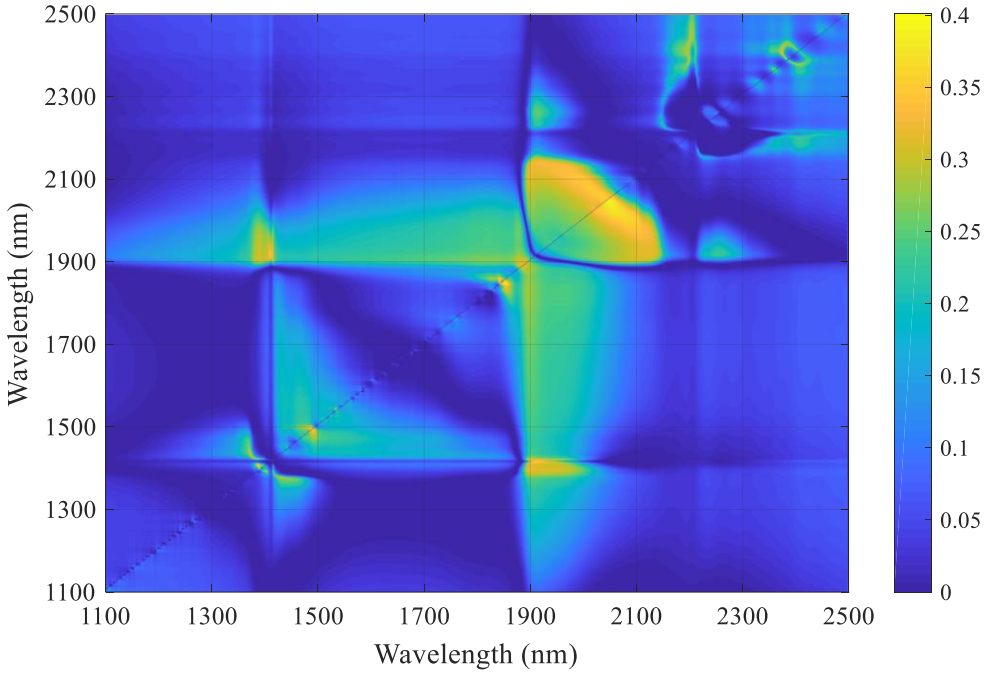

Fig. 4.6 Coefficient of determination of calibration (upper triangle) and validation (lower triangle) of simple linear regression between normalized difference reflectance indices and wilting point

were found to be relevant in case of WP. Relatively better results were obtained for the combination of reflectance values at 1844 and 1845 nm in case of FC while those at 1844 and 1856 for WP estimation (Table 4.3).

Table 4.3 Statistics of simple linear regression between the best normalized difference reflectance index (NDRI) with field capacity and wilting point

Details	**Field capacity**	**Wilting point**
Wavelengths used in NDRI	1844, 1845	1844, 1856
Intercept	21.33	5.84
Coefficient	470510.25	11236.69
F-value	204.9	116.18
p-value	3.30×10^{-33}	5.90×10^{-22}
R^2	0.49	0.40
RMSE	6.27	3.56
RPD	1.40	1.30

R^2: coefficient of determination; RMSE: root mean squared error; RPD: residual prediction deviation of validation.

The parameters of simple linear regression model between best NDRI and soil attributes together with its statistics are included in the table. When compared, the NDRI based model performance was found to be higher in case of FC than WP. The plot of observed and predicted values (using NDRI) of soil attributes is shown in Fig. 4.7.

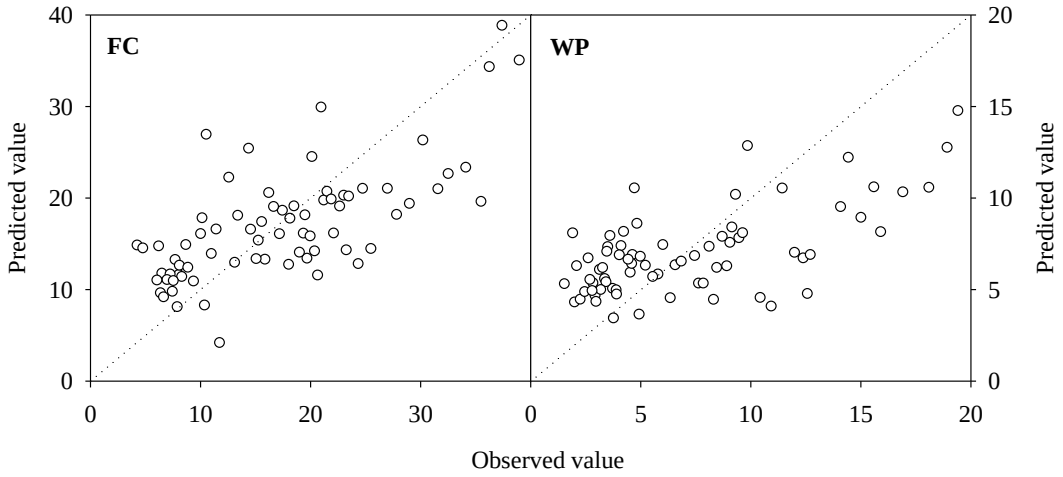

Fig. 4.7 Observed versus predicted value plots in the validation of field capacity (FC) and wilting point (WP) using the best normalized difference reflectance index. Dotted line represents the 1:1 line.

In Table 4.3, the low *p*-value ($p<0.05$) obtained for simple linear regression models suggest the rejection of null hypothesis that the coefficient of regression is zero. Further, the high *F*-value ($F>2.28$ at $\alpha=0.05$) recognized the statistical significance of predictor variable in the regression model. Although statistically significant, the regression statistics of prediction merely comply with the accuracy level (RPD<1.4) expected from laboratory based DRS. Recently, Nocita *et al.* (2013) reported that NSMI based on wavelengths 1800 and 2119 nm (Haubrock *et al.*, 2008) can be used to estimate soil moisture content ($R^2=0.60$). In their study, spectral measurements were made from soil samples wetted to pre-defined moisture levels. The moisture content being a primary chromophore, its variation significantly affects the water absorption feature around 1900 nm with shoulders near 1800 and 2119 nm. This would have enabled NSMI based estimation of soil moisture content in their study. In contrast, this study was based on NDRI computed from spectral reflectance of

dried soil alone. Even then, a significant correlation was noted between the best NDRI with soil moisture constants examined in this study. This relationship so noted for both FC and WP and moderate prediction of FC (RPD=1.40) favor the investigation on the utility of identified NDRI indices as proxy for soil moisture constants in future studies.

4.3.3.2 Full-spectrum based Calibration Functions

In general, any soil spectra would be considered to be an integration of absorption characteristics (spectrally active chromophores), scattering effects (due to orientation and packing of particles in the soil matrix effects) and undesired variations (due to measurement and operational conditions of the instrument). Both the scattering and undesired variations in the spectra be effectively removed by pre-processing to improve the reliability and robustness of calibration functions (Casa *et al.*, 2013). For the purpose, both the reflectance and absorbance spectra of soils were subjected to different transformations generally used in soil spectroscopic literature namely, *SNV*, *MSC* (Ji *et al.*, 2016; Shi *et al.*, 2016; Li *et al.*, 2015), *DT* (Buddenbaum and Steffens, 2012), *FD* (Raj *et al.*, 2018; Sarathjith *et al.*, 2014; Waiser *et al.*, 2007) and *SD* (Hong-Yan *et al.*, 2009; Ben-Dor *et al.*, 1997) and their selected pairwise combinations yielding a total of 24 spectra pre-treatments. The pre-processed spectra obtained by transforming reflectance and absorbance of a soil sample are shown as an example in Fig. 4.8 and 4.9, respectively. It can be seen that scatter correction techniques (*SNV*, *MSC* and *DT*) individually transformed the spectrum into appropriate units while the spectral pattern remain least affected as that of the untransformed counterpart. Although, the spectral pattern was not preserved in case of derivatives transformations, they were appropriate to well-portray the absorption features in the spectrum. Even visually obscured absorption features (for example, 2300-2450 nm) in the untransformed spectrum were ‘amplified’ by pre-processing using derivatives. The combined treatments of derivatives after scatter correction methods benefitted the advantages of each technique (when used individually) and hence expected to yield better results.

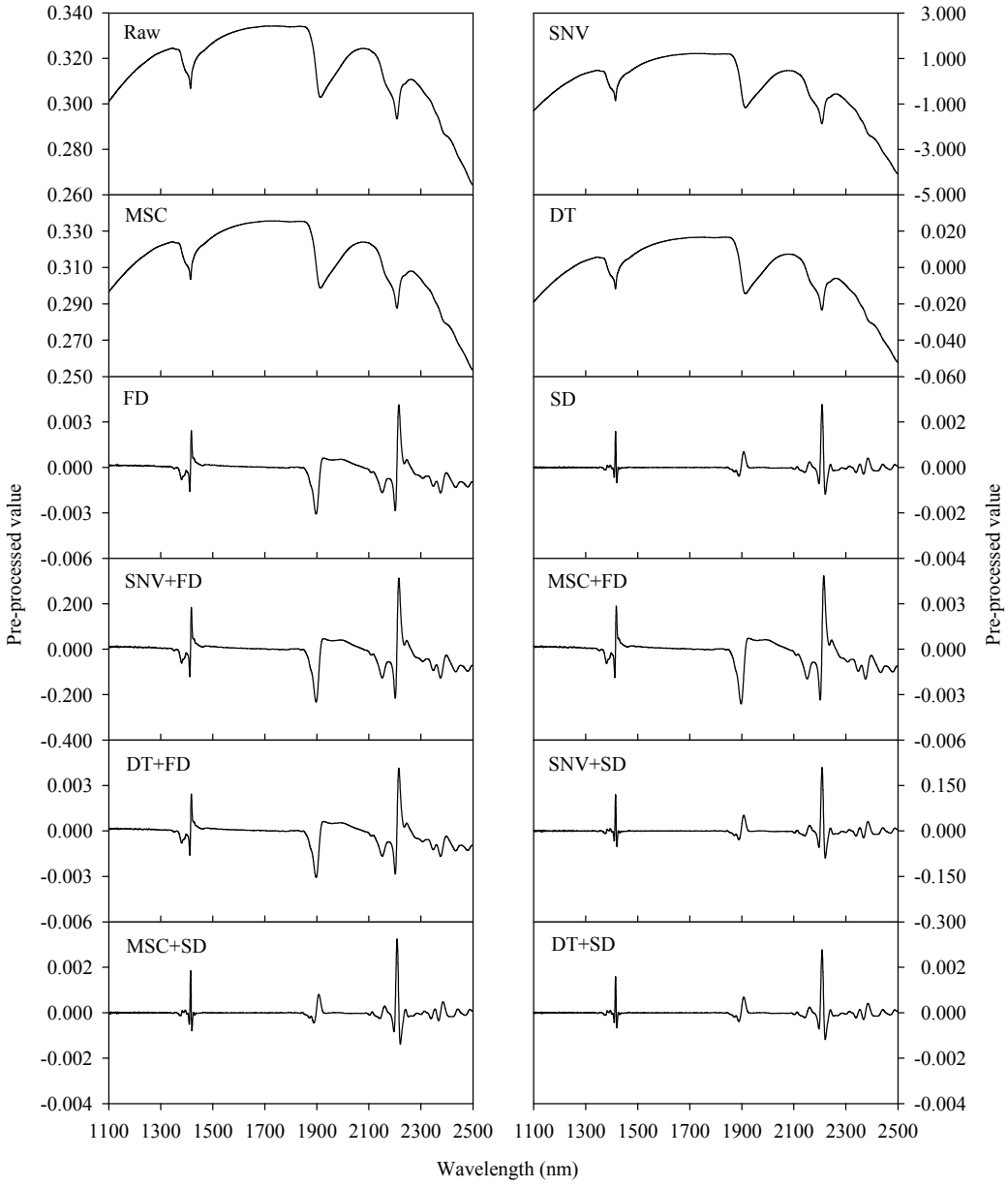

Fig. 4.8 Pre-processing of reflectance spectra of soil using different transformations. Raw: untransformed, SNV: standard normal variate, MSC: multiplicative scatter correction, DT: de-trending, FD: first derivative, SD: second derivative.

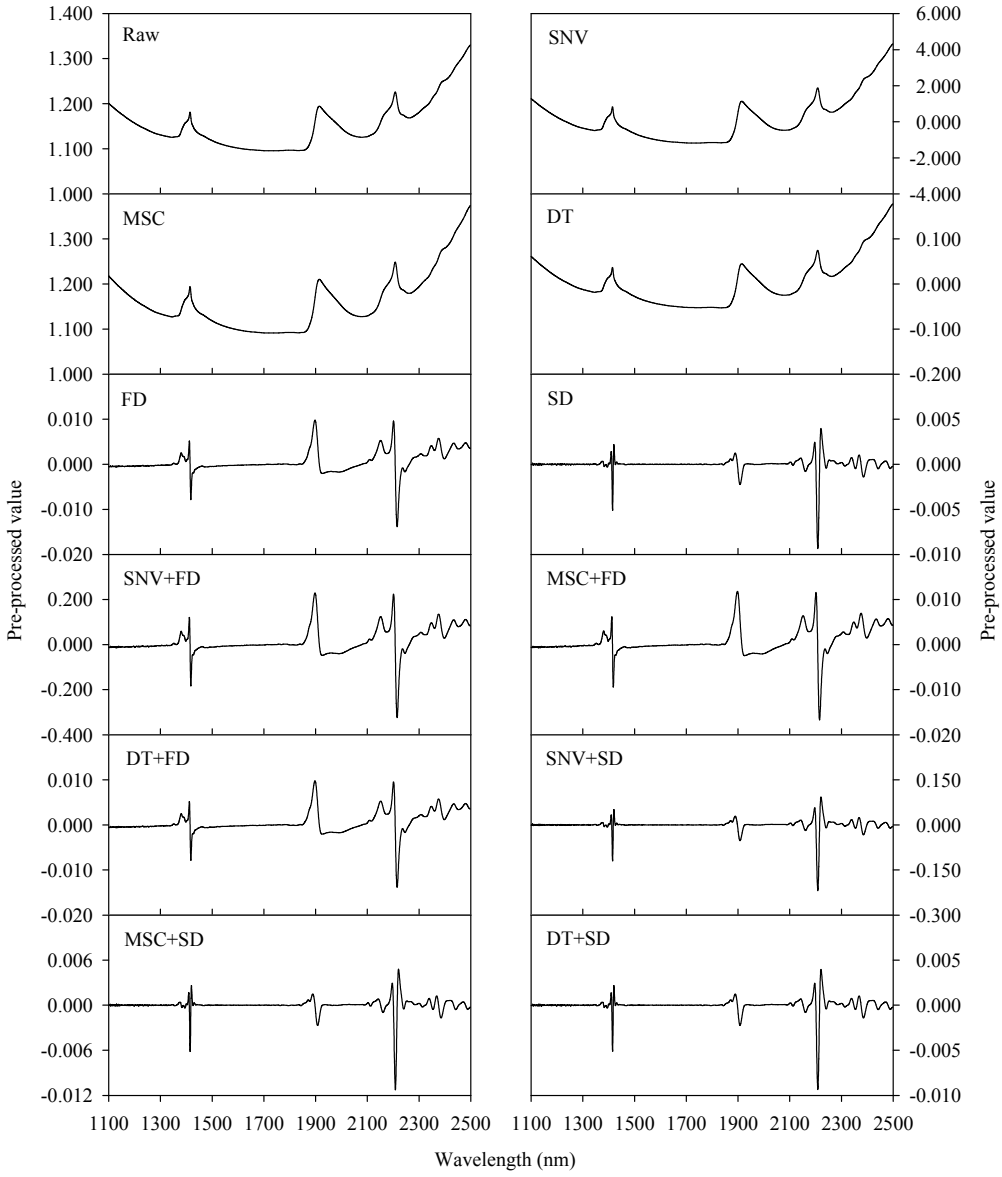

Fig. 4.9 Pre-processing of absorbance spectra of soil using different transformations. Raw: untransformed, SNV: standard normal variate, MSC: multiplicative scatter correction, DT: de-trending, FD: first derivative, SD: second derivative.

The pre-processed spectra (1100–2500 nm) obtained by aforesaid mathematical treatments of samples used for calibration were then linked with respective soil attribute (FC and WP) by means of PLSR algorithm. The performance of calibration functions so developed were later tested using the validation subset.

A prerequisite to implement PLSR was to determine an appropriate number of LV to optimize and validate calibration functions (Gholizadeh *et al*., 2016). It also help to avoid under-fitting or over-fitting of calibration functions (Kawamura et al., 2017; Viscarra Rossel, 2008). Among different approaches for the selection of number of *LV*, leave-one-out cross validation appeared to be the most effective and hence widely used in soil spectroscopic studies (Chakraborty *et al.*, 2015; Ji *et al.*, 2015; Vohland *et al.*, 2014; Kuang and Mouazen, 2011). Figure 4.10 illustrates an example of the plot (MSE versus number of *LV*) generated as an output of leave-one-out cross validation of FC and WP using *FD* of reflectance spectra which served in the selection of optimum number of *LV*. Based on minimum MSE criterion, 8 and 10 number of *LV* were found to be optimum to develop PLSR based calibration functions of FC and WP using *FD* of reflectance. In the same manner, optimum number of *LV* required to build calibration functions of the soil attributes using different pre-processed spectra was determined.

Table B1 and B2 of Appendix B lists the regression statistics of the calibration and validation of FC and WP using different spectral pretreatments. In case of FC, the R^2 and RMSE of calibration varied in the range 0.72–0.81 and 3.87–4.74, respectively while they were noted to be 0.68–0.76 and 4.30–4.98 in case of validation across different spectral pretreatments. The RPD of validation was found to be in 1.77–2.05 range of which the spectral pretreatments *R* (reflectance) and *R+DT+FD* yielded the lowest and highest values, respectively. In case of WP, the performance of calibration functions in terms of R^2 and RMSE of calibration and validation were in the tune of 0.74–0.80 & 2.12–2.42 and 0.61–0.73 & 2.36–2.85, respectively. The RPD of validation varied between 1.62 (*R+MSC*) and 1.96 (*A*) across different spectral pretreatment based calibration functions. The calibration

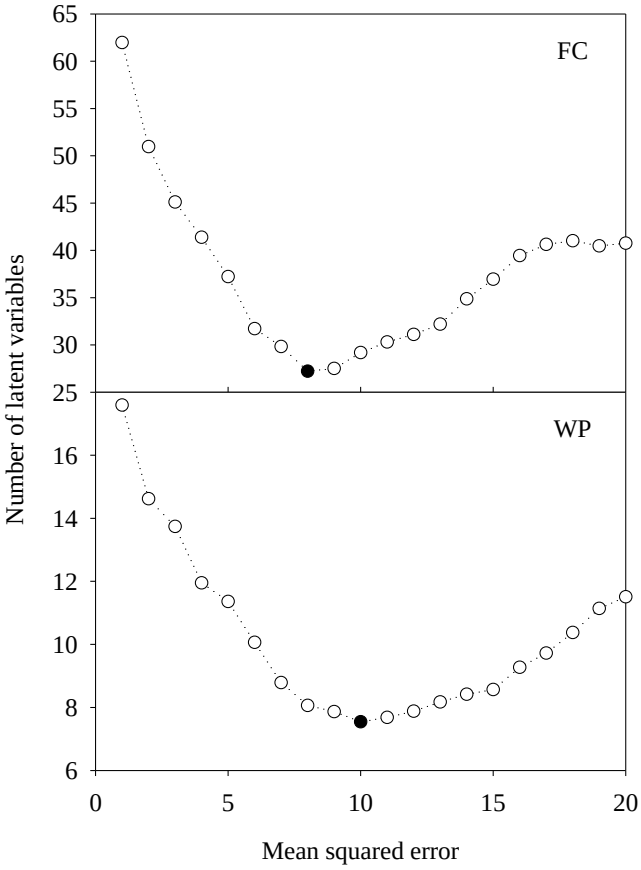

Fig. 4.10 Selection of optimum number of latent variables for PLSR based calibration functions to estimate field capacity (FC) and wilting point (WP) using first derivative of reflectance. Minimum MSE is represented as black filled marker.

functions based on *R+DT+FD* and *A* (absorbance) can be chosen as the best for the estimation of FC and WP, respectively if the selection criteria was merely based on accuracy criteria (RPD of validation). However, in this study, the selection of best spectral preprocessing considered both accuracy and complexity of calibration functions together using AIC as a combined indicator. The AIC being an estimate of information loss of a statistical model, its lower value represent a good quality model and vice versa. Thus, minimum AIC criterion was followed in the study to identify the best spectral preprocessing based calibration function of soil attributes. Accordingly, *R+SD* was found to be the best for both FC and WP among different spectral pretreatments examined in this study (Fig. 4.11). It was also noted that both

R+SD and *R+DT+SD* yielded same transformed spectra and hence the use of latter preprocessing may be replaced by the former in future studies.

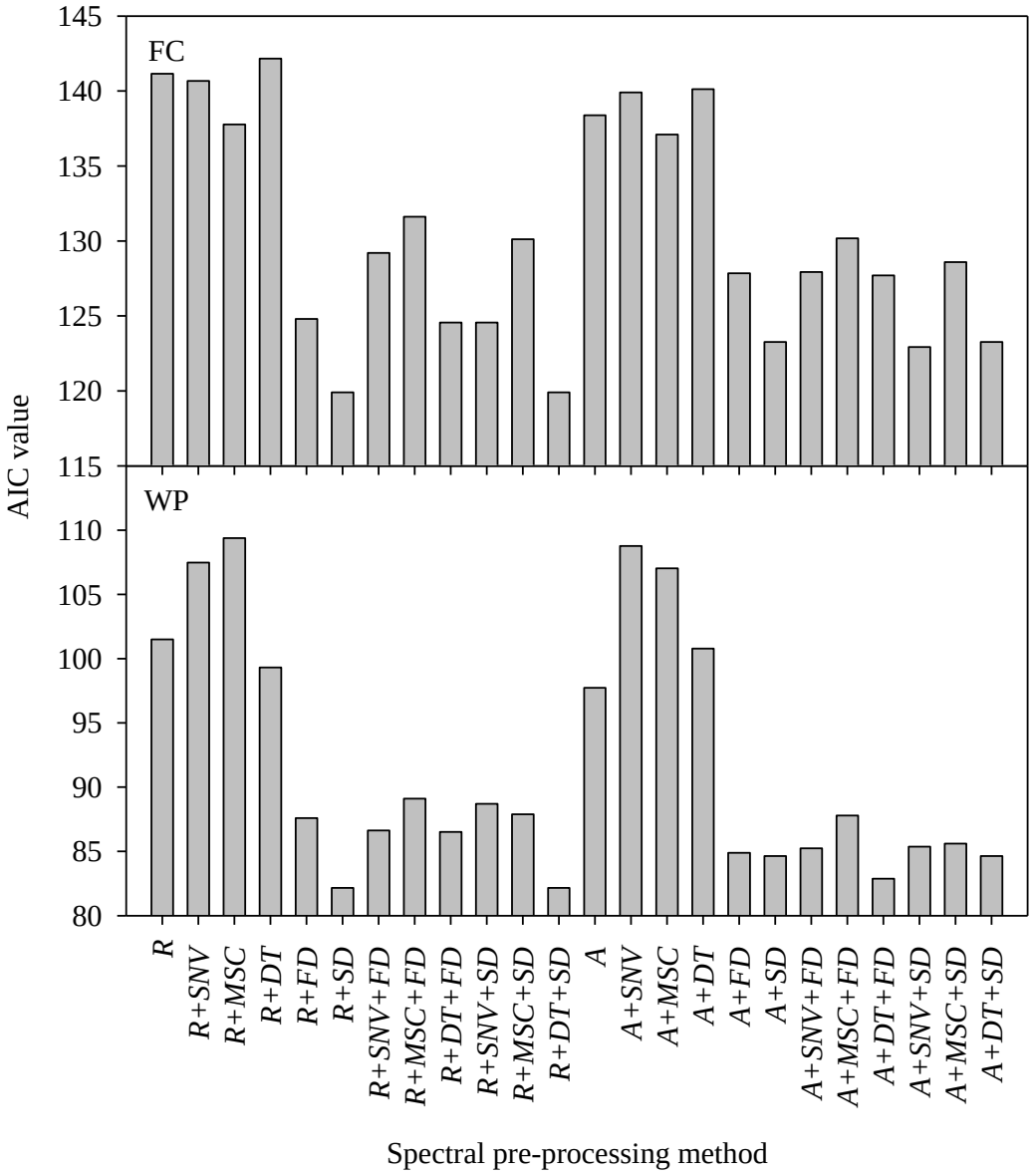

Fig. 4.11 Akaike Information Criteria (AIC) obtained using different spectra pre-processing methods for the estimation of field capacity (FC) and wilting point (WP). *R*: reflectance, *SNV*: standard normal variate, *MSC*: multiplicative scatter correction, *DT*: detrending, *FD*: first derivative, *SD*: second derivative.

The regression statistics of the best calibration functions in calibration and validation of FC and WP are given in Table 4.4 and the respective plots of observed versus predicted values are shown in Fig. 4.12.

Table 4.4 Regression statistics of prediction of field capacity (FC) and wilting point (WP) using best spectral pre-processing method

Attribute	Preprocess		Calibration			Validation			
		LV	*n*	R^2	RMSE	*n*	R^2	RMSE	RPD
FC	*R+SD*	6	220	0.77	4.30	73	0.75	4.38	2.01
WP	*R+SD*	6	218	0.76	2.34	72	0.67	2.65	1.74

n: number of soils, *LV*: number of latent variables, R^2: coefficient of determination, RMSE: root mean squared error, RPD: residual prediction deviation, *R+SD*: pairwise combination of reflectance and second derivative

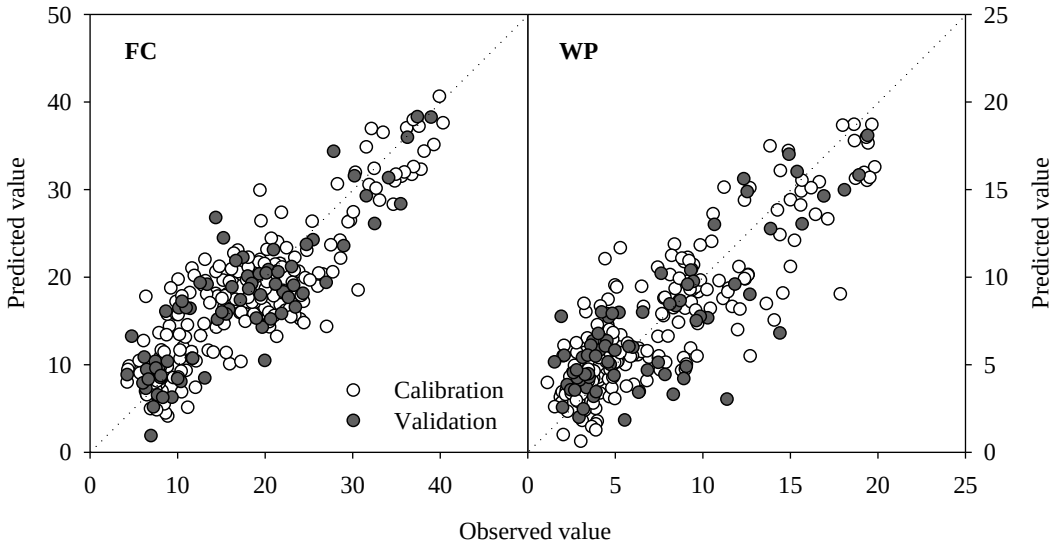

Fig. 4.12 Observed versus predicted value plots of field capacity (FC) and wilting point (WP) using full-spectrum. Dotted line represents the 1:1 line.

Based on the model accuracy criteria suggested by Chang *et al.* (2001), the best calibration function was found to be accurate (RPD>2) in the estimation of FC while moderate performance (1.4<RPD<2) was noted for WP. The regression statistics obtained in this study for FC and WP are comparable or even better to that reported in soil spectroscopic literature (Kinoshita *et al.*, 2012; Janik *et al.*, 2009). For instance, Janik *et al.* (2009) used PLSR with 18 number of *LV* to validate the

prediction of FC and WP of 249 number of soils. The accuracy of PLSR models examined in their study are relatively lower for both FC (RMSE=5.5) and WP (RMSE=3.4) than that obtained in this study (Table 4.4). Also, the regression statistics reported by Kinoshita *et al.* (2012) in the estimation of FC (R^2=0.66; RPD=1.81) and WP (R^2=0.76; RPD=1.97) are analogous to the results of this study.

The best calibration function or regression coefficient (β) that describe the linkage between *R+SD* of spectra and soil attributes are shown in Fig 4.13. To have better visual distinction on the significance of different wavelengths, a variable indicator described by the product of absolute values of β and *VIP* after normalization (β-*VIP*) was used for feature identification (Fig. 4.13).

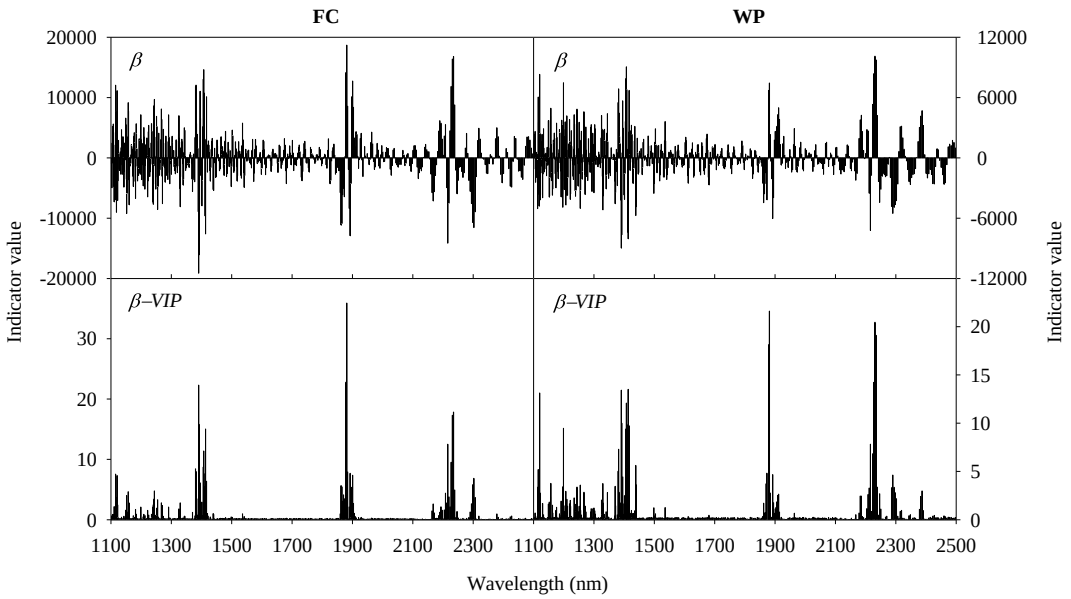

Fig. 4.13 Variable indicators describing significant wavelengths for the prediction of field capacity (FC) and wilting point (WP). β and β-*VIP* indicates the regression coefficient and its combination with variable importance of projection, respectively.

The magnitude of β-*VIP* values describe the relevance of spectral feature at respective wavelengths in prediction of soil attributes. Accordingly, water absorption features located around 1400, 1900 and 2200 nm appeared to have significant contribution in the estimation of both the soil attributes. Among them,

spectral features around 1900 nm was noted to be most prominent in both soil attributes followed by 1400 nm and 2200 nm in case of FC and WP, respectively. Apart from these wavelengths, spectral features in 1100–1400 nm range and around 2300 nm also have evident contribution in the estimation both FC and WP. The spectral characteristics around 2400 nm has noticeable contribution in the prediction of WP but not in case of FC. The most relevant wavelengths in the prediction of soil attributes based on the magnitude of regression coefficient are presented in Table 4.5. In addition, the possible functional groups related to organic carbon, clay content and iron oxides compiled from Bayer *et al.* (2012); Ben-Dor *et al.* (1997); Clark (1999) and Viscarra Rossel and Behrens (2010) which may be related with the significant wavelengths identified in this study are also included in the table. Some significant wavelengths identified *viz*. 1116, 1120, 1157, 1244, 1255, 1391, 1395, 1407, 1414, 1417, 1881, 1892 and 2216 nm appeared to be common in the estimation of both FC and WP. The reason for similar wavelengths in prediction may be attributed to the inherent correlation between FC and WP values. In contrast, the spectral features at 1330, 1370, 1439, 2290 and 2387 nm appeared to be most prominent in case of WP while no relevant wavelengths were found near to them for the estimation of FC. Similarly, the wavelengths at 1862, 1901, 2208 and 2302 nm were most significant for the estimation of FC but have less relevance in case of WP.

4.3.3.3 Variable Selection based Calibration Functions

This study deployed the utility of the variable indicators based OPS approach for spectral variable selection. In this approach, absolute magnitude of variable indicators (after normalization) were used in conjunction with EDF to select spectral variables for subset models. Among the subset models (with different NSV) generated using a particular variable indicator, the optimum subset model and respective optimum NSV were identified based on the minimum RMSE criteria. An illustrative example of OPS plot that aid the optimum subset model selection is shown in Fig. 4.14 for the case of FC using *r* as the variable indicator.

Table 4.5 Functional groups assigned for most significant wavelengths in the estimation of field capacity (λ_{FC}) and wilting point (λ_{WP})

λ_{FC}	λ_{WP}	$\lambda_{reported}$	Constituent	Functional group assigned	Reference[†]
1116, 1118, 1120	1114, 1116, 1120	1100	Organics	3*v*1 of aromatics	[1]
1152, 1157	1157	1170	Organics	3*v*1 of asymmetric–symmetric doublet	[1]
-	1197, 1199	1201, 1203	Organics	Oil/cellulose/wax	[2]
1244, 1255	1244, 1255	-	-	-	-
-	1330	-	-	-	-
-	1370	1367	Organics	OH in water of cellulose/lignin/starch	[2]
1381	1382	1380	Iron oxides	*v*1+*v*3 of water	[1]
1391, 1395	1391, 1395	1395	Clay	2*v*1a of kaolin doublet	[1]
1407	1407	1400	Iron oxides	2*v*1 of hydroxyl	[1]
1414, 1417	1414, 1417	1415	Clay	2*v*1b of kaolin doublet	[1]
-	1439	1449	Organics	4*v*1 of carboxylic acids	[1]
1862	-	1870	Carbonates	-	[3]
1871	-	1870	Carbonates	-	[3]
-	1873	1870	Carbonates	-	[3]
1881	1881	-	-	-	-
1892	1892	-	-	-	-
1901	-	1915	Iron oxides	*v*2+*v*3 of water	[1]
2208	-	2208	Clay	Doublet and *v*+*v* of OH stretch of	[4]
2216	2216	2216	Clay	Illite	[4]
2235	2231	2230	Clay	*v*1+*δ*b of smectite	[1]
-	2290	2279	Organics	3*v* of CH_2, CH_3	[2]
2302	-	2300	Organics	*v*+*v*4 of CH stretch	[4]
-	2387	2386	Organics	3*v* of COO–,CH3 of pectin/protein	[2]

[†] [1] Viscarra Rossel and Behrens (2010); [2] Ben-Dor *et al.* (1997); [3] Clark (1999); [4] Bayer *et al.* (2012)
v and *δ* represents stretching and bending modes, respectively.

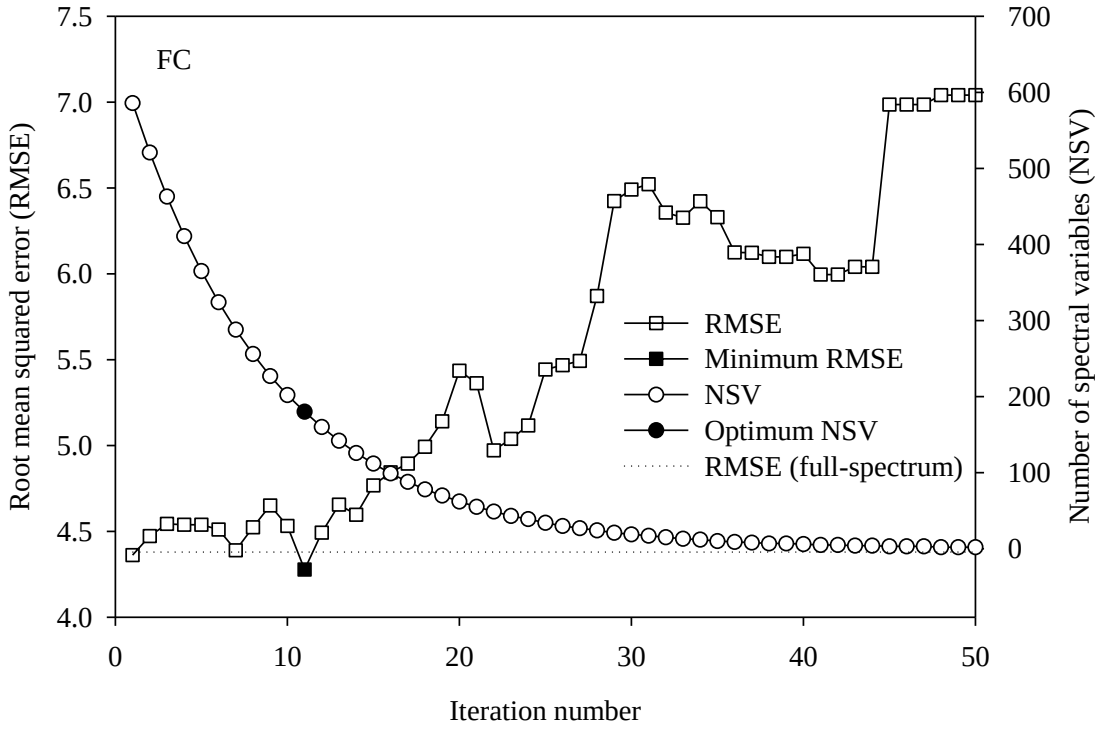

Fig. 4.14 OPS plot based on regression coefficient for the case of field capacity (FC)

The figure depicts the plot of RMSE values in the validation of subset models together with the NSV which were used as predictors of PLSR model. The trend in RMSE values obtained in this study appeared to be similar to that reported by Sarathjith *et al.* (2016). It may be noted that the RMSE values of subset models generated during initial iterations (1 to 14) were comparable to that of full-spectrum counterpart (Table 4.4). Among them, subset models at iteration 1 and 11 yielded better performance (lower RMSE value) than full-spectrum model. The EDF assisted removal of irrelevant/noisy wavelengths from data modeling might have resulted in better performance of these subset models. In case of other subset models (up to iteration 14), the wavelength removal has marginally affected their performance as indicated by slight high RMSE values than that of full-spectrum model. After iteration 14, all the subset models generated appeared to have considerable reduction in performance for the cause being the elimination of relevant and informative wavelengths. It may be noted that all the subset models

with most significant wavelengths (iteration 29 to 50) were not satisfactory. This revealed the importance of combination of most significant wavelengths with others for better model performance (Yun *et al.*, 2014).

In the example shown in Fig. 4.14, subset model generated at iteration 11 appeared to have lowest RMSE among others and hence chosen as the optimum subset model (NSV=180; R^2=0.76; RMSE=4.28; RPD=2.06). In the same manner, the optimum subset model and optimum NSV for all variable indicators were identified. Later, the regression statistics of optimum subset models of both FC (Table C1 of Appendix C) and WP (Table C2 of Appendix C) were compiled to be used to identify the best subset model and variable indicator for each soil attribute separately.

Figure 4.15 depicts the percent difference in RMSE of validation values of optimum subset models with respect to that of full-spectrum models (represented as zero reference line) of FC and WP. The NSV used to build optimum subset models are also indicated in the figure. The positive and negative bars represents optimum subset models of inferior (high RMSE) and superior (low RMSE) performance as that of full-spectrum counterpart, respectively. From the figure, the following remarks may be made, a) the variable indicators namely, *β-CovProc, β-r*, *β-StN, bicor-CovProc, SqRes-StN, StN, StN-CovProc* and *VIP-StN* were capable to yield optimum subset models with better performance than full-spectrum model in case of both FC and WP, b) none of the PLSR-dependent variable indicators yielded satisfactory results when used individually in both soil attributes while the performance of *StN* (PLSR-independent) was successful, c) pairwise combination of variable indicators appeared to be more reliable than they used individually, d) among the successful optimum subset models in both soil attributes, pairwise combinations of PLSR-dependent with PLSR-independent variable indicators appeared to yield better results in most cases (except *CovProc* and *bicor-CovProc*), e) the variable indicator based OPS approach appeared to be successful in both soil attributes as at least one variable indicator among others yielded optimum subset models with improved performance.

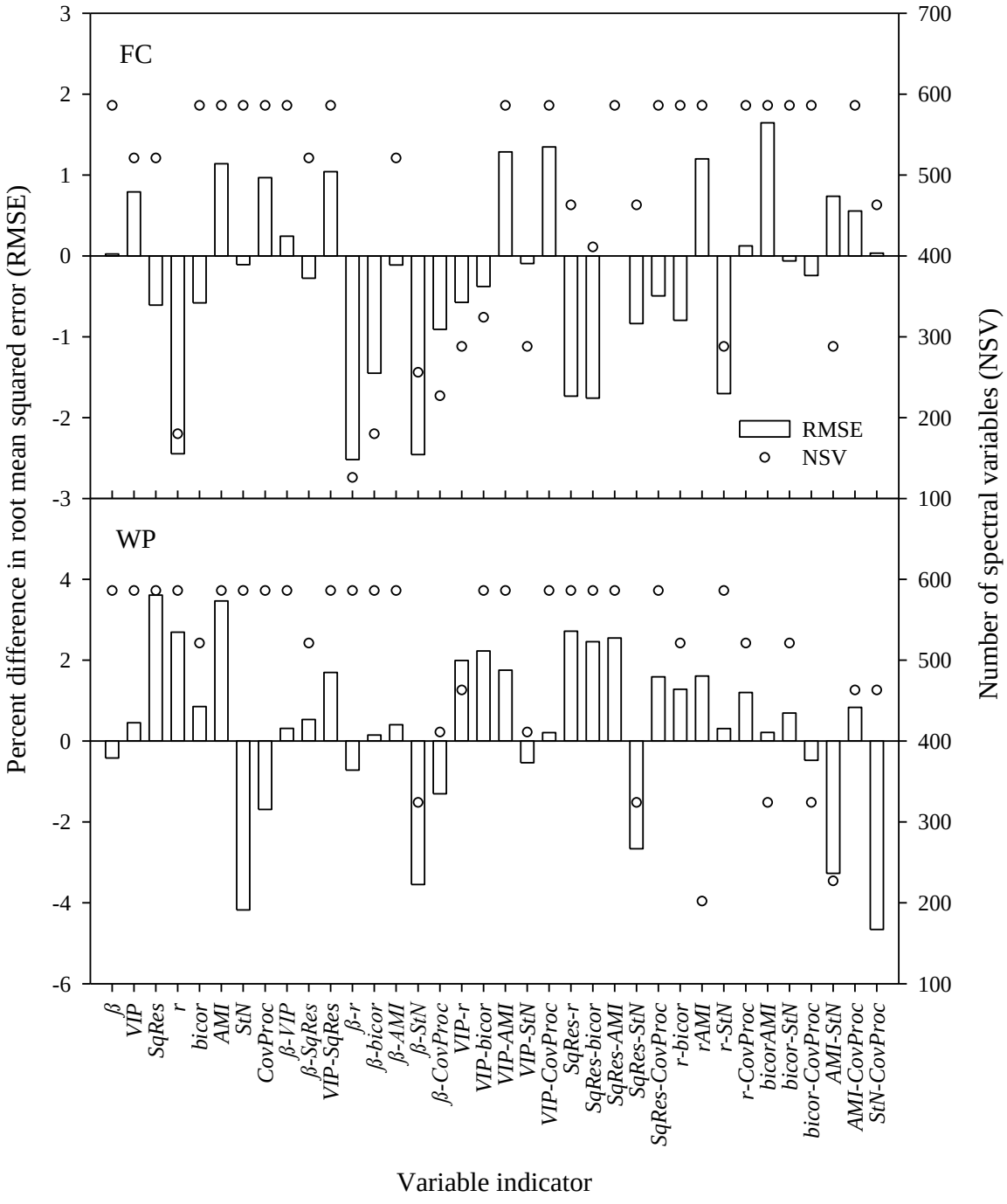

Fig. 4.15 Percent difference in root mean squared error of optimum subset models compared to full-spectrum models

With a view to identify the best subset model among optimum subset models, a two-fold criteria with regard to measures of accuracy (RMSE) and complexity (NSV) was implemented. Initially, all the optimum subset models with low RMSE

values (negative bars in Fig. 4.15) were considered. Then, the one among the selected optimum subset models with lowest NSV value was regarded as the best subset model. Also, the variable indicator associated with the best subset model was considered to the best variable indicator for the soil attribute of interest. Accordingly, optimum subset model based on *β-r* (NSV=126; RMSE=4.27) was found to be the best subset model for the estimation of FC and *AMI-StN* (NSV=227; RMSE=2.56) appeared to be appropriate for WP.

Further, we inspected whether the performance noted for optimum subset models (increase, decrease or similar) were statistically significant with respect to respective full-spectrum model. For the purpose, a distribution of RMSE of validation was generated for each optimum subset model and also for full-spectrum model by means of bootstrapping with replacement for 500 times. Then, the distributions of optimum subset models and full-spectrum model were examined for their similarity with respect to mean value via the implementation of both right-tail and left-tail Student's *t*-tests at 5% level of significance (α=0.05). The equality of mean values of these distributions formed the null hypothesis of the tests. The generated distributions to be compared were assumed to be normal with unknown and equal variances. The results of both the tests were used conjunctively to examine the similarity/dissimilarity in prediction performance between full spectrum and optimum subset models.

An illustrative example to depict the nature of generated distributions (in terms of kernel smoothing density estimates) and the results of Student's *t*-tests is shown in Fig. 4.16. Based on the *p*-value of right-tailed (p_r) and left-tailed (p_l) tests, the following interpretations were made, a) if $p_r > \alpha$ and $p_l < \alpha$, then optimum subset model outperform full-spectrum model. In other words, the mean value of RMSE distribution of optimum subset model is less than that of full-spectrum model (eg. Fig. 4.16a), b) $p_r > \alpha$ and $p_l > \alpha$ imply that both full-spectrum and optimum subset model have similar performance (eg. Fig. 4.16b), c) $p_r < \alpha$ and $p_l > \alpha$ indicate the inferior performance of optimum subset model with regard to full-spectrum model (eg. Fig. 4.16c).The p_r and p_l values of all optimum subset models were generated

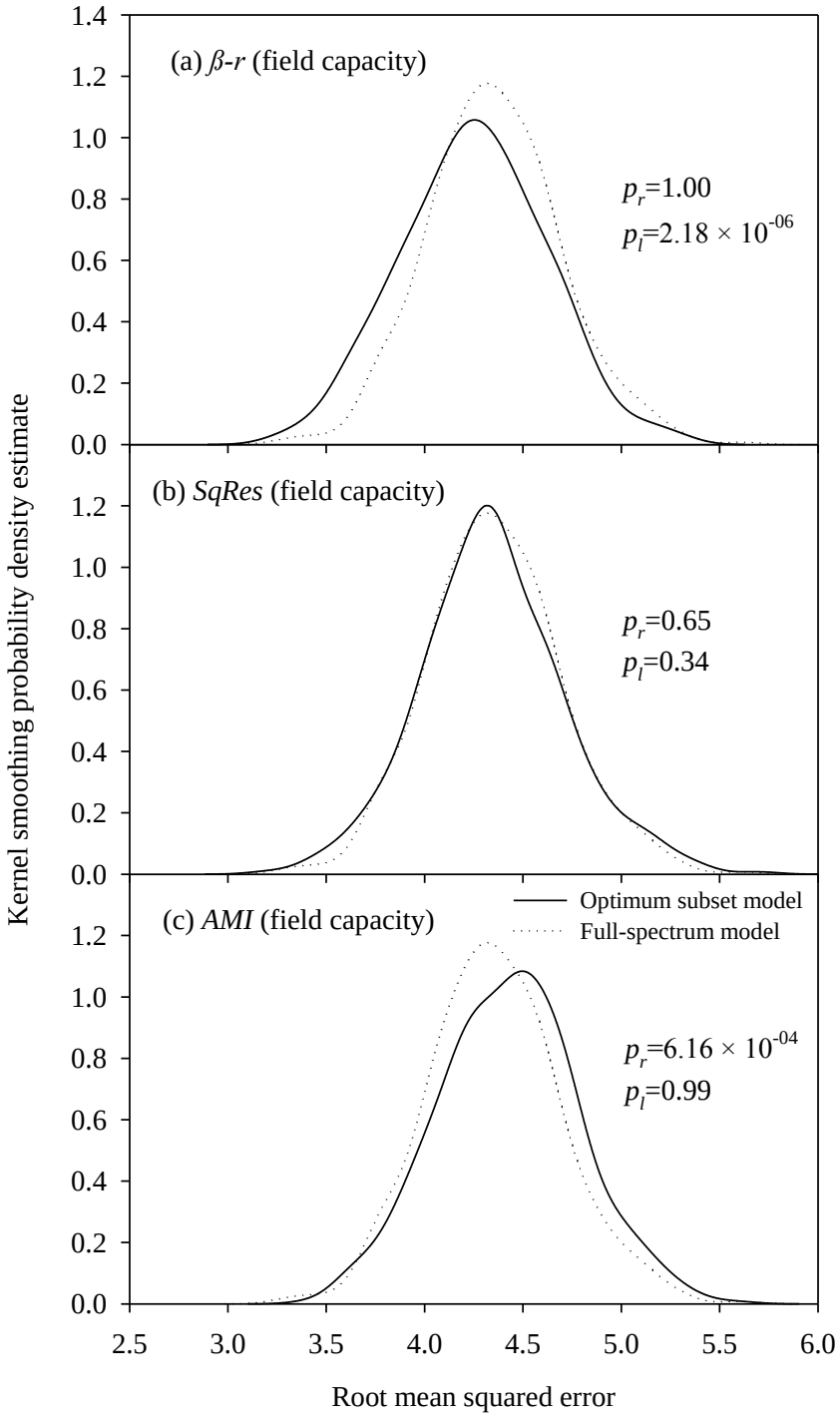

Fig. 4.16 Kernel smoothing density estimates of the distribution of root mean squared error of validation; p_r and p_l indicate the *p*-value of right- and left-tailed tests, *ß-r* (pairwise combination of coefficients of regression and correlation), S*qRe*s (squared residual) and *AMI* (adjacency values of mutual information) are variable indicators.

and examined for their similarity/dissimilarity with full-spectrum models (Fig. 4.17). Based on the three criteria discussed above, the optimum subset models were classified as better, similar or poor in performance compared to full spectrum models. Accordingly, optimum subset models of FC based on variable indicators namely, *r*, *r-StN*, *SqRes-r*, *SqRes-bicor*, *β-r*, *β-bicor* and *β-StN* were found to be better than full-spectrum model. In case of WP, *StN*, *CovProc*, *β-StN*, *β-CovProc*, *SqRes-StN*, *AMI-StN* and *StN-CovProc* yielded optimum subset models of better performance. Several variable indicators yielded optimum subset models with similar performance as that of full-spectrum model which include, *bicor, bicor-CovProc, bicor-StN*, *r-bicor*, *r-CovProc*, *SqRes*, *SqRes-CovProc*, *SqRes-StN*, *StN*, *StN-CovProc*, *VIP-bicor*, *VIP-r*, *VIP-StN*, *β, β-AMI, β-CovProc, β-SqRes* in case of FC and *β, VIP, β-VIP, β-SqRes, β-r, β-bicor, β-AMI*, *VIP-StN*, *VIP-CovProc*, *r-StN*, *bicor-AMI*, *bicor-StN*, *bicor-CovProc* in case of WP. All the remaining variable indicators in each soil attribute yielded poor optimum subset models. According to Fernández Pierna *et al.* (2009), the basic objective of variable selection strategy is to generate models with better/similar performance using less NSV (optimum subset model) than the original set of spectral variables (full-spectrum model). The results of the variable selection approach implemented in this study is in agreement with the argument supported by the number of satisfactory optimum subset models of FC (24 out of 36 cases) and WP (20 out of 36 cases). Among them, optimum subset model generated using spectral variables identified based on *β-r* and *AMI-StN* was found to the best for the estimation of FC and WP, respectively as endorsed by the results of statistical analysis.

Figure 4.18 depicts the wavelengths identified by *β-r* and *AMI-StN* to build the best subset models of FC and WP, respectively. A few remarkable observations may be made from the figure, a) the spectral information content in four broad wavelength bands, roughly 1570–1670, 1730–1825, 2030–2160 and 2300–2500 nm was found to be irrelevant for the estimation of FC, b) wavelengths from all the spectral regions were found to be important for the estimation of WP, c) water absorption features related to clay mineral around 2200 nm appeared to relevant in case of FC

while the spectral region was discarded in case of WP to attain the noted level of accuracy. It may also be noted that the best subset models used only 19.09% (NSV=126 in case of FC) and 34.39% (NSV=227 in case of WP) of the total NSV to yield performance better than full-spectrum model. This demonstrated the ability of variable indicator based OPS approach implemented in this study to develop parsimonious models for reliable estimation of FC and WP of soil.

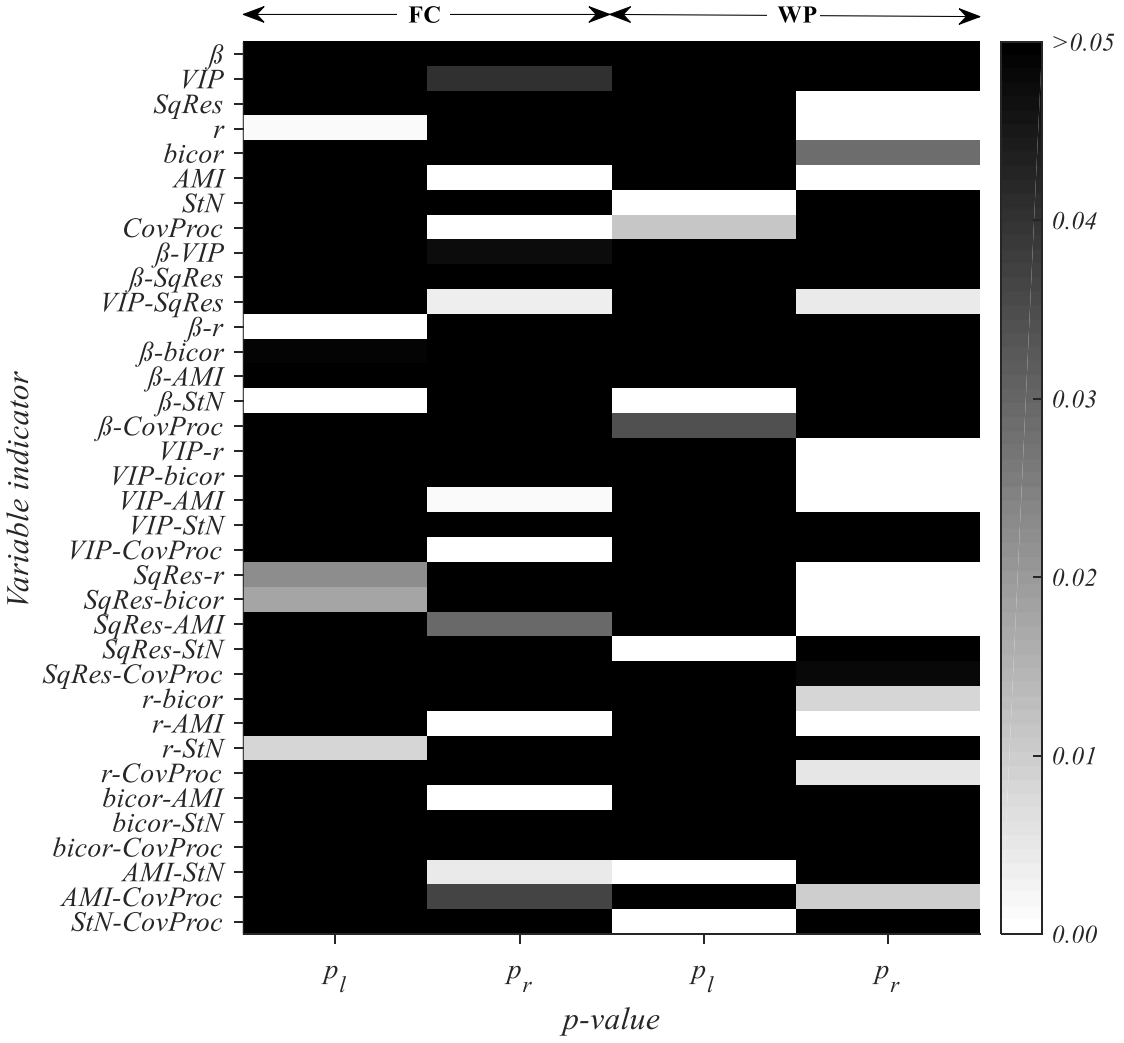

Fig. 4.17 Result of comparison (*p*-value) between generated distributions of root mean squared error of optimum subset and full-spectrum models upon validation for field capacity (FC) and wilting point (WP). p_l and p_r denotes the *p*-value obtained for left-and right-tailed Student's *t*-tests, respectively.

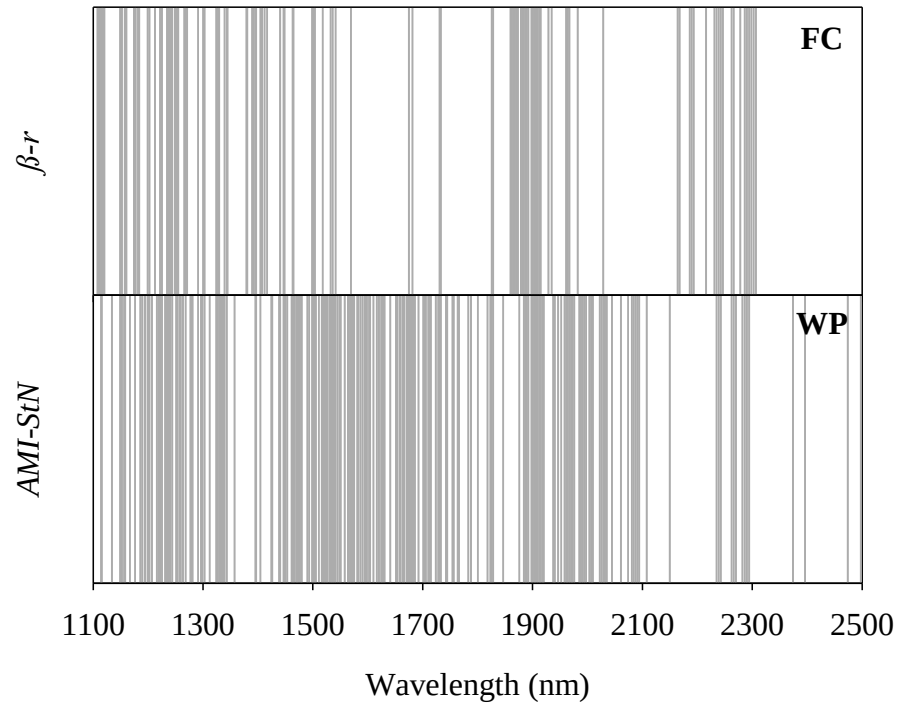

Fig. 4.18 Spectral variables associated with best subset models of field capacity (FC) and wilting point (WP). *β-r* (pairwise combination of coefficients of regression and correlation), *AMI-StN* (pairwise combination of adjacency values of mutual information with signal-to-noise vector) are the best variable indicators identified.

4.3.3.4 Comparison of Data Modelling Approaches

The present study evaluated the utility of three main approaches typically used in DRS of soils viz. spectral index (NDRI), full-spectrum based PLSR and variable selection (OPS). Under each approach, several calibration functions were evaluated and the best among others was identified as discussed in sections 4.3.3.1, 4.3.3.2 and 4.3.3.3. The performance of the best calibration functions were then compared. For the purpose, distribution of RPD values in the validation of the best calibration functions were generated by bootstrapping with replacement for 500 iterations. Figure 4.19 depicts the Kernel smoothing density estimates of the generated RPD distributions of FC and WP. The accuracy classification scheme

suggested by Chang *et al.* (2001) is also provided in the figure to aid model performance evaluation.

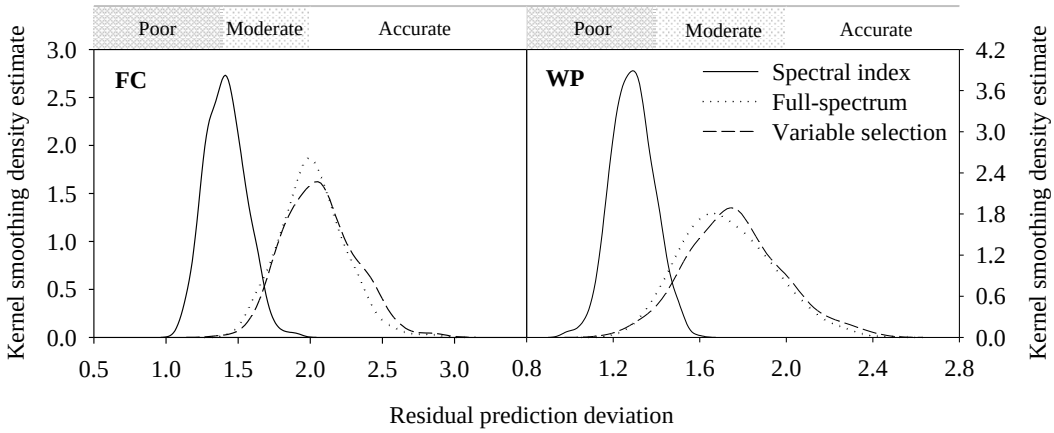

Fig. 4.19 Kernel smoothing density estimates of the distribution of residual prediction deviation of validation of field capacity (FC) and wilting point (WP) obtained using the best calibration function of spectral index, full-spectrum and variable selection approaches.

The mean value of RPD distribution in case of FC and WP in ordered pairs was found to be (1.4, 1.3), (2.0, 1.7) and (2.1, 1.8) as obtained using NDRI, full-spectrum and OPS approaches, respectively. Accordingly, calibration functions developed using NDRI were found to have poor estimation of soil moisture constants. In contrast, both full-spectrum and OPS based calibration functions were regarded as accurate in case of FC while they yielded moderate estimation of WP. Among them, OPS based calibration functions yielded better RPD value invariably across both the soil moisture constants. Moreover, the OPS based calibration functions has fewer NSV (less complex) than that of full-spectrum. Hence, the OPS approach was regarded as the best in this study to estimate the soil moisture constants from spectral reflectance of dry soil.

SUMMARY AND CONCLUSIONS

CHAPTER 5

SUMMARY AND CONCLUSIONS

Rapid and accurate assessment of soil moisture constants namely, FC and WP as descriptors of WHC is a major pre-requisite to scientific irrigation scheduling. Although several conventional methods exists for their estimation, majority of them are cumbersome and time-consuming and not appropriate for their use at varying space and time scales. Over the past few decades, DRS has gained popularity as an effective tool for soil analysis and digital soil mapping. However, many studies have revealed that the approach is not always satisfactory. Despite this shortcoming, the approach retained its relevance by its ability to rapidly and cost-effectively assess multiple soil attributes in both non-destructive and non-invasive manner. This study investigated the utility of DRS approach in estimating both FC and WP directly from spectral signature of dry soil. This study was performed using an available soil database comprised of spectral signature and moisture content values at 0.33 bar (FC) and 15 bar (WP) of about 300 soil samples.

Calibration functions describing the linkage between soil spectra (1100–2500 nm) and soil moisture constants were developed via three different statistical approaches typically employed in DRS studies of soil. In the first approach based on spectral indices, NDRI were computed for all the pairwise combinations of wavelengths and subsequently related to soil moisture constant values by means of simple linear regression. The NDRI which combined the spectral features at 1844 and 1845 nm (in case of FC) and at 1844 and 1856 nm (in case of WP) yielded statistically significant relationship between spectra and soil attributes. However, the results do not comply with the minimum accuracy level needed for DRS models of soils.

In the second approach based on full-spectrum, PLSR was employed to develop the desired calibrations. The spectra was subjected to various pre-processing techniques and the best one was identified based on AIC values. Accordingly, *R+SD* was found to be the more appropriate pre-processing method

and hence respective calibration functions were treated as the best for both FC and WP. The performance of the full-spectrum calibration function was found to be accurate in case of FC (RPD=2.01) while moderate performance was noted in case of WP (RPD=1.74) based on the accuracy classification suggested by Chang *et al.* (2001).

In the third approach, spectral variable selection was performed as a way to develop simple, reliable and parsimonious calibration functions. An OPS approach based on variable indicators was implemented for the purpose. A variety of PLSR-dependent and -independent variable indicators and their pairwise combinations were tested. The OPS approach made use of an EDF in an iterative manner to select different NSV based on the absolute values of variable indicator. At each iteration, the selected NSV were related to soil attribute via PLSR (subset model). Later, the optimum model among the subset models was identified based on mimimum-RMSE criteria of validation. Later, distributions of RMSE in the validation of optimum subset and full-spectrum models were generated and compared for their similarity with respect to mean value using both left and right tailed Student's *t*-tests. Based on the *p*-value of the test, optimum models that yielded either similar or better performance to that of full-spectrum were selected and the one among them with lowest NSV was regarded as the best. Accordingly, best subset model of FC and WP was based on *β-r* (NSV=126; RMSE=4.27) and *AMI-StN* (NSV=227; RMSE=2.56), respectively.

Among the different approaches compared, variable indicator based OPS approach outperformed both spectral indices and full-spectrum counterparts with regard to accuracy, simplicity and parsimony together of the developed calibration functions of FC and WP. Hence, the overall results of this study suggest OPS approach to develop simple, effective and parsimonious calibration functions that directly translate spectral information into soil moisture constant values.

Research Perspectives

This study has demonstrated the utility of DRS approach to estimate soil moisture constants namely, FC and WP from spectral signature of dry soil. Among the different approaches compared, OPS based calibration functions outperformed those developed using spectral indices and full-spectrum. Thus, the OPS approach implemented in this study may be advocated in conjunction with existing or future spectral libraries to have relatively simple and parsimonious DRS models to estimate soil moisture constants. Nevertheless, the efficacy of other calibration algorithms not implemented in this study may also be examined for improved results. To generate more robust DRS calibration functions of FC and WP, the approach implemented in the study may be extended using large spectral database of different soils. For more practical utility of the DRS technique, the feasibility of transferring the developed laboratory based calibration functions for their use in conjunction with ground, airborne and space borne hyperspectral measurements may be investigated in future studies.

REFERENCES

REFERENCES

Abdi, D., Tremblay, G.F., Ziadi, N., Bélanger, G., and Parent, L.-É. 2012. Predicting soil phosphorus-related properties using near-infrared reflectance spectroscopy. *Soil Sci. Soc. Am. J.*, 76(6):2318–2326.

Akaike, H. 1973. Information theory and an extension of maximum likelihood principle., *in Second International Symposium on Information Theory, B.N. Petrov, F. Csáki, eds,* Akadémia Kiadó, Budapest, Hungary, pp. 267–281.

Alamar, M.C., Bobelyn, E., Lammertyn, J., Nicolaï, B.M., and Moltó, E. 2007. Calibration transfer between NIR diode array and FT-NIR spectrophotometers for measuring the soluble solids contents of apple. *Postharvest Biol. Technol.*, 45(1):38–45.

Babaeian, E., Homaee, M., Vereecken, H., Montzka, C., Norouzi, A.A., and van Genuchten, M.T. 2015. A comparative study of multiple approaches for predicting the soil–water retention curve: hyperspectral information vs. basic soil properties. *Soil Sci. Soc. Am. J.*, 79(4):1043–1058.

Barnes, R., Dhanoa, M., and Lister, S. 1989. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.*, 43(5):772–777.

Battiti, R. 1994. Using mutual information for selecting features in supervised neural net learning. *IEEE Trans. Neural Netw.*, 5(4):537–550.

Bayer, A., Bachmann, M., Müller, A., and Kaufmann, H. 2012. A comparison of feature-based MLR and PLS regression techniques for the prediction of three soil constituents in a degraded South African ecosystem. *Appl. Environ. Soil Sci.*, 2012(1):1–20.

Bellon-Maurel, V. and McBratney, A. 2011. Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils – Critical review and research perspectives. *Soil Biol. Biochem.*, 43(7):1398–1410.

Ben-Dor, E. 2011. Characterization of Soil Properties Using Reflectance Spectroscopy., *in Hyperspectral Remote Sensing of Vegetation, P.S. Thenkabail, J.G. Lyon, A. Huete, eds,* CRC Press, pp. 513–558.

Ben-Dor, E., Chabrillat, S., Demattê, J.A.M., Taylor, G.R., Hill, J., Whiting, M.L., and Sommer, S. 2009. Using Imaging Spectroscopy to study soil properties. *Remote Sens. Environ.,* 113:S38–S55.

Ben-Dor, E., Inbar, Y., and Chen, Y. 1997. The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400–2500 nm) during a controlled decomposition process. *Remote Sens. Environ.,* 61(1):1–15.

Ben-Dor, E., Irons, J.R., and Epema, G.F. 1999. Soil reflectance., *in Remote Sensing for the Earth Sciences: Manual of Remote Sensing, N.A. Rencz, ed*, John Wiley & Sons, New York, pp. 111–188.

Bilgili, A.V., Es, H.M. Van, Akbas, F., Durak, A., and Hively, W.D. 2010. Visible-near infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey. *J. Arid Environ.,* 74(2):229–238.

Bokobza, L. 2002. Origin of near-infrared absorption bands., *in Near Infrared Spectroscopy: Principles, Instruments, Applications, H.W. Siesler, Y. Ozaki, S. Kawata, H.M. Heise, eds*, Wiley-VCH, Weinheim, Germany, pp. 11–41.

Boulesteix, A. and Strimmer, K. 2007. Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. *Brief. Bioinform.,* 8(1):32–44.

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

Brown, D.J. 2007. Using a global VNIR soil-spectral library for local soil characterization and landscape modeling in a 2nd-order Uganda watershed. 140:444–453.

Brown, D.J., Shepherd, K.D., Walsh, M.G., Dewayne Mays, M., and Reinsch, T.G. 2006a. Global soil characterization with VNIR diffuse reflectance spectroscopy. *Geoderma,* 132(3–4):273–290.

Brown, D.J., Shepherd, K.D., Walsh, M.G., Mays, M.D., and Reinsch, T.G. 2006b. Global soil characterization with VNIR diffuse reflectance spectroscopy. *Geoderma,* 132(3–4):273–290.

Bryant, R., Thoma, D., Moran, S., Goodrich, D., Keefer, T., Paige, G., and Skirvin, S. 2003. Evaluation of hyperspectral, infrared temperature and radar measurements for monitoring surface soil moisture., *in 1st Interagency Conference on Research in the Watersheds,* pp. 528–533.

Buddenbaum, H. and Steffens, M. 2012. The effects of spectral pretreatments on chemometric analyses of soil profiles using laboratory imaging spectroscopy. *Appl. Environ. Soil Sci.,* 2012:1–12.

Campbell, G.S. and Campbell, M.D. 1982. Irrigation scheduling using soil moisture measurements: theory and practice., *in Advances in Irrigation, D. Hillel, ed,* Academic Press, New York, pp. 25–42.

Casa, R., Castaldi, F., Pascucci, S., Palombo, A., and Pignatti, S. 2013. A comparison of sensor resolution and calibration strategies for soil texture estimation from hyperspectral remote sensing. *Geoderma,* 197–198:17–26.

Castaldi, F., Palombo, A., Santini, F., Pascucci, S., Pignatti, S., and Casa, R. 2016. Evaluation of the potential of the current and forthcoming multispectral and hyperspectral imagers to estimate soil texture and organic carbon. *Remote Sens. Environ.,* 179:54–65.

Chabrillat, S., Goetz, A.F.., Krosley, L., and Olsen, H.W. 2002. Use of hyperspectral images in the identification and mapping of expansive clay soils and the role of spatial resolution. *Remote Sens. Environ.,* 82(2–3):431–445.

Chakraborty, S., Weindorf, D.C., Paul, S., Ghosh, B., Li, B., Ali, N., Kumar, R., Ray, D.P., and Majumdar, K. 2015. Diffuse reflectance spectroscopy for monitoring lead in land fill agricultural soils of India. *Geoderma Reg.,* 5:77–85.

Chang, C.-W., Laird, D.A., Mausbach, M.J., and Hurburgh, C.R. 2001. Near-infrared reflectance spectroscopy–Principal components regression analyses of soil properties. *Soil Sci. Soc. Am. J.,* 65(2):480–490.

Chang, C. and Laird, D. 2002. Near-infrared reflectance spectroscopic analysis of soil C and N. *Soil Sci.,* 167(2):110–116.

Clark, R.N. 1999. Spectroscopy of rocks andminerals, and principles of spectrocopy., *in Remote Sensing for the Earth Sciences: Manual of Remote Sensing, A.N. Rencz, ed,* American Society for Photogrammetry and Remote Sensing, pp. 3–58.

Clark, R.N., King, T.V. V, Klejwa, M., and Swayze, G.A. 1990. High spectral resolution reflectance spectroscopy of minerals. *J. Geophys. Res.*, 95(89):12653–12680.

Cornelis, W.M., Ronsyn, J., Meirvenne, M. Van, and Hartmann, R. 2001. Evaluation of pedotransfer functions for predicting the soil moisture retention curve. *Soil Sci. Soc. Am. J.*, 65:638–648.

Coûteaux, M.-M., Berg, B., and Rovira, P. 2003. Near infrared reflectance spectroscopy for determination of organic matter fractions including microbial biomass in coniferous forest soils. *Soil Biol. Biochem.*, 35(12):1587–1600.

Dahiya, B.Y.I.S., Dahiya, D.J., Kuhad, M.S., and Karwasra, S.P.S. 1988. Statistical equations for estimating field capacity, wilting point and available water capacity of soils from their saturation percentage. *J. Agric. Sci.*, 110(3):515–520.

Daniel, K.W., Tripathi, N.K., and Honda, K. 2003. Artificial neural network analysis of laboratory and in situ spectra for the estimation of macronutrients in soils of Lop Buri (Thailand. *Aust. J. Soil Res.*, 41(1):47–59.

Dunn, B., Batten, G., Beecher, H., and Ciavarella, S. 2002. The potential of near-infrared reflectance spectroscopy for soil analysis—a case study from the Riverine Plain of south-eastern Australia. *Anim. Prod. Sci.*, 01.

Fabre, S., Briottet, X., Lesaignoux, A., Avenue, B.P., Belin, E., and Cedex, F.-T. 2015. Estimation of soil moisture content from the spectral reflectance of bare soils in the 0.4–2.5 µm domain. *Sensors*, 15:3262–3281.

Farifteh, J., Tolpekin, V., Meer, F. Van Der, and Sukchan, S. 2010. Salinity modelling by inverted Gaussian parameters of soil reflectance spectra. *Int. J. Remote Sens.*, 31(12):3195–3210.

Fernández Pierna, J.A., Abbas, O., Baeten, V., and Dardenne, P. 2009. A Backward Variable Selection method for PLS regression (BVSPLS. *Anal. Chim. Acta*, 642:89–93.

Fox, G.A. and Metla, R. 2005. Soil property analysis using principal components analysis, soil line, and regression models. *Soil Sci. Soc. Am. J.*, 69(6):1782–1788.

Galvão, R.K.H., Araujo, M.C.U., José, G.E., Pontes, M.J.C., Silva, E.C., and Saldanha, T.C.B. 2005. A method for calibration and validation subset partitioning. *Talanta*, 67(4):736–40.

Gee, G.W. and Bauder, J.W. 1986. Particle-size analysis., *in Methods of Soil Analysis. Part 1. Physical and Mineralogical Methods, SSSA Book Series, A. Klute, ed*, Soil Science Society of America, American Society of Agronomy, Madison, WI, pp. 383–412.

Genot, V., Colinet, G., Laurent, B., Dominique, V., and Dardenne, P. 2011. Near infrared reflectance spectroscopy for estimating soil characteristics valuable in the diagnosis of soil fertility. *J. Near Infrared Spectrosc.*, 19:117–138.

Gholizadeh, A., Borůvka, L., Saberioon, M., and Vašát, R. 2016. A memory-based learning approach as compared to other data mining algorithms for the prediction of soil texture using diffuse reflectance spectra. *Remote Sens.*, 8(4):341.

Gholizadeh, A., Borůvka, L., Saberioon, M., and Vašát, R. 2013. Visible, near-infrared, and mid-infrared spectroscopy applications for soil assessment with emphasis on soil organic matter content and quality: State-of-the-art and key issues. *Appl. Spectrosc.*, 67(12):1349–1362.

Givi, J., Prasher, S.O., and Patel, R.M. 2004. Evaluation of pedotransfer functions in predicting the soil water contents at field capacity and wilting point. *Agric. Water Manag.*, 70(2):83–96.

Goldshleger, N., Chudnovsky, A., and Ben-Dor, E. 2012. Using reflectance spectroscopy and artificial neural network to assess water infiltration rate into the soil profile. *Appl. Environ. Soil Sci.*, 2012:1–9.

Grewal, K.S., Buchan, G.D., and Tonkin, P.J. 1990. Estimation of field capacity and wilting point of some New Zealand soils from their saturation percentages. *New Zeal. J. Crop Hortic. Sci. ISSN*, 18(4):241–246.

Gupta, M., Srivastava, P.K., and Islam, T. 2016. Integrative use of near-surface satellite soil moisture and precipitation for estimation of improved irrigation scheduling parameters., *in Satellite Soil Moisture Retrieval,* Elsevier Inc., pp. 271–288.

Gupta, S.C. and Larson, W.E. 1979. Estimating soil water retention characteristics from particle size distribution, organic matter percent, and bulk density., *in Water Resources Research,* pp. 1633–1635.

Haubrock, S. -N., Chabrillat, S., Lemmnitz, C., and Kaufmann, H. 2008. Surface soil moisture quantification models from reflectance data under field conditions. *Int. J. Remote Sens.*, 29(1):3–29.

He, Y., Song, H., Pereira, A.G., and Gómez, A.H. 2005. A new approach to predict N, P, K and OM content in a loamy mixed soil by using near infrared reflectance. 859–867.

Hill, J., Udelhoven, T., Vohland, M., and Stevens, A. 2010. The use of laboratory spectroscopy and optical remote sensing for estimating soil properties., *in Precision Crop Protection-the Challenge and Use of Heterogeneity, E.-C. Oerke, R. Gerhards, G. Menz, A.S. Richard, eds,* Springer, Dordrecht, pp. 67–86.

Hong-Yan, R., Da-fang, Z., Singh, A.N., Jian-jun, P., Dong-sheng, Q., and Run-He, S. 2009. Estimation of As and Cu contamination in agricultural soils around a mining area by reflectance spectroscopy : A case study. *Pedosphere*, 19(6):719–726.

Islam, K., Singh, B., and Bratney, A.M. 2006. Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. 1101–1114.

Israelsen, O.W. and Hansen, V.E. 1962. Irrigation Principles and Practices, 3rd ed. John Wiley & Sons, Inc, New York.

Janik, L.J., Forrester, S.T., and Rawson, A. 2009. The prediction of soil chemical and physical properties from mid-infrared spectroscopy and combined partial least-squares regression and neural networks (PLS-NN) analysis. *Chemom. Intell. Lab. Syst.,* 97(2):179–188.

Ji, W., Li, S., Chen, S., Shi, Z., Viscarra Rossel, R.A., and Mouazen, A.M. 2016. Prediction of soil attributes using the Chinese soil spectral library and standardized spectra recorded at field conditions. *Soil Tillage Res.,* 155:492–500.

Ji, W., Viscarra Rossel, R.A., and Shi, Z. 2015. Accounting for the effects of water and the environment on proximally sensed vis-NIR soil spectra and their calibrations. *Eur. J. Soil Sci.,* 66(3):555–565.

Johnston, C.T. and Aochi, Y.O. 1996. Fourier transform infrared and Raman spectroscopy., *in Methods of Soil Analysis. Part 3. Chemical Methods, D.L. Sparks, ed,* SSSA and ASA, Madison, WI, USA, pp. 269–321.

Karkanis, P.G. 1983. Determining field capacity and wilting point using soil saturation by capillary rise. *Can. Agric. Eng.,* 25(1):19–21.

Kawamura, K., Tsujimoto, Y., Rabenarivo, M., and Asai, H. 2017. Vis-NIR spectroscopy and PLS regression with waveband selection for estimating the total C and N of paddy soils in Madagascar.

Kennard, R.W. and Stone, L.A. 1969. Computer aided design of experiments. *Technometrics,* 11(1):137–148.

Kinoshita, R., Moebius-Clune, B.N., van Es, H.M., Hively, W.D., and Bilgilis, A.V. 2012. Strategies for soil quality assessment using visible and near-infrared reflectance spectroscopy in a western Kenya chronosequence. *Soil Sci. Soc. Am. J.,* 76(5):1776–1788.

Kirkham, M.B. 2014. Field capacity, wilting point, available water, and the nonlimiting water range., *in Principles of Soil and Plant Water Relations,* USA, pp. 153–170.

Kuang, B. and Mouazen, A.M. 2011. Calibration of visible and near infrared spectroscopy for soil analysis at the field scale on three European farms. August):629–636.

Lacerda, M., Demattê, J., Sato, M., Fongaro, C., Gallo, B., and Souza, A. 2016. Tropical texture determination by proximal sensing using a regional spectral library and its relationship with soil classification. *Remote Sens.*, 8(9):701.

Ladoni, M., Ali, Æ.H., Sayed, B.Æ., and Alavipanah, K. 2010. Estimating soil organic carbon from soil reflectance : a review. 82–99.

Lagacherie, P., Baret, F., Feret, J., Madeira, J., and Robbez-masson, J.M. 2008. Estimation of soil clay and calcium carbonate using laboratory , field and airborne hyperspectral measurements. *Remote Sens. Environ.*, 112:825–835.

Lehane, J.J. and Staple, W.J. 1951. Desiccator method for determining permanent wilting percentages of soils. *Soil Sci.*, 72(6):429–434.

Li, H., Liang, Y., Xu, Q., and Cao, D. 2009. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. *Anal. Chim. Acta*, 648(1):77–84.

Li, S., Shi, Z., Chen, S., Ji, W., Zhou, L., Yu, W., and Webster, R. 2015. In situ measurements of organic carbon in soil profiles using vis-NIR spectroscopy on the Qinghai − Tibet Plateau. *Environ. Sci. Technol.*, 49(8):4980–4987.

Ludwig, B., Khanna, P.K., Bauhus, J., and Hopmans, P. 2002. Near infrared spectroscopy of forest soils to determine chemical and biological properties related to soil sustainability. 171:121–132.

Malley, D.F. and Martin, P.D. 2003. The use of near-infrared spectroscopy for soil analysis., *in Tools for Nutrient and Pollutant Management: Application to Agriculture and Environmental Quality, L. Currie, J. Hanly, eds*, Fertilizer and Lime Research Centre,Massey University, Palmerston North, New Zealand, pp. 371–404.

Malley, D.F., Martin, P.D., and Ben-Dor, E. 2004. Application in analysis of soils., *in Near-Infrared Spectroscopy in Agriculture, C.A. Roberts, J. Workman, Jr, J.B. Reeves, III, eds*, Agronomy monograph 44,ASA, CSSA, and SSSA, Madison, WI, USA, pp. 729–784.

Martens, H., Jensen, S.A., and Geladi, P. 1983. Multivariate linearity transformation for near-infrared reflectance spectrometry., *in Proceedings of the Nordic Symposium on Applied Statistics,* pp. 205–234.

Martin, P.D., Malley, D.F., Manning, G., and Fuller, L. 2002. Determination of soil organic carbon and nitrogen at the field level using near-infrared spectroscopy. 7.

Mehmood, T., Liland, K.H., Snipen, L., and Sæbø, S. 2012. A review of variable selection methods in Partial Least Squares Regression. *Chemom. Intell. Lab. Syst.,* 118:62–69.

Morellos, A., Pantazi, X.-E., Moshou, D., Alexandridis, T., Whetton, R., Tziotzios, G., Wiebensohn, J., Bill, R., and Mouazen, A.M. 2016. Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. *Biosyst. Eng.,* 1–13.

Mouazen, A.M., Kuang, B., De Baerdemaeker, J., and Ramon, H. 2010. Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. *Geoderma,* 158(1–2):23–31.

Mouazen, A.M., Maleki, M.R., De Baerdemaeker, J., and Ramon, H. 2007. On-line measurement of some selected soil properties using a VIS–NIR sensor. *Soil Tillage Res.,* 93(1):13–27.

Nadler, B. and Coifman, R.R. 2005. The prediction error in CLS and PLS: the importance of feature selection prior to multivariate calibration. *J. Chemom.,* 19(2):107–118.

Nawar, S. and Mouazen, A.M. 2017. Predictive performance of mobile vis-near infrared spectroscopy for key soil properties at different geographical scales by using spiking and data mining techniques. *Catena,* 151:118–129.

Nocita, M., Stevens, A., Noon, C., and Van Wesemael, B. 2013. Prediction of soil organic carbon for different levels of soil moisture using Vis-NIR spectroscopy. *Geoderma,* 199:37–42.

Nouri, M., Gomez, C., Gorretta, N., and Roger, J.M. 2017. Clay content mapping from airborne hyperspectral Vis-NIR data by transferring a laboratory regression model. *Geoderma,* 298:54–66.

Nowkandeh, S.M., Noroozi, A.A., and Homaee, M. 2018. Estimating soil organic matter content from Hyperion reflectance images using PLSR, PCR, MinR and SWR models in semi-arid regions of Iran. *Environ. Dev.,* 25:23–32.

Pachepsky, Y., Timlin, D., and Varallyay, G. 1996. Artificial neural networks to estimate soil water retention from easily measurable data. *Soil Sci. Soc. Am. J.,* 60:727–733.

Pittaki-Chrysodonta, Z., Moldrup, P., Knadel, M., Iversen, B. V, Hermansen, C., Greve, M.H., and Jonge, L.W. de (2018. Predicting the Campbell soil water retention function : Comparing visible–near-infrared spectroscopy with classical pedotransfer function. *Vadose Zo. J.,* 17:170169.

Post, J. and Noble, P. 1993. The near-infrared combination band frequencies of dioctahedral smectites, micas, and illites. *Clays Clay Miner.,* 41(6):639–644.

Raj, A., Chakraborty, S., Duda, B.M., Weindorf, D.C., Li, B., Roy, S., Sarathjith, M.C., Das, B.S., and Paulette, L. 2018. Soil mapping via diffuse reflectance spectroscopy based on variable indicators: An ordered predictor selection approach. *Geoderma,* 314:146–159.

Rajer-Kanduč, K., Zupan, J., and Majcen, N. 2003. Separation of data on the training and test set for modelling: a case study for modelling of five colour properties of a white pigment. *Chemom. Intell. Lab. Syst.,* 65(2):221–229.

Reeves III, J.. 2010. Near- versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis : Where are we and what needs to be done ? *Geoderma,* 158(1–2):3–14.

Richards, L.A. and Fireman, M. 1943. Pressure-plate apparatus for measuring moisture sorption and transmission by soils. *Soil Sci.,* 56(6):395–404.

Richards, L.A. and Weaver, L.R. 1943. Fifteen-atmosphere percentage as related to the permanent wilting percentage. *Soil Sci.,* 56(5):331–340.

Rinnan, Å., Berg, F. Van Den, and Engelsen, S.B. 2009. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends Anal. Chem.,* 28(10):1201–1222.

Ritchie, J.T. 1981. Soil water availability., *in Plant and Soil,* Dr. W. Junk Publishers, The Netherlands, pp. 327–338.

Saeys, W., Mouazen, A.M., and Ramon, H. 2005. Potential for onsite and online analysis of pig manure using visible and near infrared reflectance spectroscopy. *Biosyst. Eng.,* 91(4):393–402.

Salter, P.J. 1967. Methods of determining the moisture characteristics of soils. *Exp. Agric.,* 3(2):163–173.

Salter, P.J. and Williams, J.B. 1965a. The influence of texture on the moisture characteristics of soils: I. a Critical comparison of techniques for determining the available-water capacity and moisture characteristic curve of a soil. *J. Soil Sci.,* 16(1):1–15.

Salter, P.J. and Williams, J.B. 1965b. The influence of texture on the moisture characteristics of soils: II. Available-water capacity and moisture release characteristics. *J. Soil Sci.,* 16(2):310–317.

Santra, P., Sahoo, R.N., Das, B.S., Samal, R.N., Pattanaik, A.K., and Gupta, V.K. 2009. Estimation of soil hydraulic properties using proximal spectral reflectance in visible, near-infrared, and shortwave-infrared (VIS–NIR–SWIR) region. *Geoderma,* 152(3–4):338–349.

Santra, P., Singh, R., Sarathjith, M.C., Panwar, N.R., Varghese, P., and Das, B.S. 2015. Reflectance spectroscopic approach for estimation of soil properties in hot arid western Rajasthan, India. *Environ. Earth Sci.,* 74(5):4233–4245.

Sarathjith, M.C., Das, B.S., Vasava, H.B., Mohanty, B., Sahadevan, A.S., Wani, S.P., and Sahrawat, K.L. 2014. Diffuse reflectance spectroscopic approach for the characterization of soil aggregate size distribution. *Soil Sci. Soc. Am. J.,* 78(2):369–376.

Sarathjith, M.C., Das, B.S., Wani, S.P., and Sahrawat, K.L. 2016a. Variable indicators for optimum wavelength selection in diffuse reflectance

spectroscopy of soils. *Geoderma*, 267:1–9.

Sarathjith, M.C., Das, B.S., Wani, S.P., Sahrawat, K.L., and Gupta, A. 2016b. Comparison of data mining approaches for estimating soil nutrient contents using diffuse reflectance spectroscopy. *Curr. Sci.*, 110(6):1031–1037.

Schaap, M.G., Leij, F.J., and Genuchten, M.T. Van (2001. Rosetta : a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. 251:163–176.

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

Shi, T., Liu, H., Chen, Y., Wang, J., and Wu, G. 2016. Estimation of arsenic in agricultural soils using hyperspectral vegetation indices of rice. *J. Hazard. Mater.*, 308:243–252.

Shrestha, R. 2006. Relating soil electrical conductivity to remote sensing and other soil properties for assessing soil salinity in northeast Thailand. *L. Degrad. Dev.*, 17(6):677–689.

Singh, K., Murphy, B., and Marchant, B. 2013. Towards cost-effective estimation of soil carbon stocks at the field scale. *Soil Res.*, 01:485–490.

Song, L., Langfelder, P., and Horvath, S. 2012. Comparison of co-expression measures: mutual information, correlation, and model based indices. *BMC Bioinformatics*, 13:328.

Soriano-Disla, J.M., Janik, L.J., Viscarra Rossel, R.A., Macdonald, L.M., and McLaughlin, M.J. 2014. The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. *Appl. Spectrosc. Rev.*, 49(2):139–186.

Stenberg, B. 2010. Effects of soil sample pretreatments and standardised rewetting as interacted with sand classes on Vis-NIR predictions of clay and soil organic carbon. *Geoderma*, 158(1–2):15–22.

Stenberg, B., Rossel, R.A.V., Mouazen, A.M., and Wetterlind, J. 2010. Visible and near infrared spectroscopy in soil science., *in Advances in Agronomy, Donald L. Sparks, ed*, Academic Press, Burlington, pp. 163–215.

Stuart, B.H. 2004. Infrared Spectroscopy: Fundamentals and Applications, *Methods, Analytical Techniques in the Sciences*. John Wiley & Sons, Ltd, Chichester, UK.

Taylor, S.A. and Ashcroft, G.L. 1972. Physical edaphology. The physics of irrigated and nonirrigated soils. W.H. Freeman, San Francisco.

Teófilo, R.F., Martins, J.P.A., and Ferreira, M.M.C. 2009. Sorting variables by using informative vectors as a strategy for feature selection in multivariate regression. *J. Chemom.,* 23(1):32–48.

Tietje, O. and Tapkenhinrichs, M. 1993. Evaluation of pedo-transfer functions. *Soil Sci. Soc. Am. J.,* 57:1088–1095.

Tyler, S.W. and Wheatcraft, S.W. 1989. Application of fractal mathematics to soil water retention estimation. *Soil Sci. Soc. Am. J.,* 53(4):987–996.

Vasques, G.M., Grunwald, S., and Harris, W.G. 2009a. Spectroscopic models of soil organic carbon in Florida, USA. *J. Environ. Qual.,* 39(3):923–34.

Vasques, G.M., Grunwald, S., and Sickman, J.O. 2009b. Modeling of soil organic carbon fractions using visible–near-infrared spectroscopy. *Soil Sci. Soc. Am. J.,* 73(1):176–184.

Vasques, G.M., Grunwald, S., and Sickman, J.O. 2008. Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. *Geoderma,* 146(1–2):14–25.

Vaudour, E., Gilliot, J.M., Bel, L., Lefevre, J., and Chehdi, K. 2016. Regional prediction of soil organic carbon content over temperate croplands using visible near-infrared airborne hyperspectral imagery and synchronous field spectra. *Int. J. Appl. Earth Obs. Geoinf.,* 49:24–38.

Veihmeyer, F.J. and Henderickson, A.H. 1949. Methods of measuring field capacity and permanent wilting percentage of soils. *Soil Sci.*, 68(1):75–94.

Vendrame, P.R.S., Marchao, R.L., Brunet, D., and Becquer, T. 2012. The potential of NIR spectroscopy to predict soil texture and mineralogy in Cerrado Latosols. *Eur. J. Soil Sci.*, 63:743–753.

Viscarra Rossel, R.A. 2008. ParLeS : Software for chemometric analysis of spectroscopic data. *Chemom. Intell. Lab. Syst.*, 90:72–83.

Viscarra Rossel, R.A. 2007. Robust modelling of soil diffuse reflectance spectra by bagging- partial least squares regression. *J. Near Infrared Spectrosc.*, 15(2):39–47.

Viscarra Rossel, R.A. and Behrens, T. 2010. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, 158(1–2):46–54.

Viscarra Rossel, R.A. and Lark, R.M. 2009. Improved analysis and modelling of soil diffuse reflectance spectra using wavelets. *Eur. J. Soil Sci.*, 60(3):453–464.

Viscarra Rossel, R.A., McGlynn, R.N., and McBratney, A.B. 2006a. Determining the composition of mineral-organic mixes using UV–vis–NIR diffuse reflectance spectroscopy. *Geoderma*, 137(1–2):70–82.

Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., and Skjemstad, J.O. 2006b. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma*, 131(1–2):59–75.

Viscarra Rossel, R.A. and Webster, R. 2012. Predicting soil properties from the Australian soil visible-near infrared spectroscopic database. *Eur. J. Soil Sci.*, 63(6):848–860.

Vohland, M., Ludwig, M., Thiele-Bruhn, S., and Ludwig, B. 2014. Determination of soil properties with visible to near- and mid-infrared spectroscopy: Effects of spectral variable selection. *Geoderma*, 223–225:88–96.

Waiser, T.H., Morgan, C.L.S., Brown, D.J., and Hallmark, C.T. 2007. In situ characterization of soil clay content with visible near-infrared diffuse reflectance spectroscopy. *Soil Sci. Soc. Am. J.*, 71(2):389–396.

Walkley, A.J. and Black, I.A. 1934. An examination of Degtjareff method for determining soil organic matter and a proposed modification of the chromic acid titration method. *Soil Sci.*, 37:29–38.

Wartini, N., Minasny, B., Malone, B.P., Sarathjith, M.C., and Das, B.S. 2019. Optimizing wavelength selection by using informative vectors for parsimonious infrared spectra modelling. *Comput. Electron. Agric.*,

158:201–210.

Weidong, L., Baret, F., Gu, X., Zhang, B., Tong, Q., and Zheng, L. 2003. Evaluation of methods for soil surface moisture estimation from reflectance data. *Int. J. Remote Sens.*, 24(10):2069–2083.

Weidong, L., Baret, F., Xingfa, G., Qingxi, T., Lanfen, Z., and Bing, Z. 2002. Relating soil surface moisture to reflectance. *Remote Sens. Environ.*, 81(2–3):238–246.

Wetterlind, J. and Stenberg, B. 2010. Near-infrared spectroscopy for within-field soil characterization : small local calibrations compared with national libraries spiked with local samples. *Eur. J. Soil Sci.*, 61:823–843.

Wilcox, R. 2005. Introduction to robust estimation and hypothesis testing, 2nd ed. Elsevier, Amsterdam.

Williams, P. and Norris, K. 1987. Near-infrared technology in the agricultural and food industries. American Association of Cereal Chemists, Inc., St. Paul, MN.

Wold, S., Sjöström, M., and Eriksson, L. 2001. PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.*, 109–130.

Workman, J. and Shenk, J. 2004. Understanding and using the near-infrared spectrum as an analytical method., *in Near Infrared Reflectance Spectroscopy in Agriculture, C.A. Roberts, J.W. Jr., J.B. Reeves, eds*, Agronomy monograph 44,ASA, CSSA, and SSSA, Madison, WI, USA, pp. 3–10.

Xiaobo, Z., Jiewen, Z., Povey, M.J.W., Holmes, M., and Hanpin, M. 2010. Variables selection methods in near-infrared spectroscopy. *Anal. Chim. Acta*, 667(1–2):14–32.

Yun, Y.H., Wang, W.T., Tan, M.L., Liang, Y.Z., Li, H.D., Cao, D.S., Lu, H.M., and Xu, Q.S. 2014. A strategy that iteratively retains informative variables for selecting optimal variable subset in multivariate calibration. *Anal. Chim. Acta*, 807:36–43.

APPENDICES

APPENDIX A

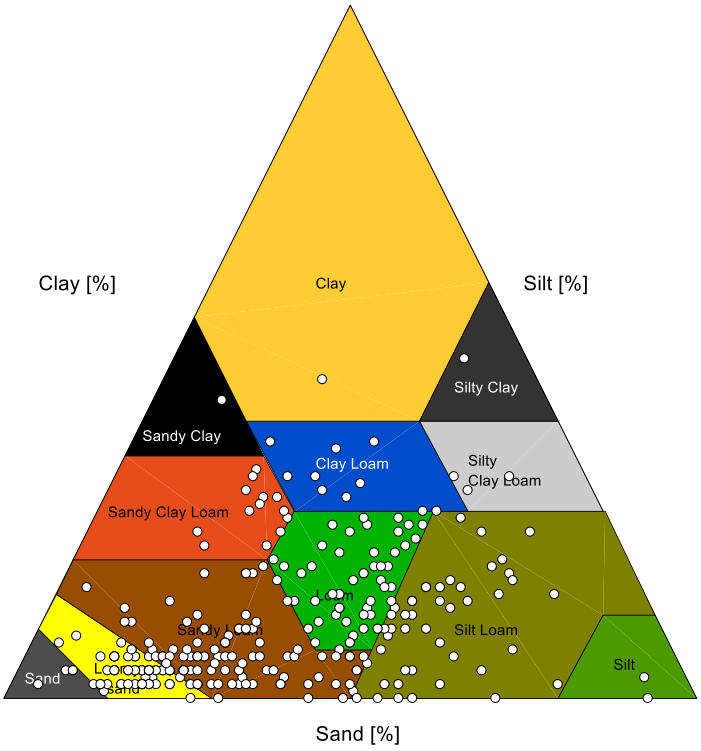

Fig. A1 USDA textural class of soil samples in the database

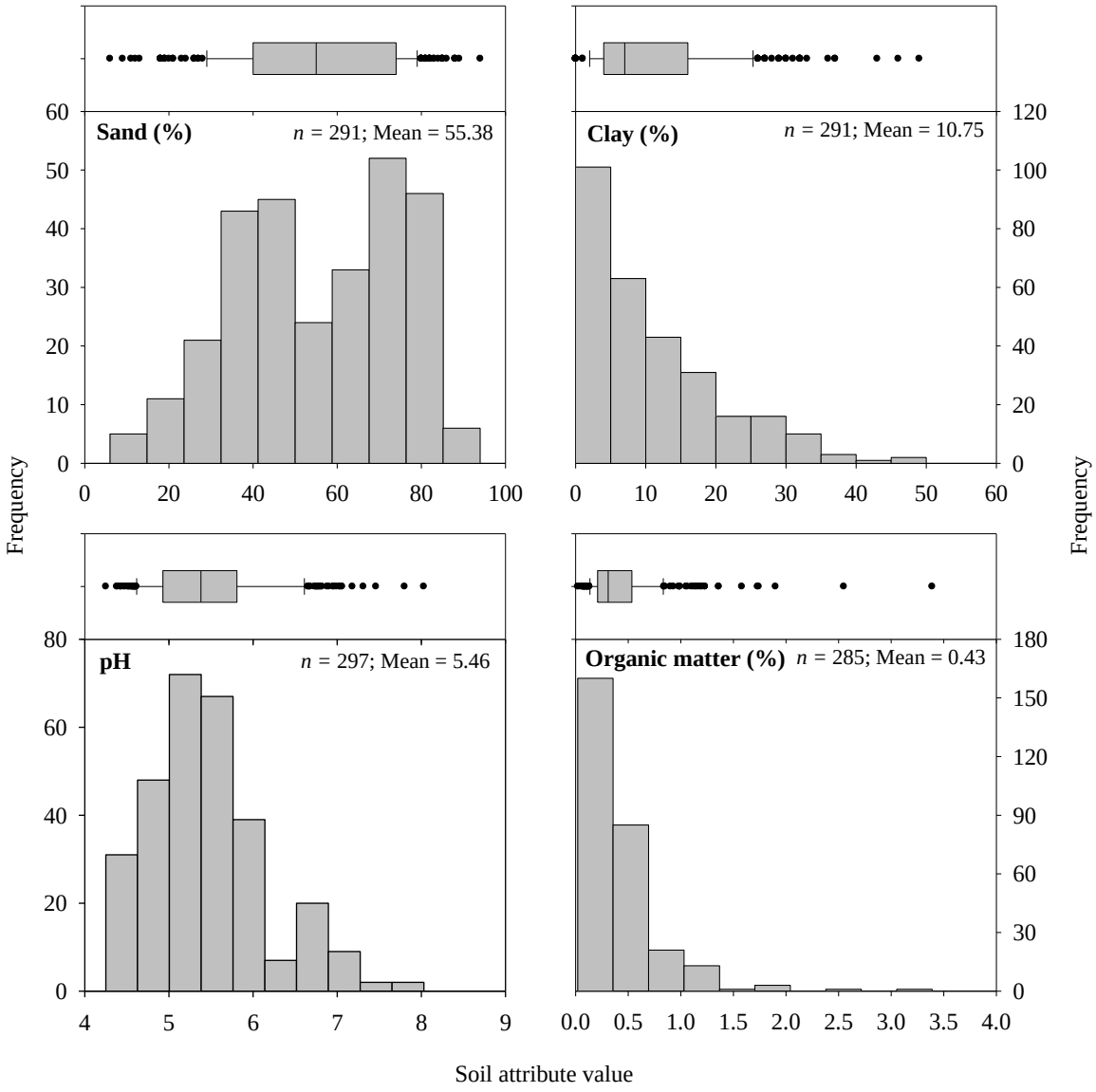

Fig. A2 Histograms and box plots of soil texture, pH and organic matter

APPENDIX B

Table B1 Regression statistics of the prediction of field capacity using different spectral pre-processing methods

Pre-process	LV	Calibration (*n*=220)		Validation (*n*=73)		
		R^2	RMSE	R^2	RMSE	RPD
R	12	0.72	4.74	0.68	4.98	1.77
R+SNV	16	0.80	4.03	0.74	4.43	1.99
R+MSC	15	0.78	4.18	0.75	4.38	2.02
R+DT	17	0.81	3.87	0.75	4.40	2.00
R+FD	8	0.77	4.29	0.74	4.44	1.99
R+SD	6	0.77	4.30	0.75	4.38	2.01
R+SNV+FD	9	0.77	4.27	0.73	4.59	1.92
R+MSC+FD	10	0.78	4.21	0.72	4.61	1.91
R+DT+FD	9	0.78	4.22	0.76	4.30	2.05
R+SNV+SD	7	0.78	4.20	0.73	4.55	1.94
R+MSC+SD	7	0.76	4.32	0.69	4.91	1.80
R+DT+SD	6	0.77	4.30	0.75	4.38	2.01
A	13	0.73	4.61	0.72	4.66	1.89
A+SNV	16	0.79	4.12	0.75	4.38	2.01
A+MSC	15	0.79	4.13	0.75	4.34	2.03
A+DT	12	0.72	4.74	0.68	4.98	1.77
A+FD	9	0.78	4.21	0.74	4.50	1.96
A+SD	6	0.75	4.41	0.73	4.59	1.92
A+SNV+FD	9	0.77	4.23	0.74	4.51	1.96
A+MSC+FD	10	0.78	4.18	0.73	4.52	1.95
A+DT+FD	9	0.79	4.04	0.74	4.49	1.96
A+SNV+SD	7	0.78	4.14	0.74	4.45	1.98
A+MSC+SD	7	0.77	4.28	0.70	4.81	1.84
A+DT+SD	6	0.75	4.41	0.73	4.59	1.92

n: number of soils, *LV*: number of latent variables, R^2: coefficient of determination, RMSE: root mean squared error, RPD: residual prediction deviation, *R*: reflectance, *SNV*: standard normal variate, *MSC*: multiplicative scatter correction, *DT*: detrending, *FD*: first derivative, *SD*: second derivative

Table B2 Regression statistics of the prediction of wilting point using different spectral pre-processing methods

Pre-process		Calibration (*n*=218)		Validation (*n*=72)		
	LV	R^2	RMSE	R^2	RMSE	RPD
R	18	0.80	2.15	0.71	2.48	1.86
R+SNV	18	0.80	2.15	0.65	2.70	1.71
R+MSC	17	0.79	2.20	0.61	2.85	1.62
R+DT	16	0.79	2.21	0.69	2.55	1.81
R+FD	10	0.80	2.13	0.69	2.56	1.80
R+SD	6	0.76	2.34	0.67	2.65	1.74
R+SNV+FD	11	0.80	2.12	0.71	2.45	1.88
R+MSC+FD	11	0.79	2.16	0.69	2.54	1.82
R+DT+FD	10	0.80	2.13	0.70	2.52	1.83
R+SNV+SD	8	0.79	2.20	0.64	2.74	1.68
R+MSC+SD	7	0.74	2.42	0.63	2.79	1.65
R+DT+SD	6	0.76	2.34	0.67	2.65	1.74
A	18	0.80	2.12	0.73	2.36	1.96
A+SNV	18	0.80	2.13	0.64	2.75	1.68
A+MSC	17	0.79	2.17	0.64	2.76	1.67
A+DT	18	0.80	2.15	0.71	2.48	1.86
A+FD	10	0.79	2.17	0.71	2.46	1.87
A+SD	7	0.79	2.19	0.66	2.67	1.73
A+SNV+FD	11	0.80	2.12	0.72	2.41	1.92
A+MSC+FD	11	0.80	2.15	0.70	2.49	1.85
A+DT+FD	10	0.80	2.13	0.73	2.39	1.93
A+SNV+SD	7	0.77	2.30	0.65	2.69	1.71
A+MSC+SD	7	0.75	2.37	0.65	2.70	1.71
A+DT+SD	7	0.79	2.19	0.66	2.67	1.73

n: number of soils, *LV*: number of latent variables, R^2: coefficient of determination, RMSE: root mean squared error, RPD: residual prediction deviation, *R*: reflectance, *SNV*: standard normal variate, *MSC*: multiplicative scatter correction, *DT*: detrending, *FD*: first derivative, *SD*: second derivative

APPENDIX C

Table C1 Optimum subset models of field capacity identified using ordered predictor selection approach

Variable indicator	NSV	Calibration (*n*=220, LV=6)		Validation (*n*=73)		
		R^2	RMSE	R^2	RMSE	RPD
β	586	0.77	4.26	0.75	4.39	2.01
VIP	521	0.76	4.37	0.75	4.42	2.00
SqRes	521	0.75	4.50	0.75	4.36	2.02
r	180	0.76	4.35	0.76	4.28	2.06
bicor	586	0.77	4.31	0.75	4.36	2.02
AMI	586	0.77	4.30	0.74	4.43	1.99
StN	586	0.76	4.34	0.75	4.38	2.01
CovProc	586	0.76	4.33	0.74	4.43	1.99
β-VIP	586	0.77	4.29	0.75	4.40	2.01
β-SqRes	521	0.77	4.29	0.75	4.37	2.02
VIP-SqRes	586	0.76	4.34	0.74	4.43	1.99
β-r	126	0.76	4.37	0.76	4.27	2.06
β-bicor	180	0.77	4.26	0.76	4.32	2.04
βAMI	521	0.77	4.28	0.75	4.38	2.01
β-StN	256	0.78	4.17	0.76	4.28	2.06
β-CovProc	227	0.77	4.31	0.75	4.35	2.03
VIP-r	288	0.75	4.43	0.75	4.36	2.02
VIP-bicor	324	0.75	4.43	0.75	4.37	2.02
VIP-AMI	586	0.77	4.32	0.74	4.44	1.99
VIP-StN	288	0.77	4.32	0.75	4.38	2.01
VIP-CovProc	586	0.76	4.34	0.74	4.44	1.98
SqRes-r	463	0.76	4.39	0.76	4.31	2.05
SqRes-bicor	411	0.75	4.42	0.76	4.31	2.05
SqRes-AMI	586	0.76	4.34	0.75	4.38	2.01
SqRes-StN	463	0.75	4.46	0.75	4.35	2.03
SqRes-CovProc	586	0.76	4.37	0.75	4.36	2.02
r-bicor	586	0.77	4.24	0.75	4.35	2.03
rAMI	586	0.77	4.31	0.74	4.44	1.99
r-StN	288	0.78	4.15	0.76	4.31	2.05
r-CovProc	586	0.76	4.33	0.75	4.39	2.01
bicor-AMI	586	0.76	4.34	0.74	4.46	1.98
bicor-StN	586	0.76	4.34	0.75	4.38	2.01
bicor-CovProc	586	0.76	4.34	0.75	4.37	2.02
AMI-StN	288	0.79	4.07	0.75	4.42	2.00
AMI-CovProc	586	0.77	4.29	0.75	4.41	2.00
StN-CovProc	463	0.75	4.49	0.75	4.39	2.01

n: number of samples; LV: number of latent variables; NSV: number of spectral variables, R^2: coefficient of determination; RMSE: root mean squared error; RPD: residual prediction deviation

Table C2 Optimum subset models of wilting point identified using ordered predictor selection approach

Variable indicator	NSV	Calibration (*n*=218, LV=6)		Validation (*n*=72)		
		R^2	RMSE	R^2	RMSE	RPD
β	586	0.76	2.32	0.67	2.64	1.75
VIP	586	0.76	2.35	0.66	2.66	1.73
SqRes	586	0.75	2.40	0.64	2.75	1.68
r	586	0.75	2.38	0.65	2.72	1.69
bicor	521	0.76	2.34	0.66	2.67	1.73
AMI	586	0.74	2.41	0.64	2.74	1.68
StN	586	0.75	2.38	0.69	2.54	1.82
CovProc	586	0.77	2.29	0.68	2.60	1.77
β-VIP	586	0.76	2.34	0.66	2.66	1.73
β-SqRes	521	0.76	2.35	0.66	2.66	1.73
VIP-SqRes	586	0.75	2.37	0.65	2.69	1.71
β-r	586	0.76	2.33	0.67	2.63	1.75
β-bicor	586	0.76	2.33	0.66	2.65	1.74
B-AMI	586	0.76	2.34	0.66	2.66	1.73
β-StN	324	0.77	2.30	0.69	2.56	1.80
β-CovProc	411	0.77	2.28	0.67	2.62	1.76
VIP-r	463	0.74	2.42	0.65	2.70	1.71
VIP-bicor	586	0.75	2.39	0.65	2.71	1.70
VIP-AMI	586	0.75	2.37	0.65	2.70	1.71
VIP-StN	411	0.75	2.37	0.67	2.64	1.75
VIP-CovProc	586	0.76	2.34	0.66	2.66	1.74
SqRes-r	586	0.75	2.38	0.65	2.72	1.69
SqRes-bicor	586	0.75	2.40	0.65	2.71	1.70
SqRes-AMI	586	0.75	2.39	0.65	2.72	1.70
SqRes-StN	324	0.71	2.57	0.68	2.58	1.79
SqRes-CovProc	586	0.75	2.39	0.65	2.69	1.71
r-bicor	521	0.76	2.34	0.66	2.68	1.72
r-AMI	202	0.73	2.48	0.65	2.69	1.71
r-StN	586	0.76	2.36	0.66	2.66	1.73
r-CovProc	521	0.75	2.38	0.66	2.68	1.72
bicorAMI	324	0.74	2.45	0.66	2.66	1.74
bicor-StN	521	0.76	2.34	0.66	2.67	1.73
Bicor-CovProc	324	0.75	2.39	0.67	2.64	1.75
AMI-StN[*]	227	0.72	2.51	0.69	2.56	1.80
AMI-CovProc	463	0.75	2.39	0.66	2.67	1.73
StN-CovProc	463	0.75	2.38	0.70	2.53	1.82

n: number of samples; LV: number of latent variables; NSV: number of spectral variables, R^2: coefficient of determination; RMSE: root mean squared error; RPD: residual prediction deviation

* LV=5.

ESTIMATION OF SOIL MOISTURE INDICES USING DIFFUSE REFLECTANCE SPECTROSCOPY

by

Sarathjith M.C.
(2017-18-003)

ABSTRACT OF THE THESIS REPORT

Submitted in partial fulfillment of the

requirement for the degree of

Master of Technology

In

Agricultural Engineering

Faculty of Agricultural Engineering and Technology

Kerala Agricultural University

Department of Irrigation and Drainage Engineering
KELAPPAJI COLLEGE OF AGRICULTURAL ENGINEERING AND TECHNOLOGY
TAVANUR - 679 573, MALAPPURAM
KERALA, INDIA
2019

ABSTRACT

Rapid and reliable estimation of soil moisture constants namely, field capacity (FC) and wilting point (WP) is significant for scientific irrigation scheduling. The conventional methods for their estimation are cumbersome, time consuming and not suitable for their estimation at different space and time domains. An alternative would be the use of diffuse reflectance spectroscopy (DRS) for which the development of calibration functions that link the soil attributes with spectral signature is a major pre-requisite. In this study, the utility of spectral index, feature projection of full-spectrum and variable selection approaches namely, normalized difference reflectance index (NDRI), partial least squares regression (PLSR) and ordered predictor selection (OPS), respectively to build accurate and less complex calibration functions was evaluated. The performance of calibration functions were judged in terms residual prediction deviation (RPD) criteria. The NDRI based calibration functions developed in this study do not comply with the minimum accuracy level (RPD<1.4) expected from DRS analysis. In contrast, both full-spectrum based PLSR and OPS approaches yielded calibration functions which were capable for accurate (RPD>2.0) and moderate (1.4<RPD>2.0) estimation of FC and WP, respectively. Specifically, the full-spectrum based calibration function developed using second derivative of reflectance was found to be the best for both FC (RPD=2.01) and WP (RPD=1.74). The OPS approach in conjunction with variable indicators namely, combination of regression & correlation coefficient (β-r) and combination of adjacency values of mutual information & signal-to-noise vector (*AMI-StN*) yielded best calibration functions in case of FC and WP, respectively. The calibration functions so developed consisted of only 19.09% (FC) and 34.39% (WP) of total number of spectral variables as that in full-spectrum. Thus, the result of the study advocate the use of OPS approach to develop simple and parsimonious calibration functions to estimate FC and WP from spectral signature of soil.