Remote Sensing of Soil Physicochemical Properties with Sentinel-2A and ASD Field Spec-4 using Multiple Linear Regression Analysis and Artificial Neural Network

By

Sania Parveen

(2018-NUST-MS-GIS-276473)

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Remote Sensing and GIS

Institute of Geographical Information Systems
School of Civil and Environmental Engineering
National University of Sciences and Technology
Islamabad, Pakistan

July 2022

CERTIFICATE

Certified that the contents and form of thesis entitled **"Remote Sensing of Soil Physicochemical Properties with Sentinel-2A and ASD Field Spec-4 using Multiple Linear Regression Analysis and Artificial Neural Network"** submitted by Ms. Sania Parveen (Registration No. 2018-NUST-MS-RS&GIS-276473) has been found satisfactory for the requirements of Master of Science degree in Geographical Information Systems and Remote Sensing.

Supervisor: _____

Dr. Javed Iqbal

HoD, IGIS, NUST

Member: _____

Dr. Ejaz Hussain

Associate Professor

IGIS - SCEE, NUST

Member: _____

Dr. Muhammad Hassan Ali Baig

Associate Professor

IGEO, AAUR

Member: _____

Mr. Junaid Aziz Khan

Lecturer

IGIS - SCEE, NUST

DEDICATION

To

My Beloved Father & Mother

Thanks for their love, care, and motivation all the way since the start of my studies and to my sweet sister and brother including all those who encouraged me and prayed for me for the completion of this thesis.

ACADEMIC THESIS: DECLARATION OF AUTHORSHIP

I, Sania Parveen declare that this thesis and the work presented in it are my own and have been generated by me as the result of my own original research.

"Remote Sensing of Soil Physicochemical Properties with Sentinel-2A and ASD Field Spec-4 using Multiple Linear Regression Analysis and Artificial Neural Network"

I confirm that:

1. This work was done wholly by me in candidature for an MS research degree at the National University of Sciences and Technology, Islamabad.
2. Wherever I have consulted the published work of others, it has been clearly attributed.
3. Wherever I have quoted from the work of others, the source has been always cited.
4. I have acknowledged all main sources of help.
5. Where the work of thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.
6. None of this work has been published before submission. This work is not plagiarized under the HEC plagiarism policy.

Signed: ……

Date: ……...

ACKNOWLEDGEMENTS

All praises to Allah Almighty, and blessings upon Prophet Muhammad (peace be upon him). I owe my sincere gratitude and pleasure to my respected research supervisor Dr. Javed Iqbal for his devoted guidance and encouraging attitude throughout the research. I am also grateful to guidance and examination committee members for their coordination. I am thankful to IGIS-NUST for the research fund granted for completion of all research activities. This research would have not been possible without the sincere cooperation and consistent support of IGIS NUST, which supported me in all activities. I am really thankful to Sir Junaid Aziz khan and Sir Naqash Taj Abbasi who helped me out in the collection of soil spectral signatures and to all the staff that was there in assisting. I really want to pay my sincere gratitude to Soil Physical lab of Soil Science department of Arid Agriculture University, Rawalpindi to perform soil texture analysis there and Waste Water Lab of IESE, NUST who allowed me to perform soil chemical analysis there. It is a pleasure to pay tribute to all my sincere friends and family members especially my father who helped me in collection of soil samples and stood by my side in each and every situation, my mother, my beloved sister and brother for their sincere attitude, support, patience, and prayers that really helped to successfully complete my tasks.

Sania Parveen

Table of Contents

List of figures

List of Tables

Abbreviations

ABSTRACT

In site-specific agriculture, mapping fine-scale spatial changes in soil characteristics are critical. Traditional soil analysis methods for mapping soil properties on a large scale are expensive, time-taking and not feasible. The current research looks at how remote sensing (RS) and geographic information system (GIS) approaches can be used to investigate the spatial variability of surface soil properties. An average of 51 samples were taken from tehsil Talagang of district Chakwal, Punjab, Pakistan. Lab analysis were carried out to determine soil texture and soil chemical properties. The objectives were to generate the predictive multiple linear regression modeling for soil physicochemical properties using Sentinel-2 and ASD Field Spec 4 data. The MLR has shown a significant ($p<0.05$) relationship of band-2, band-7, band-8 and band-8A with sand (%) ($R^2 = 0.19$), OM (%) showed a significant ($p<0.05$) relationship with band-2, band-4, band-5 and band-11 with an R^2 value of 0.23 and phosphorus obtained an R^2 value of 0.22 while other properties did not show significant results. In the same way, OLS had obtained R^2 value 0.11 for sand, K with R^2 value of 0.19 and P had obtained R^2 value of 0.10. Hence, both the techniques have not obtained significant results. Random Forest Regression has performed better than these for all soil properties. ASD data was modeled using SMLR and PLSR with sand properties. SMLR has sand (%) R^2 value of 0.85, silt (%) of 0.71 and clay (%) 0.51 while with PLSR sand (%) has R^2 value of 0.90, silt (%) 0.89 and clay (%) 0.83. Hence, PLSR performed better than SMLR in the case of hyperspectral data. Interpolation techniques were also used to predict soil properties, among which IDW has performed best. In conclusion, the DRS can be successfully used in the detection of soil properties as compared to multispectral data. Further, hyperspectral imagery can be used for most accurate results. Soil spectral libraries can be created using large amount of samples from the study area.

CHAPTER 1

INTRODUCTION

Global food security challenge (Gebbers & Adamchuk, 2010), due to increasing growth of urban areas (Su et al., 2011), degradation of soil (Lerch et al., 2005), and the agro-ecological balance is out of whack (E. Birch et al., 2011) has led to significant importance of precision agriculture in 21st century. An effective solution to combat this global challenge is the integration of Global Positioning System (GPS), Remote Sensing (RS) (Khanal et al., 2017), hyper spectral RS, field sensors, unmanned aerial vehicles (UAV) (Tokekar et al., 2016), automated machinery, and Internet of Things (Khanna & Kaur, 2019) through PA (Precision Agriculture). The scientists have reduced the cost and increased (Schimmelpfennig & Ebel, 2016) production through the ideal use of fertilizers using Variable Rate Technology (VRT) (Fabiani et al., 2020). The concept of precision agriculture is based on the knowledge of soil physical and mechanical properties including their spatial variability (Hanquet et al., 2004). Precision agriculture's (PA) performance is relies heavily on an efficient and precise method for determining soil properties in the field. Farmers need this information to calculate the right amount of inputs for the highest crop production and the least amount of environmental impact (Ge et al., 2011). A complex interaction of seed, soil, water, and agrochemicals results in the agricultural production system (including fertilizer). The moment has come to use all technological instruments at our disposal by combining information technology and agricultural science for more efficient and environmentally sustainable crop production. Precision agriculture (PA) integrates new information-age technologies with a mature agricultural industry. It was first introduced in the 1980s in the United States in response to calls to address agricultural and environmental issues such as fertilizer and pesticide pollution. It adjusts inputs and management procedures based on variables like

soil/landscape features, pest presence, and microclimate. It takes a more holistic approach to agriculture by combining data from numerous sources to influence agricultural production, logistics, marketing, finance, and personnel decisions. PA is a data-driven approach to farming made possible by several quickly evolving technology. Although these technologies have been widely used, it is still unclear how to fully utilize their potential in precision agriculture. Additional precision farming research should be driven by need. We can begin by identifying these critical requirements. To fulfill these demands, appropriate research projects be proposed (Afroj et al., 2016). Agriculture is responsible for roughly 13.5 percent of total worldwide anthropogenic Greenhouse Gas (GHG) emissions, making it a major contributor to climate change (Montzka et al., 2011). Precision agriculture (PA) methods, which employ a huge reservoir of Precision Agriculture Technologies (PATs) in agricultural field operations, could help to reduce GHG emissions in the following ways: (a) the improvement of soils' ability to function as carbon stock reserves through reduced tillage (Angers & Eriksen-Hamel, 2008) and nitrogen fertilization (Khan et al., 2007; Waldrop et al., 2004), (b) the reduction of fuel consumption through fewer tractor in-field operations (direct GHG reduction); and (c) the reduction of inputs for agricultural field operations (indirect GHG reduction) (Plant et al., 2000). PATs can help achieve agricultural sustainability by improving the efficiency of most agricultural activities by lowering or redistributing inputs to meet the crop's actual needs. The new Common Agricultural Policy (CAP) is expected to support more PATs as a way to boost or sustain production while lowering environmental consequences and, in particular, GHG emissions (Balafoutis et al., 2017). Crop yield variability is affected by soil physicochemical properties (Lee et al., 2012) and nutrients availability (Krishna, 2003).

1.1 Modelling Soil Physicochemical Properties using Remote Sensing

The collection of precise information about an object from a distance without coming into contact with it is referred to as remote sensing. Although RS has been used for a decade, its application in agriculture for the management of spatial variability is new. To monitor temporally dynamic plant and soil conditions, RS measures visible and invisible features of a field or set of fields and translates point measurements into spatial information (Afroj et al., 2016). Remote sensing (RS) is a new technology for acquiring PA data that has the benefits of cheap cost, speed, and relatively high resolution spatially. In-field soil property determination using RS has made significant progress (Ge et al., 2011).

Laboratory methods to model soil physicochemical properties are time consuming (Angelopoulou et al., 2020), successful inclusion of remote sensing data with laboratory measurements (Preissler & Loercher, 1995) while effective modeling of different attributes can be achieved by applying quantitative procedures (Al-Quraishi et al., 2019). Using a variety of remote sensing data and different approaches, scientists have utilized this methodology to model and map different attributes (Farifteh et al., 2006; Grunwald et al., 2015; King et al., 2005; Peng et al., 2019; Potopová et al., 2020; Sadeghi et al., 2015; B. Wang et al., 2018). The widely used remote sensing sensors for this purpose are ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) (Nawar et al., 2015), MODIS (Vågen et al., 2014), Sentinel (Vaudour et al., 2019), Landsat (Al-Quraishi et al., 2019; Zhang et al., 2019). Different spatial, spectral, and radiometric resolutions of sensors have an impact on the accuracy of modeling. A 185 km x 180 km Landsat-8 picture with a spatial resolution in the range of (30,100 meters), equipped with Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) has been used in a number of studies (Gorji et al., 2020; Srisomkiew et al., 2021;

Zhang et al., 2019). Researchers have also utilized Sentinel-2 to predict soil parameters as it has a 290 km swath width and has a wavelength range of (442.2, 2185.7 nm) with a spatial resolution of (10-60meters) (Lin et al., 2020; Vaudour et al., 2019; J. Wang et al., 2019; Zhou et al., 2020). Innovative researchers are comparing the effectiveness of several sensors for predicting soil qualities using a variety of methodologies and algorithms (Davis et al., 2019; Gorji et al., 2020; Silvero et al., 2021). Regression can be used to model information from RS data (Bhunia et al., 2019). In order to model spatial data, classical statistics have been employed extensively (Simmonds et al., 2014). However, as knowledge of GIS and RS has grown, spatial statistics have evolved over time to account for the shortcomings of classical statistics. Soil properties are widely modelled now a days using Artificial Neural Network (ANN) and Machine Learning (ML) techniques (Ahmad et al., 2010; Behrens et al., 2005; Pellegrini et al., 2021). Also they have been modelled by using spatial interpolation techniques such as Inverse Distance Weighting (IDW) (Sheng et al., 2021) and kriging (Mueller et al., 2004; Panday et al., 2018; Qiao et al., 2018; Robinson & Metternicht, 2006).

1.2 Efficient Land usage through Precision Agriculture

It helps in efficient use of land through

1. GPS aids in the selection of suitable land for specific crops.
2. GIS aids in data analysis.
3. VART aids in the application of the proper amount of pesticide to the crops, keeping the soil fruitful.

1.3 Spatial Variability of Soil

One of the spatial variable is soil variation. Topography, along with water-holding capacity or organic matter variation, creates an even more intriguing field

view when the inputs are inserted or either the soil is disturbed by a producer. Wollenhaupt and colleagues (1997) gave a comprehensive overview of soil sampling and interpolation methods that may be used to assess soil variation. Judgmental sampling, simple random sampling, stratified sampling, cluster sampling, nested or multistage sampling, systematic sampling, stratified systematic unaligned sampling, and search sampling were the approaches they defined (Hatfield, 2000). Farmers have been using in-field soil data to make crop management decisions for generations. Traditionally, grid sampling was done to estimate soil properties, but it has a disadvantage due to the cost it takes while sampling a large scale area so the difficulties of grid sampling have become more evident with the introduction of PA, which might demand high spatial resolution of in-field soil properties, therefore agricultural scientists and engineers have turned to RS for in-field soil property characterization (Ge et al., 2011). Many soil surveys, mapping, and quantitative soil-property characterization investigations have made use of both bare-soil pictures acquired through remote sensing and spectroscopic reflectance of soil samples (Agbu et al., 1990; Ben-Dor & Banin, 1994, 1995; Dalal & Henry, 1986). For several applications, the cheapest, fastest and most accurate approach is remote sensing than traditional ones (e.g., the soil textures determination through pipette method and soil organic carbon concentration by dry combustion method) (Ge et al., 2011). In general, bare-soil satellite and aerial photos provide coverage on large extent and with reasonably detailed spatial information of study regions but with limited spectrum information. In the recent year's soil property characterization in the field through remote sensing has increased dramatically in PA (Precision Agriculture). Soil property determination in PA is carried out by the established methods of RS images and spectroscopy (Ge et al., 2011). In the 1930s, the first attempt was made to employ RS for soil research, when the base maps were created by black and white aerial images for soil surveys in the United States

(Baumgardner et al., 1985). In the late 1060s, the soil scientists began to investigate the use of multispectral sensor (MSS) data to identify variations in surface soils in the late 1960s and early 1970s (Kristof, 1971).

1.4 Soil Spectroscopy

The science of spectroscopy is concerned with the interaction of matter with its electromagnetic radiation (Skoog, D. A., Holler, F. J., & Crouch, 2007). Spectral reflectance of some of the materials is shown in the figure 1.1. The history of spectroscopy dates back to Isaac Newton's discoveries of the light nature and color fundamentals in the seventeenth century. The underlying premise of Vis-NIR is based on molecular distinctions, with different material's spectral signatures being classified based on their reflectance and absorbance spectra (Mohamed et al., 2018). Development of instrumentation and its advancement in industrial applications in the second part of the twentieth century is shown in figure 1.2.

The use of laboratory spectrometers for hyperspectral soil data gathering is a recent development. It is expected that soil parameters can be quantitatively measured through adequate data manipulation due to the highly precise spectrum information (Ge et al., 2011). Visible near-infrared spectroscopy (vis-NIR) has grown in popularity as a cost-effective alternative to traditional laboratory analyses because it is quick, non-destructive and does not require hazardous chemicals. It also allows multiple soil properties to be estimated simultaneously from a single spectrum (Vohland et al., 2011). Soil chemical and physical properties, including SOM, texture and clay mineralogy has been predicted by using visible-near-infrared spectrum (Stenberg et al., 2010). Curcio et al.2013 used VNIR-SWIR band reflectance spectroscopy to predict soil texture (sand, clay, and silt) (Curcio et al., 2013). Hyperspectral soil data captured by spectrometers is frequently

noisy and challenging to analyze, even in carefully controlled laboratory settings. Due to this, data preparation processes are used to "clean" them. Smoothing is most generally used method for noise reduction which includes Savitzky-Golay smoothing, running average, mean, and median filtering. Other extraneous elements that can alter hyperspectral soil data are usually reduced through derivative analysis. The first derivative, for example, is effective at accounting for sun angles and viewing geometry (e et al., 2011).

The most commonly used statistical models for predicting different soil characteristics using spectral reflections are stepwise multiple linear regression (SMLR), partial least squares regression (PLSR), multivariate adaptive regression splines (MARS), principal component regression (PCR), and artificial neural networks (ANN) (Mohamed et al., 2018).

1.5 Objectives

The objectives of the study are as follows:

1) Investigation and generation of a predictive multilinear regression model for sand, silt, clay, OM, phosphorus and potassium content using multispectral Sentinel-2A satellite data and hyperspectral ASD Field Spec 4 spectroradiometer data.

2) Statistically comparison of the ASD and the Sentinel-2A predictive models.

3) Generation of regional soil texture maps using spatial interpolation techniques.

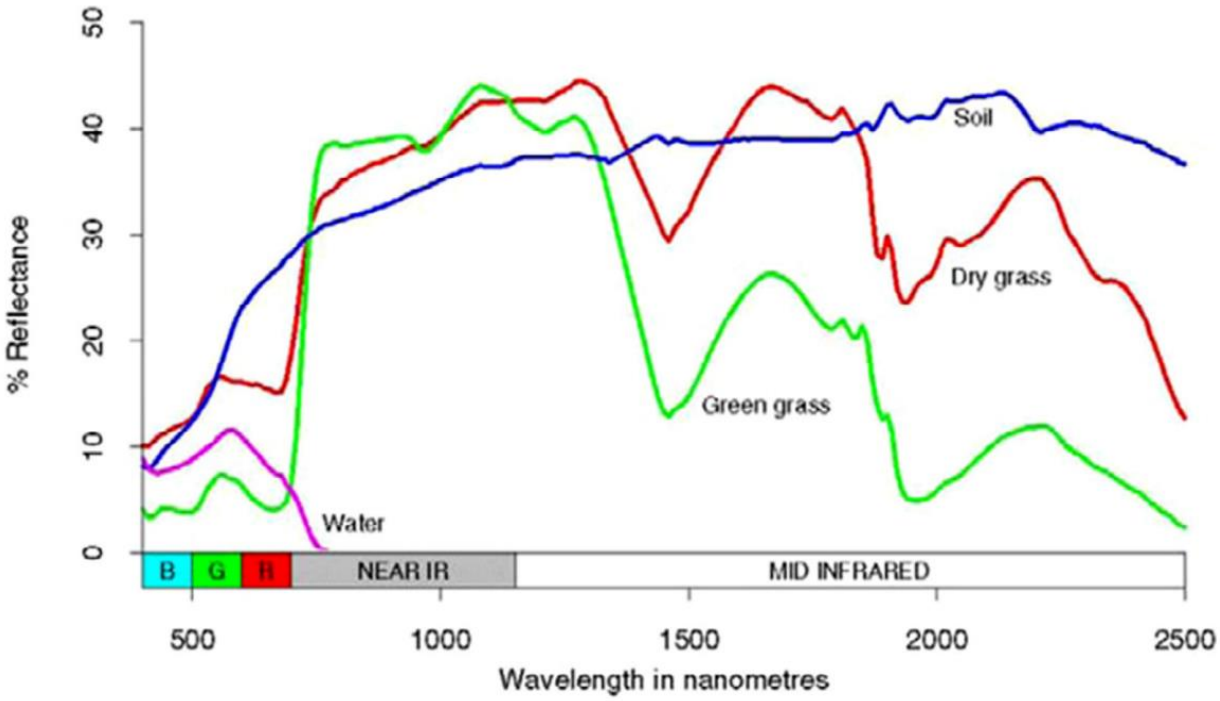

Figure 1.1. Spectral reflectance of materials (Short, 1982).

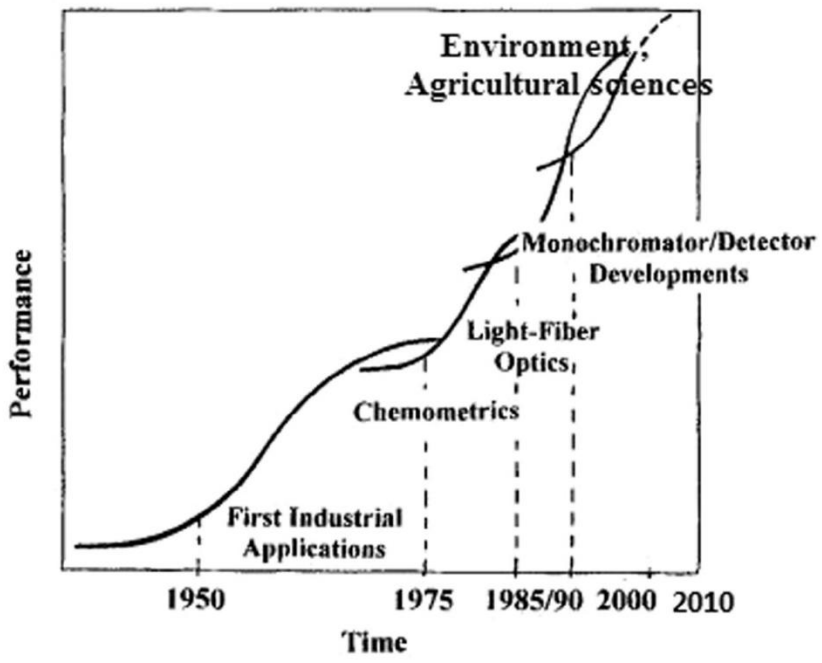

Figure 1.2. Near-infrared spectroscopy development (W. Siesler, Y. Ozaki, 2002).

CHAPTER 2

LITERATURE REVIEW

Land use planning and other agricultural management and environmental protection operations require a thorough understanding of the spatial distribution and variability of soil texture. Using a small number of soil samples collected from a site in Pingdu, Shandong Province of China, the study evaluated Landsat Enhanced Thematic Mapper (ETM) remote sensing data as auxiliary variables for spatial estimate of surface soil texture. Estimation was done through three methods firstly using multiple stepwise regression (MSR) based on the link between surface soil sand, silt, and clay concentrations and remote sensing data, secondly preparation of kriging maps of surface soil sand, silt, and clay content and thirdly cokriging using remote sensing data. Results showed that the surface soil sand, silt, and clay concentrations were strongly correlated with Landsat ETM digital number (DN) of six bands (Bands 1–5 and Band 7) and the majority of the variability in soil sand, silt, and clay contents explained by the DN of Band 7. Compared to MSR and kriging, cokriging using remote sensing data improves estimations of surface soil texture significantly (Liao et al., 2013).

For site-specific agriculture, mapping fine-scale spatial changes in soil characteristics is critical. The current research looks into the possibilities of using remote sensing (RS) and geographic information system (GIS) approaches to investigate the spatial variability of surface soil properties. A total of 170 surface (0-30 cm) soil samples were collected from Shorkot Tehsil, Punjab, Pakistan, and examined for surface soil texture and organic matter (OM). To correlate surface soil factors with spectral data from the Landsat TM5 satellite, a multivariate linear regression (MLR) analysis technique was used. The MLR equations were then utilized to model these properties spatially throughout the full study area. The USDA textural triangle restrictions for clay percent

and silt percent were used to construct a code in Visual Basic Language in an ArcGIS environment to produce a surface soil texture map. The most abundant textural class in the area was ‘sandy clay loam,' followed by ‘sandy loam,' and 'clay loam,' according to the findings. Furthermore, OM status in the entire study area soils was very poor (less than 1%). The findings suggest that RS and GIS approaches could be utilized to map fine-scale soil texture and OM across a greater area (Ahmed & Iqbal, 2014).

Soil reflectance data can be acquired quickly and cheaply through remote sensing. Knowing which soil factors impact bare soil imaging most would help farmers employ remote sensing more effectively for precision crop management. The study was conducted with the purpose to (a) identifying the most significant impact of measured soil qualities on the remotely sensed bare soil reflectance and (ii) discovering the spectral band or combination of spectral bands that appropriately define specific soil attributes. Three study sites in northeastern Colorado were used for this research. Irrigated continuous corn (Zea mays L.) cropping systems were used at all sites. Prior to planting, aircraft gathered remote sensing imagery. Bulk density, soil color (wet and dry), organic matter, organic carbon, soil texture, and cone index were all measured in soil samples. The image’s green, red, and near-infrared (NIR) bands were subjected to principal component analysis (PCA). The link between remote sensing and soil data was investigated using least-squares regression analysis. The first main components of the green, red, and NIR bands were statistically significant associations with organic carbon and the sand, silt, and clay fractions across all research locations. In this study soil spatial variability is assessed by the use of remote sensing (Mzuku et al., 2015).

The samples were collected and analyzed in the laboratory to determine or study the chemical properties of the selected sites, which were seven sites in Kirkuk city of Iraq and the other one to predict soil maps using GIS techniques. Geotechnical maps

were produced using Inverse Distance Weighting (IDW) interpolation technique (Sulyman et al., 2020)

Geostatistics offers useful methods for describing the spatial distribution of soil properties. Kriging techniques rely on the spatial dependence between observations to forecast attribute values at un-sampled places. Similar study was conducted at Mansoura University's experimental farm in Kalapshow, Bilqas District, Dakahlia Governorate, Egypt, these techniques were utilized to investigate the geographical distribution of specific soil physicochemical parameters. Surface interpolation of soil clay, available water, EC, bulk density, soil organic matter (SOM), soluble K, exchangeable K, and available K was done using Ordinary Kriging (OK) (El-Sirafy et al., 2011).

Soil property mapping is crucial because it contributes to our understanding of soil qualities and how they can be used sustainably. The research was conducted in a local government area in Oyo state to map out some soil features and examine their variability throughout the area. The cluster sampling technique was used to collect soil samples at three places around the local government region. The soil samples were air-dried, crushed, and sieved before their chemical analysis is carried out in the laboratory. Classical statistics were utilized to describe soil properties and the spatial variability is highlighted using kriging interpolation techniques in a GIS setting (Denton et al., 2017).

As the soil’s total nitrogen (TN), organic carbon (OC), and moisture content (MC) has a direct spectral response in the near-infrared (NIR) region shows the potential of visible and near infrared (VIS-NIR) spectroscopy to estimate them. The study aimed of examining the predictive performance of two linear multivariates and two machine learning approaches. Principal component regression (PCR) and partial least squares

regression (PLSR) were the two multivariate approaches explored, while least squares support vector machines (LS-SVM) and Cubist were the machine learning methods. Soil spectra (350 - 2200 nm) in diffuse reflectance mode were recorded from 140 wet soil samples collected from one field in Germany using a mobile, fibre-type VIS-NIR spectrophotometer. The findings show that machine learning algorithms can solve non-linear problems in the dataset. For the prediction of all three soil parameters investigated, LS-SVMs and the Cubist technique outperformed linear multivariate methods (Morellos et al., 2016).

For sustainable land management; there is a need to develop fast, accurate and inexpensive methods for determining soil physicochemical properties. Vis-NIR spectroscopy has become an exciting and alternative method for the determination of these properties from the last two to three decades. Multivariate calibration approaches such as Partial Least Squares Regression (PLSR) are often used to correlate spectra with chemical, physical, and mineralogical aspects of soils to produce reliable predictions of soil properties. This study aimed to see how well Vis-NIR spectroscopy combined with PLSR could be used to evaluate soil chemical and physical parameters such organic carbon (SOC), sand, silt, clay, and calcium carbonate ($CaCO_3$) in a sample location in southern Italy. Based on chemical and physical parameters, spectral curves revealed that the soils may be spectrally separated. PLSR calibration models were developed and validated for each soil parameter using an independent data set. Leave-one-out-cross-validation was used to find the optimal number of components to keep in the calibration models. The coefficient of determination (R^2) and the root mean squared error were used to assess the accuracy of the calibration and validation models for the various soil parameters (RMSE). As a result, a combination of Vis-NIR spectroscopy and multivariate

statistical approaches can be utilized to characterize the soils of southern Italy in a quick, low-cost, and quantitative manner (Massimo Conforti & Gabriele Buttafuoco, 2014).

Standard laboratory procedures for estimating soil texture are difficult, expensive, and time-consuming and require significant work. Reflectance spectroscopy is a low-cost, quick, and repeatable analytical approach for predicting various soil physical parameters. The study was conducted with a goal to see if visible (VIS: 350-700 nm) and near-infrared and short-wave-infrared (NIRS: 701-2500 nm) spectroscopy could be used to forecast and map the clay, silt, and sand fractions in the soils of the Triffa plain (northeast Morocco). A total of 100 soil samples (0-20 cm) were collected and texture evaluated using VIS-NIRS spectroscopy and the traditional laboratory approach. The capacity of spectral data to predict soil texture was evaluated using the partial least squares regression (PLSR) technique. The results of the prediction models revealed outstanding performance for the VIS-NIRS spectroscopy in predicting the sand fraction (R^2 = 0.93 and RMSE = 3.72), good prediction for the silt fraction (R^2 = 0.87; RMSE = 4.55), and acceptable prediction for the clay fraction (R^2 = 0.53; RMSE = 3.72). Furthermore, the spectral range between 2150 and 2450 nm is most important for forecasting the sand and silt fractions, while the region between 2200 and 2440 nm is best for estimating the clay percentage (Lazaar et al., 2021).

The research was done to show how to effectively analyze critical features of Mediterranean soils from southern Italy using visible–near infrared (vis–NIR) reflectance spectroscopy and partial least squares regression (PLSR). Understanding soil qualities is a necessary precondition for long-term land management. Vis–NIR reflectance spectroscopy and chemometrics have received a lot of attention in recent years. The potential of vis–NIR spectroscopy and PLSR for predicting chemical and physical properties [sand, silt, and clay, organic carbon (OC), total nitrogen (N), cation exchange

capacity (CEC), and calcium carbonate ($CaCO_3$)] of soils representative of three Mediterranean agro-ecosystems from the Campania region in southern Italy was investigated in this study. PLSR is one of the most widely used chemometrics modeling approaches, and it is frequently used for quantitative spectroscopic analysis. PLSR models were created and was then tested using a separate set of data not utilized in the modeling. The root mean squared error (RMSE) and the relative percent deviation (RPD) were used to measure the correctness of the calibrations and validations for the various soil parameters (P. Leone et al., 2012).

In soil research, detecting precise and accurate soil physicochemical attributes (SPAs) is a difficult task. With the variety of nature, the SPA can be altered spatially and temporally. SPA detection has previously been accomplished using standard soil chemical and physical laboratory testing. These laboratory procedures, however, do not meet the time constraints. As a result, non-destructive method to identify and describe soil properties is diffuse reflectance spectroscopy (DRS). 74 soil samples were agglomerated by farming sectors in Phulambri Tehsil in the Aurangabad region of Maharashtra, India, using spectral curves in the visible (350–700 nm) and near-infrared (700–2500 nm) (VNIR) regions. The VNIR spectrum was quantitatively analyzed. The Analytical Spectral Device (ASD) Field spec 4 spectroradiometer was used to collect the spectra of agglomerated farmed soils. The calibration models were built using the PLS regression methodology, which were independently tested for predicting SPA from the soil spectrum. A correlation study was performed between measured SPA and spectral reflectance in order to create a model. Soil organic carbon, nitrogen, soil organic matter, pH levels, electrical conductivity, phosphorus, potassium, iron, sand, silt, and clay were detected. The findings of this study have been found to be useful for precision farming and decision-making (Vibhute et al., 2018).

Along with appropriate farming techniques and soil quality, soil organic matter (SOM) plays a significant role in plant growth. Due to its significant geographical variability and chemical treatments, assessment of SOM is a time-consuming task. Vis-NIR reflectance spectroscopy has traditionally been used to measure the organic content of soil without the use of hazardous chemicals. As a result, VNIR spectrum reflectance is expected to be in high demand for precision farming. The reflectance spectra of 30 soil samples were collected from agricultural areas in the Phulambri Tehsil in the Aurangabad region of Maharashtra, India, using the Analytical Spectral Device (ASD) Field spec 4 spectroradiometer. The fringe channels were removed, and the absorption channels of the 400-2450nm wavebands were detected using a continuum-removed approach. The first-derivative transformation was used to smooth the spectra using the Savitzky-Golay (SG) method (FDT). Correlation study between spectral reflectance and SOM contents was used to forecast the SOM using the partial least squares regression (PLSR) model. Wavelengths of 441, 517, 527, 648, and 1000 nm are found to be sensitive to SOM channels. The research will help farmers and decision-makers for more efficient and cost-effective farming and decision making (Vibhute et al., 2018).

The study was aimed to forecast organic carbon content of soil in the hilly terrain of eastern Lesotho, southern Africa, which is home to abundant indigenous biodiversity and is heavily utilized for small-scale agriculture. Measurements of soil reflectance spectra with an Analytical Spectral Device (ASD) Field Spec® 4 optical sensor was carried out in an integrated field and laboratory to assess field spectroscopy-based model's accuracy, soil spectra were obtained on the land surface under field conditions and then on soil in the laboratory. Prediction effectiveness of two alternative models such as random forest and PLS regression were tested. The results reveal that utilizing field spectroscopic data, random forest regression can most accurately estimate

the soil organic carbon levels on an independent dataset. The partial least square regression model, on the other hand, over fits the calibration dataset. The visible range (400–700 nm) contained important wavelengths for predicting soil organic content. This work shows that using derivative field spectroscopic observations and random forest regression, soil organic carbon may be reliably estimated (Bangelesa et al., 2020).

For soil quality assessment and precision soil management, a quick and convenient soil analysis technique is required. The research was carried out to see how well visible (Vis) and near-infrared (NIR) spectroscopy could predict paddy soil parameters in a typical Malaysian rice field by sampling 118 soil samples. Laboratory analysis and optical measurements in the Vis-NIR region using an analytical spectral device (ASD) FieldSpec spectroradiometer (350–2500 nm) was done to assess the utility of spectroscopy for predicting soil physical characteristics (bulk density, moisture content, clay, silt, and sand). Savitzky–Golay method and stepwise multiple linear regression (SMLR) were used to preprocess, model, and predict the properties based on their spectral reflectance in the Vis-NIR range. The study found that for all of the measured soil physical characteristics, Vis and NIR spectroscopy calibration models provided a good fit ($R^2 > 0.78$); thus, Vis and NIR (specifically NIR reflectance) can be considered a reliable tool for assessing soil physical properties in Malaysian paddy fields. The findings of this study could be very useful in designing site-specific management (Gholizadeh et al., 2014).

Two different methods were compared for the estimation of soil texture using VNIR-SWIR reflectance measurements. First one, the Continuum Removal (CR) technique, which was used to correlate specific spectral absorption features with clay, silt, and sand content, and secondly, the Partial Least-Squares Regression (PLSR) method, which is a classic statistical multivariate technique that uses full-spectrum data. The

surface reflectance of 100 soil samples was taken in the lab that were collected from various locations across Sicily encompassing a wide range of textures using an ASD Field Spec Pro spectroradiometer to achieve this goal (350-2500 nm). The PLSR strategy outperformed the CR approach, according to our findings. The use of root mean squared error (RMSE) and coefficient of determination (R^2) was done to analyze soil texture accuracy and it revealed that the CR technique only allowed for a reasonable prediction for the clay texture component (Curcio et al., 2013).

CHAPTER 3

MATERIALS AND METHODS

3.1 Study Area

The study area selected was tehsil Talagang of district Chakwal, is in the Pothohar plateau of Punjab, Pakistan. It covers an area of approximately 2,022 km^2 (figure 3.1). Chakwal is an agricultural area where the production of wheat, barley, sugarcane and many other fruits especially the international standard oranges and vegetables are the cause of its popularity. It's called a 'Barani' area and has no proper canal system. Talagang lies at the latitude of 32.91°, longitude of 72.40° E and at an altitude of 460m. Talagang Tehsil is one of the largest tehsils in South Asia as it contains 102 villages, consisting of 23 union councils. Major crops grown here wheat, barley, mongphali (peanuts), Jowar, Bajra. Annual temperature ranges from 2.77°C to 38. 8°C. It has a semi-arid climate.

3.2 Soil Sampling

A random soil sampling was carried out excluding the uncultivated areas such as that are sloppy, urbanized and with dense vegetation etc. The accurate location of sampling points was located using Garmin global positioning system (Vasileios C. Drosos & Chrisvaladis Malesios, 2012). Total of 51 samples were collected from the study area considering the soil depth of 0-20 cm (Vibhute et al., 2018). Samples were collected using trowel. Before collecting the soil from the area, the plant residue or debris was removed including any stones if present. Sampling was done during the month of Jan-Feb, 2021 and under clear sky conditions. The samples were immediately packed in plastic bags and were labelled with sample number and coordinate points. Lab analysis were carried out in the soil physical lab of PMAS Arid Agriculture university.

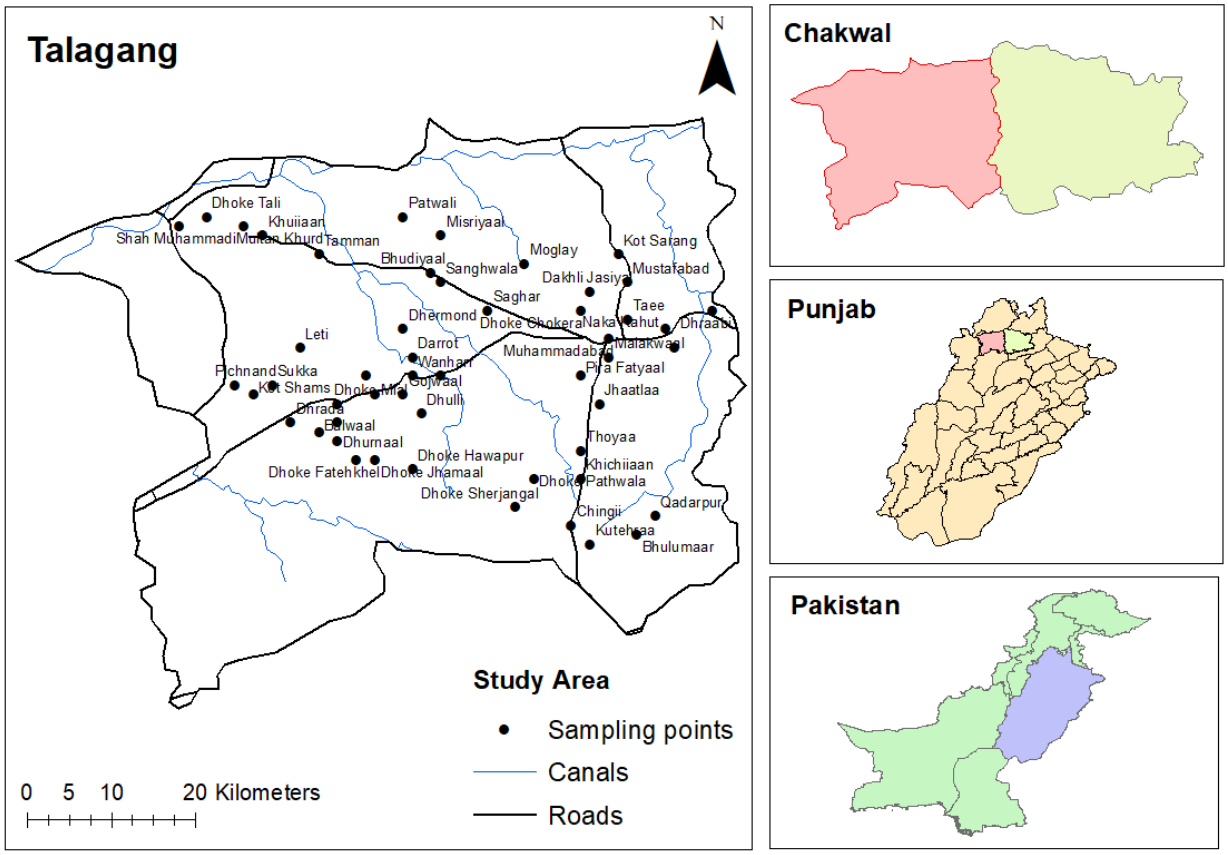

Figure 3.1: Study area map.

3.3 Soil Analysis

Soil samples were sun and air-dried after being transported into the laboratory. Then were crushed and sieved (Beuselinck et al., 1998) using 2mm sieve to remove all the debris or plant residues if remained. Hydrometer method (Bouyoucos, 1927) was used for the determination of soil texture following the ICARDA manual. Figure 3.2 and 3.3 shows the methodology adopted for the determination of the texture of soil through hydrometer method. Each sample's chemical characteristics (OM, EC, P, K and pH) were calculated. In the laboratory, soil OM was estimated using the Walkley-Black chromic acid wet oxidation method (EPA, 2002). Soil P with Olsen's Method (Sims, 2000), soil K with a Flamephotometer and other equipment, and soil pH with a spatula and pH meter (Bremmer y Mulvaney, 1982). At 25°C and 1 atmospheric pressure, the E.C. of soil was determined using a conductivity meter and associated equipment (Smith & Doran, 2015).

3.4 Collection of Soil Spectral Signatures

The spectral signatures of 51 soil samples were taken under clear weather condition on the sunny day in the month of August, 2021. Signatures of soil samples were taken by ASD Field Spec 4 Spectroradiometer (Analytical Spectral Devices Inc., Boulder, Colorado, USA) as shown in figure 3.4. It's a compact, field portable, and precision instrument with a spectral range of 350–2500 nm and has a rapid data collection time of 0.2 second per spectrum. Prior spectral analysis, the optimization and calibration of the instrument was done using white reference spectralon panel to make sure no impurities or noise is recorded and absolute reflectance is obtained. It was also made sure that the soil samples were fully dried so that they don't contain any moisture otherwise accurate spectral signatures would not be obtained. To ensure that 100% of light is reflected back, the 100 percent white reference was manufactured from lime material. The soil was placed on white paper and was smoothed before taking signature to ensure

uniformity. An average of 5 spectra was obtained for each soil sample. The instrument was calibrated with white reference panel after 5 to 10 minutes to avoid any noise being recorded. The inbuilt RS3 software was used to record spectral signatures. Signatures were recorded in ASD binary file format. Spectral library created from these spectral signatures of soil samples is shown in figure 3.5.

3.5 ASD Field Spec 4 Data

3.5.1 Pre-processing of Spectra

For the preprocessing of spectra firstly there is need to convert the .asd files into readable format such as in text form. Hence, to achieve this objective the ASD binary files were converted to text using the View Spec Pro (version 6.2.0) software (Analytical Spectral Devices, Inc., Boulder, CO, 80301). The five consecutive scans were then averaged for each soil sample and spectral signature was obtained to create a spectral library. The fringe spectra consisting of wavelength from 350-399 nm and 2350-2500 nm were removed before statistical modelling as shown in figure 3.6. And also the wavelength ranges from 1350-1460 nm and 1790-1960 nm were also eliminated because they show scattering due to atmospheric water absorption as they would affect the modeling.

3.6 Statistical Analysis

3.6.1 Stepwise Multiple Linear Regression (SMLR)

The spectra were brought into the SAS software after all the pre and post processing to develop a linear regression model. As the spectral data has the great amount of multi-collinearity so there is need to consider those wavebands in the model that show significance with the soil physicochemical properties. In order to find those wavebands

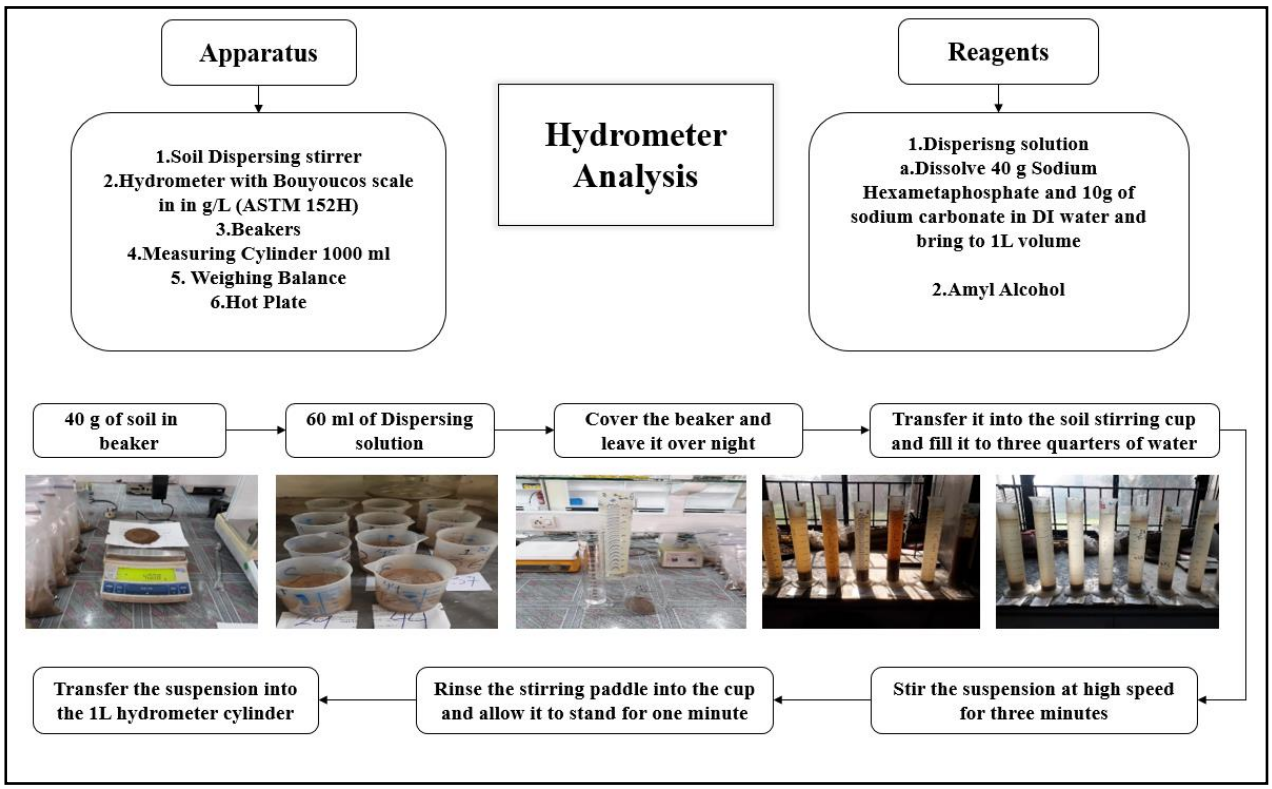

Figure 3.2. Methodology of hydrometer analysis.

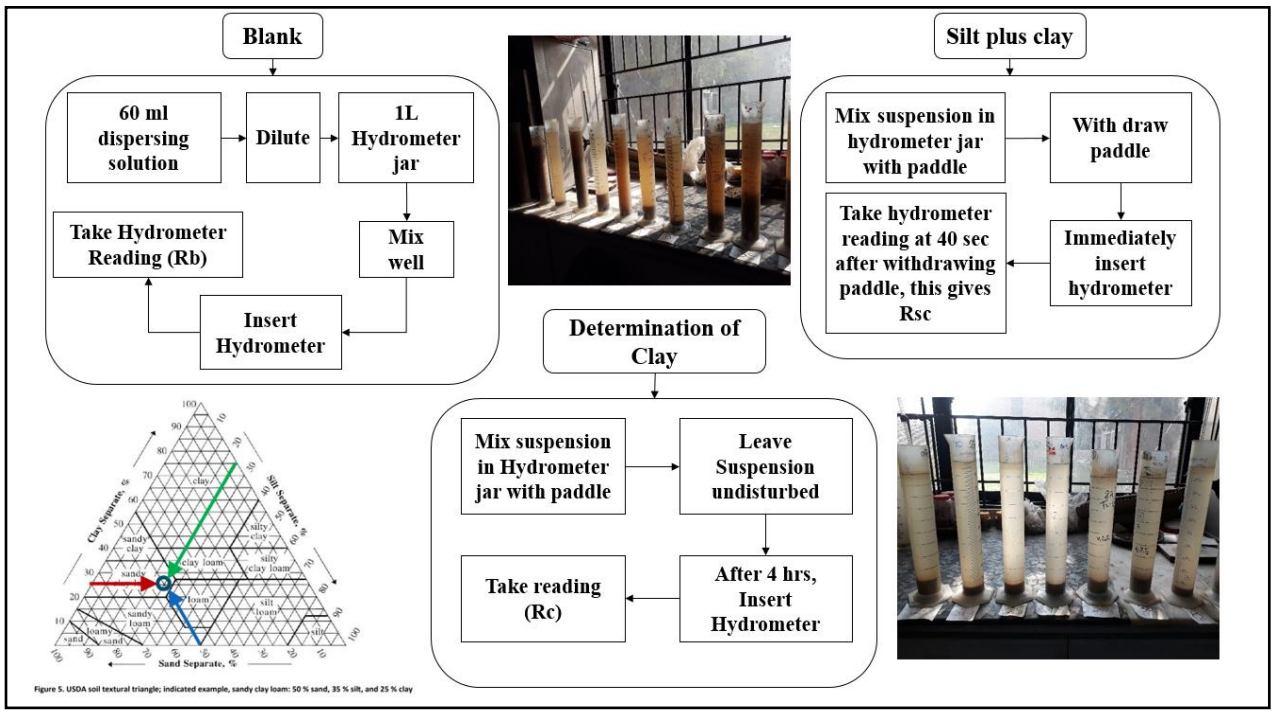

Figure 3.3. Shows the steps to determine sand, silt and clay.

Figure 3.4. Collection of soil spectral signatures.

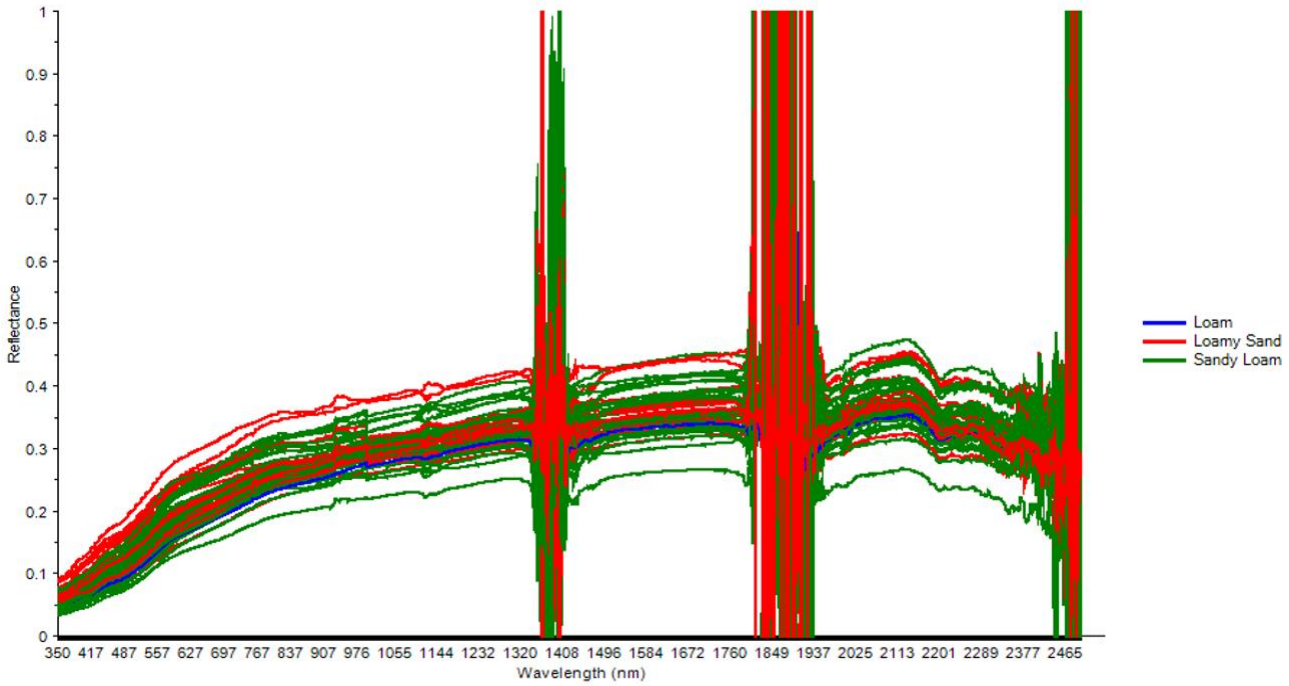

Figure 3.5. Soil spectral library.

stepwise regression was applied. The models developed are explained in the section 4.5.

3.6.2 Partial Least Squares Regression (PLSR)

After removing noisy and unnecessary bands, the spectra were imported into Unscrambler X (version 10.4) software CAMO, Norway for statistical analysis. Prior to building a model, the spectra were transformed by applying Savitzky-Golay smoothing with 1st derivative and 7 smoothing points, as shown in figure 3.7. The necessary transformation is the smoothing of spectra and to remove particle size effects and noise produced due to illumination variations as stated by (Volkan Bilgili et al., 2010) and (Tsai & Philpot, 1998).

3.7 Satellite Sentinel-2A Data

Remote sensing data, including multispectral images like Sentinel 2A were used for the analysis. The images were downloaded from Copernicus Open Access Hub. Bare soil images were downloaded and were kept cloud-free. Starting with the In-Orbit Commissioning Review (IOCR), the Copernicus Open Access Hub (formerly known as Sentinels Scientific Data Hub) gives complete, free, and open access to Sentinel-1, Sentinel-2, Sentinel-3, and Sentinel-5P user products. Image preprocessing was done using SCP (Semi-Automated Classification Plug-in in QGIS (version 3.4.6). The processed images were mosaicked and the study area was extracted in ArcMap (version 10.8.1). Sentinel-2 MSI (Multi-Spectral Instrument) has a total of 13 spectral bands; four bands at 10 m, six bands at 20 m and three bands at 60 m resolution. Before analysis the band-1, band-9 and band-10 were removed as they are sensitive to aerosol scattering, water vapour correction and clouds at high altitudes respectively. Indices including such as Normalized Difference Vegetation Index (NDVI), Modified Soil Adjusted Vegetation Index (MSAVI), Bare soil Index (BSI), Soil Adjusted Vegetation Index (SAVI) were also

calculated in ArcMap (version 10.8.1). Field survey data points with all physicochemical properties, Sentinel-2 bands, Digital Elevation Model (DEM), SAVI, MSAVI, BSI and NDVI values were compiled into an integrated Esri File Geodatabase. Esri File Geodatabase was created by using a tool called 'Extract multi-values to points' in ArcMap (version 10.8.1). Table 3.1 explains the datasets used in the research. Table 3.2 shows the software used in this research for analysis purposes. Figure 3.8 depicts a general overview of methodology used, which is explained in detail in later sections.

3.8 Statistical Exploration and Multivariate Regression

The descriptive statistics of soil physical and chemical properties using classical statistics minimum, maximum, average, standard deviation and skewness were calculated as listed in Table 4.1. Data was examined for anomalies and outliers (Esfandiarpoor Borujeni et al., 2010). A correlation matrix was calculated for the soil properties. It was also calculated among soil properties and band reflectance data. Multiple linear regression (MLR) was applied on each of the soil property considering them as y (dependent variable) and x as (independent variable) (Forkuor et al., 2017) defined as y = a + x 1 + x 2 + x 3 + ……. + x n + Σ e, where Σ e, is residual. To measure the prediction accuracy, the coefficient of determination R^2 and the Root Mean Square Error (RMSE) were considered. The results are presented in depth at section 4.31.

3.9 Geostatistics & Ordinary Least Squares Regression

As we know the classical statistics does not account for the spatial variability, so it has some limitations while using spatial data. So spatial data mapping cannot be done using correlation and linear regression. This is where Geospatial statistics comes (López-Granados et al., 2005). ***"Geostatistics is a branch of statistics that studies and forecasts the values of spatial and spatiotemporal events".*** Dependent variables are not a

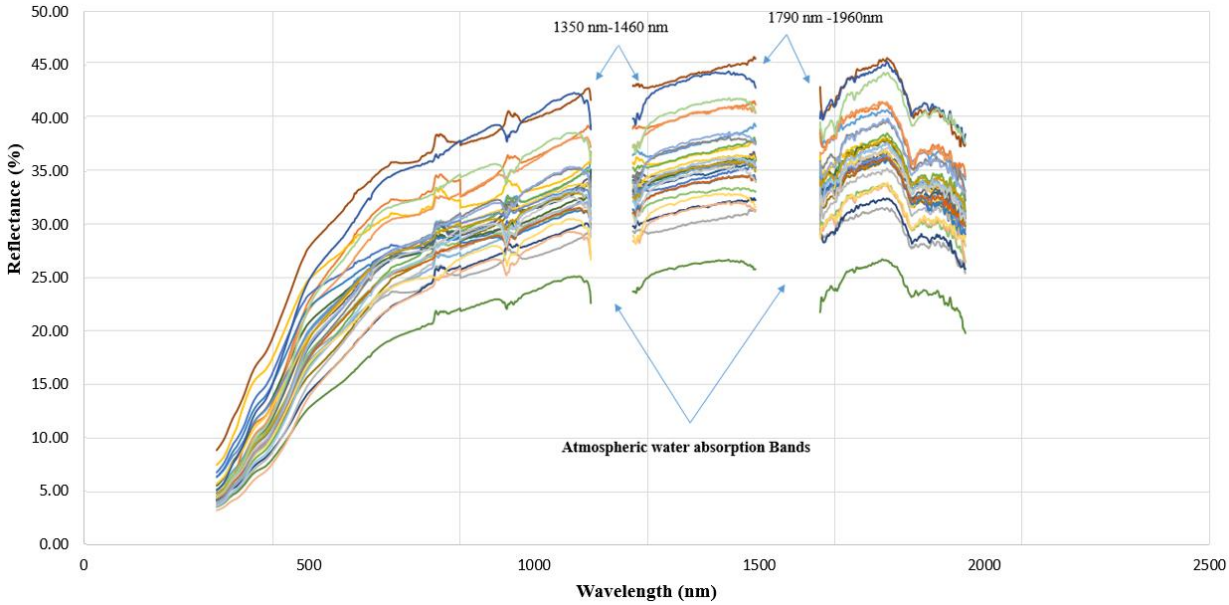

Figure 3.6. Spectral data with removed noisy regions such as from 350-399nm, 2350-2500nm,1350-1460nm and 1790-1960 nm.

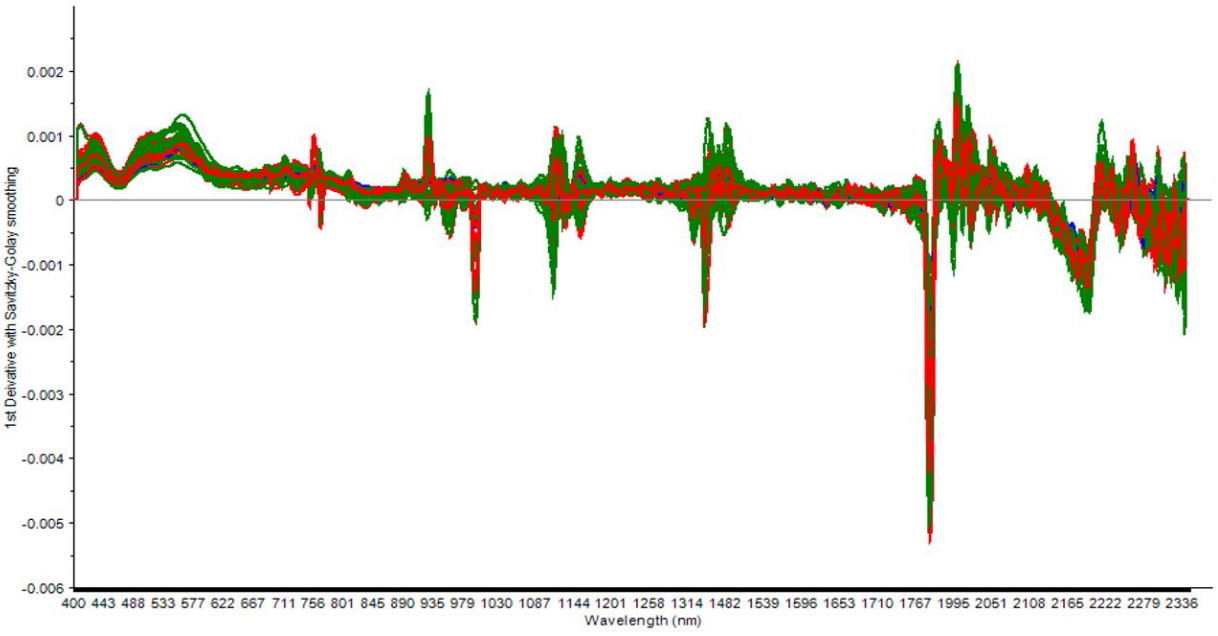

Figure 3.7. First derivative Savitzky-Golay transformation.

Table 3.1. Datasets used in this research.

Dataset	Description	Source
Soil physicochemical properties	a. Texture determination through Hydrometer Analysis b. Lab Analysis of soil chemical properties	Samples collected from the study area while doing field survey
Satellite Imagery	Sentinel-2A (13 spectral bands)	Copernicus Open Access Hub https://scihub.copernicus.eu/dhus/#/home (Atmospherically corrected images)
DEM (Digital Elevation Model)	ALOS PALSAR (12.5m resolution)	Alaska Satellite Facility https://asf.alaska.edu/data-sets/sar-data-sets/alos-palsar/
GPS Co-ordinates	Georeferencing of soil samples	The location of soil sampling was determined using a Garmin GPS.

Table 3.2. Softwares used in this research.

Software	Description	Source
ArcGIS 10.8.1	Geospatial analysis	ESRI
Microsoft Excel	Data Analysis	Microsoft
R Studio and R Program	For Correlation matrix and Random Forest Regression	https://www.rstudio.com/ https://www.r-project.org/ (open source)
SAS	On Demand for Academics	https://www.sas.com/en_us/software/on-demand-for-academics.html

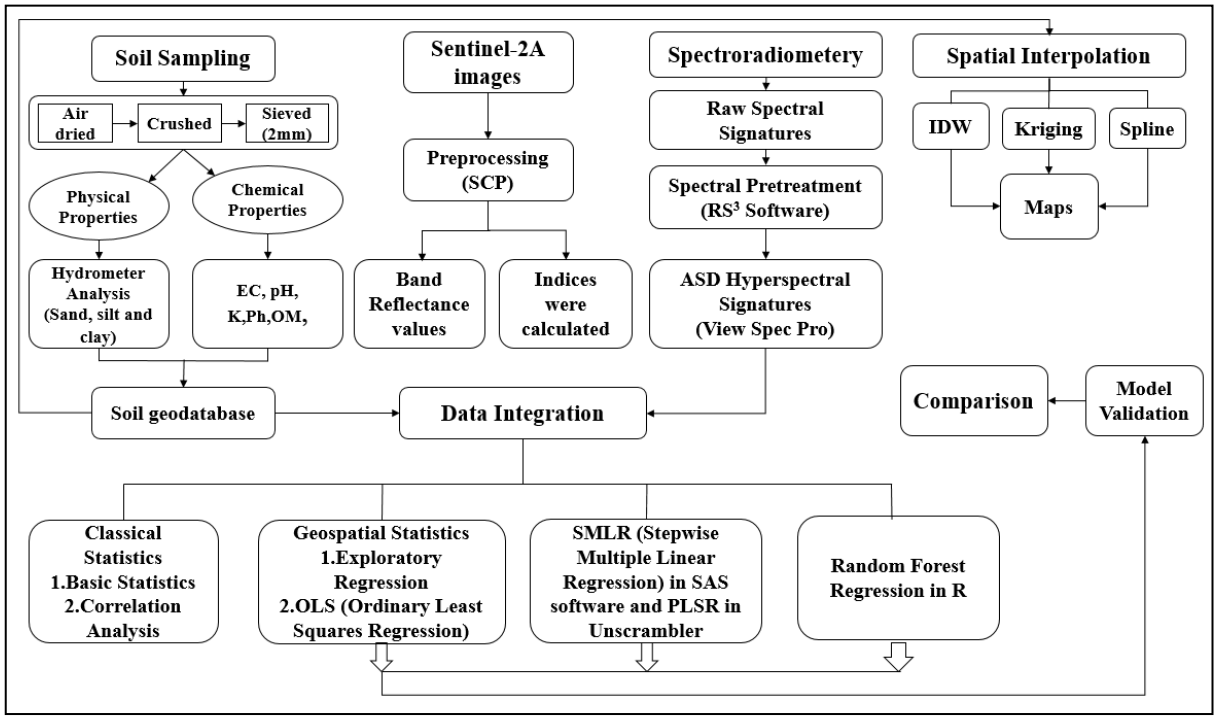

Figure 3.8. Methodological flow chart.

problem when using OLS, but explanatory variables are; thus the first step is to figure out which explanatory variable will best explain the models. Hence, to find out best explanatory variables, exploratory regression was performed. ***"Exploratory regression would help us to know the strength of the relationship***". The OLS (Ordinary Least Square Regression) was then applied to the filtered parameters (Godinho Silva et al., 2016). Ordinary minimum square regression (Mirchooli et al., 2020) is a sort of spatial modeling that investigates the relationship between spatial variables. The general form of OLS equation is described as

$$y = \beta o + \sum_{j=1}^{n} \beta_j x_i + \Sigma\, e \qquad \textit{Equation– 2.5}$$

$$y = dependant\ variable,\ x_i = \ independant\ variable, \beta o = intercept,\ \beta_j = coefficient, \Sigma\, e = Error$$

The terms RMSE, R^2, Adjusted R^2, and Histogram of standard residuals were used to evaluate the results of geospatial modeling, which were then studied and compared to classical modeling. The comparison of results is discussed in detail in section 4.3.2.

3.10 Artificial Neural Network for Soil Mapping

The Artificial Neural Network (ANN) (Khaledian & Miller, 2020) was used to explore data in the last phase of the research. ANN is built on the human brain intricate and interconnected conceptual structure (Hassoun et al., 1996). Random forest regression was used to explore the different soil properties. Random forest regression is a supervised learning approach for regression that use the ensemble learning method. The ensemble learning method combines predictions from several machine learning algorithms to produce a more accurate forecast than a single model. During training, a random forest constructs many decision trees and outputs the classes' mean as the prediction of all the trees. Random forest (RF) is a classifier or regression model that consists of a large number of decision or regression trees, each of which is reliant on the values of a random

vector generated separately and with the same distribution for all trees in the data (Liaw & Wiener, 2002). Breiman (2001) created the Random Forest (RF) technique for data mining. It can be used to eliminate redundancy in hyperspectral datasets with several dimensions. The findings of (Abdel-Rahman et al., 2013) reveal that RF regression applied to hyperspectral data may accurately forecast sugarcane leaf N content, enabling in making wise decisions about site-specific N fertilizer application. Figure 3.9 shows the structure of random forest. Results are shown in section 4.7.

The 501 sample points were generated for applying random forest regression on soil data to model soil properties with the help of a soil sampling tool that used a plug in ArcMap 10.8.1. The IDW interpolated maps of sand, silt, clay, OM, P and K were used to extract values at generated points. The R program was used for random forest regression modeling.

3.11 Geospatial Interpolation for Soil Mapping

In addition to the above modeling approaches, geospatial interpolation was used on soil data. In contrast to traditional modeling approaches, spatial prediction techniques, also known as spatial interpolation techniques, incorporate information on the physical location of the sample data points (Ver Hoef & Cressie, 1993). Spatial predictions can be used to describe a wide range of responses at various spatial sizes (Schloeder et al., 2001). Geostatistical kriging-based tools for spatial analysis, such as Simple and Ordinary Kriging, Universal Kriging, and Simple Cokriging (Ver Hoef & Cressie, 1993; Ziegel et al., 1998) have been widely employed.

3.11.1 Inverse Distance Weighted Interpolation

Inverse Distance Weighted interpolation and its variations (Franke, 1982) are the most commonly used deterministic approaches. The IDW interpolation uses a linearly

weighted combination of sample points to determine cell values. The inverse distance between the points determines the weight. This method presupposes that the influence of the variable being mapped out decreases as the distance from the sampled place increases (Oshunsanya et al., 2017). (Ajaj et al., 2018) states that to predict values at unknown positions, inverse distance weighted (IDW) interpolation uses a weighted average of values at known sample locations. Mathematically, the IDW equation is represented below where m is the number of closest points, and p is the parameter usually 2 (Watson, 1992).

$$F(r) = \sum_{i=1}^{m} w_i z(r_i) = \frac{\frac{\sum_{i=1}^{m} z(r_i)}{|r-r_i|^p}}{\sum_{j=1}^{m} * \frac{1}{|r-r_j|^p}} \qquad Equation\ 3.1$$

3.11.2 Kriging

Kriging is a form of interpolation technique that predicts unknown positions using distance and degree of variation of known samples. The general types of kriging are Universal Kriging and Ordinary Kriging.

Mathematically kriging is defined as

$$\hat{Z}(s_0) = \sum_{i=1}^{n} \lambda_i Z(s_i) \qquad Equation\ 3.2$$

$Z(s_i)$ = the measured value at the ith location; λ_i = an unknown weight for the measured value at the ith location; s_0 = the prediction location; N = the number of measured values.

Various types of kriging methods are there. Among them the widely used method is Ordinary Kriging, which is defined as $Z(s) = \mu + \varepsilon(s)$ μ: constant.

3.11.3 Spline Interpolation

In this interpolation method, the cell values are approximated using a mathematical function that minimizes total surface curvature, resulting in a smooth surface that passes through the input point locations exactly.

This study explored IDW, OK and spline Interpolation for soil chemical and physical properties. IDW, OK and spline were compared based on RMSE (Setianto & Triandini, 2015). Their results have been discussed in the section 4.5. The RMSE of each point between observed and anticipated data was calculated and used to evaluate geospatial modeling performance.

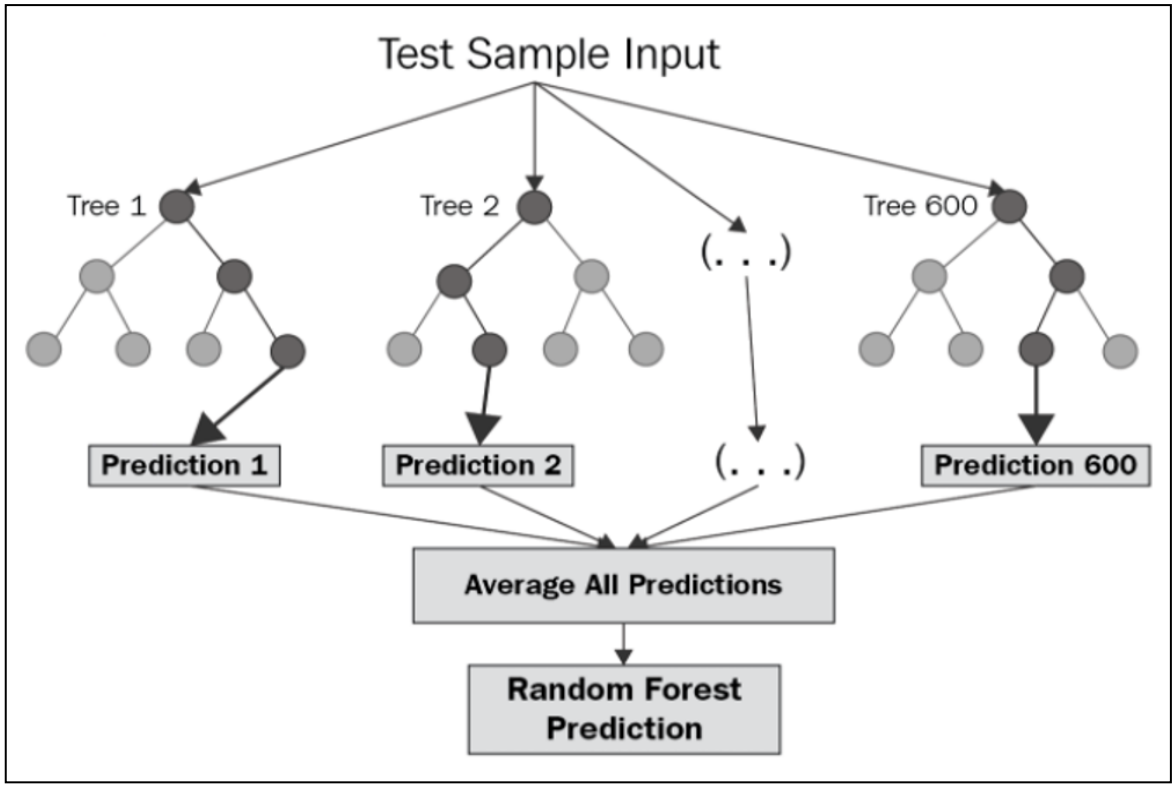

Figure 3.9. The structure of a random forest.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Descriptive Statistical Analysis

Table 4.1 show results of soil properties using basic classical statistics where minimum, maximum, mean, standard deviation, skewness and kurtosis were used to examine 51 samples. It was observed that sand ranges from 47.5% to 87.5 % with a mean value of 72.6 %. Silt ranges from 1.5% to 46.5% with a mean value of 18.4%. Clay ranges from 5% to 13.5% with a mean value of 8.85%. EC values ranges from 0.76 dSm^{-1} to 1.52 dSm^{-1} with a mean value of 0.90 dSm^{-1}. The soil pH values ranges from 7.61 to 8.15 with a mean value of 7.77, this shows the soil of the study area is good for farming practices. OM ranges from 0.41% to 0.92% with a mean value of 0.65%, these values shows that the soil is having less amount of organic matter that might not be sufficient for growth of crops. K (potassium) ranges from 6 $mgkg^{-1}$ to 145 $mgkg^{-1}$ with a mean value of 75.3 $mgkg^{-1}$. P (phosphorus) ranges from 3 $mgkg^{-1}$ to 8 $mgkg^{-1}$ with a mean value of 4.96 $mgkg^{-1}$. Figure 4.1 shows the correlation matrix of soil properties. Its shows significance at 0.05 significance level. It is seen that highest positive correlation exists between OM and P while slight positive correlation exists between OM and EC, P and EC and K and silt. Negative correlation exists pH between and EC, EC and clay, silt and Clay.

4.2 Analysis of Remote Sensing Data

Table 4.2 shows only significant remote sensing wavelength ranges with soil physicochemical properties for sentinel-2A images. Soil sand particle was found to be statistically significant with sentinel band-2 (0.490 µm), band-7 (0.783 µm), band-8 (0.842 µm) and band-8A (0.865 µm). Clay was found to be significant with sentinel band-2 (0.490 µm), band-5 (0.705 µm), band-8A (0.865 µm) and band-11 (1.610 µm).

Silt was found to be significant with sentinel band-7 (0.783 µm), band-8 (0.842 µm), band-8A (0.865 µm) and band-11 (1.610 µm). OM was found to be significant with sentinel band-2 (0.490 µm), band-4 (0.665 µm), band-5 (0.705 µm) and band-11 (1.610 µm). P was found to be statistically significant with band-3 (0.560 µm), band-4 (0.665 µm), band-5 (0.705 µm) and band-11 (1.610 µm). K was found to be statistically significant with band-4 (0.665 µm), band-5 (0.705 µm), band-8 (0.842 µm) and band-11 (1.610 µm).

4.3 Non-Spatial and Spatial Modelling

4.3.1 Prediction of soil properties using Multiple Linear Regression (MLR)

Francis Galton was the first to utilize MLR (Multiple Linear Regression). In soil prediction, MLR (Multiple Linear Regression) modelling is commonly employed (Abrougui et al., 2019). Multiple regression is another name for multiple linear regression (MLR). It is a statistical strategy that predicts the outcome of a responsible variable by combining numerous explanatory variables. MLR is a single-explanatory-variable extension of linear ordinary least squares regression. MLR is a method for calculating the relationship between two or more independent and dependent variables. $y = a + x\,1 + x\,2 + x\,3 + ... + x\,n + \Sigma\, e$, is general form of equation for MLR, where y is the dependent variable, 'a' is the intercept, x1, x2, x n is the collection of explanatory or independent variables, and Σ e is the residual (Azadi & Karimi-Jashni, 2016). MLR was used to model soil physicochemical properties (sand, silt clay, OM, P, K) using a set of explanatory factors (soil properties, Sentinel-2A data). Models developed are shown in table 4.3. All model accuracy is explained in section 4.9.1.

Table 4.1. Descriptive statistics of soil properties.

Soil Attributes	Sand	Silt	Clay	OM	EC	pH	K	P
	(%)				(dS m^{-1})		($mgkg^{-1}$)	
Min	47.5	1.5	5	0.76	7.61	0.41	6	3
Max	87.5	46.5	13.5	1.52	8.15	0.92	145	8
Mean	72.6	18.4	8.85	0.90	7.77	0.65	75.3	4.96
Standard Deviation	9.18	9.81	2.41	0.20	0.11	0.12	26.4	1.18
Skewness	-0.81	0.66	0.59	1.83	1.13	-0.11	-0.26	0.30
Kurtosis	0.77	0.52	-0.52	2.59	3.09	-0.33	2.27	-0.54

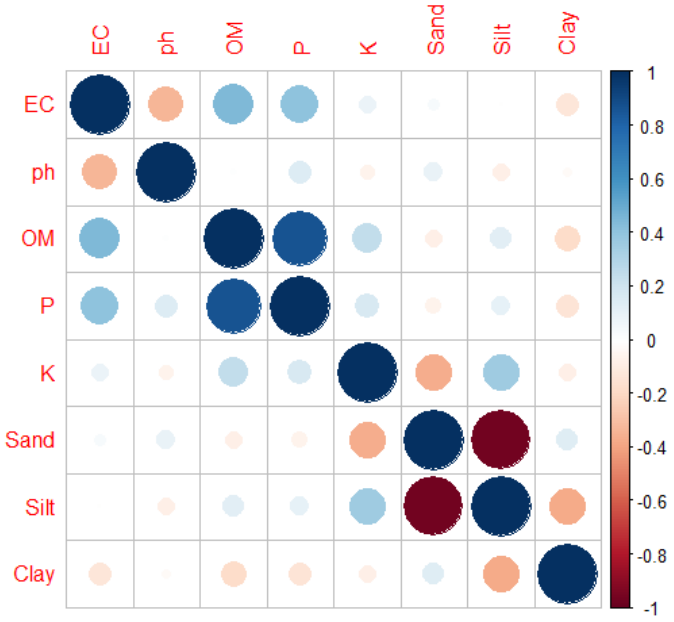

Figure 4.1. Correlation matrix.

Table 4.2. Band specification of Sentinel-2A.

Explanatory variables (Bands)	Central Wavelength (µm)	Spatial Resolution (m)	Significance with variable
Band-2 (Blue)	0.490	10	Sand, Clay, OM
Band-3 (Green)	0.560	10	P
Band-4 (Red)	0.665	10	OM, P, K
Band-5 (Vegetation Red Edge)	0.705	20	Clay, OM, P, K
Band-7 (Vegetation Red Edge)	0.783	20	Sand, Silt
Band-8 (NIR)	0.842	10	Sand, Silt, K
Band-8A (Vegetation Red Edge)	0.865	20	Sand, Silt, Clay,
Band-11 (SWIR)	1.610	20	Silt, Clay, OM, P, K

4.3.2 Prediction of Soil Properties using Exploratory Regression & Ordinary Least Squares Regression (OLS)

Exploratory regression is a data mining tool that tries all conceivable combinations of explanatory variables to see which models pass all of the OLS diagnostics. We may considerably improve our chances of finding the optimum model to solve our problem or answer our question by examining all potential combinations of the candidate explanatory variables. Exploratory regression looks for models that satisfy all of the OLS model's requirements and assumptions. Exploratory regression analysis (Kalota, 2017) identifies possible variable combinations for geographical data modeling (ESRI). For regression approaches, OLS is the most well-known. It's also where all spatial regression analysis begins. It gives us a big picture of the variable or process we seek to figure out or forecast. It represents the process using a single regression equation. Table 4.4 shows the results of exploratory regression for soil sand, silt, clay, OM, P and K using Sentinel-2A bands. Adjusted R^2, AIC (Akaike Information Criteria), JB (Jarque-Bera), VIF (Variance Inflation Factor), and SA (Spatial Autocorrelation (the Global Moran's I p-value)) are used to evaluate the model. A decent model has a high Adjusted R^2, a low AIC, (JB-nonsignificant), VIF 10 OR 7, and SA (NO pattern should exist). The validity and correctness are discussed in length in Section 4.9.2. Soil physicochemical properties were predicted using OLS (Ordinary Least-Squares Regression) after being explored through exploratory regression is shown in table 4.5.

4.3.3 Prediction of Soil Properties using Hyperspectral Data

Multivariate calibrations are required to investigate of soil diffuse reflectance spectra (Martens & Naes, 1989). Stepwise multiple linear regression (SMLR) (Ben-Dor & Banin, 1995), principal component regression (PCR), and partial least squares regression are the most used calibration methods for soil applications (PLSR). The fundamental rationale

for employing SMLR is that the traditional regression techniques like multiple linear regression (MLR) are insufficient and the soil scientists are unaware of full-spectrum data compression techniques like PCR and PLSR (Stenberg et al., 2010). Although PLS is effective for modeling spectral data, Bajwa et al., (2010) argue that it is too complicated for producers and crop consultants to adopt, as opposed to SMLR, which is considerably simpler and more flexible.

4.4 Correlation of Soil Spectra

After the pre-processing of spectra, the wavebands were correlated. The correlation analysis was carried out between sand, silt, clay and reflectance values obtained at each waveband in order to find significant wavelength ranges for sand, silt and clay. The bands ranges from 1001-1121, 1128-1344, 1494-1766, 1769-1773, 1984-1987, 1989-1993, 2030-2037, 2045-2047, 2126-2133, 2136-2150 and 2158-2349 nm were positively correlated to silt except for the band range 400-423 nm are negatively correlated at 0.05 % significance level. And all these bands showed opposite behavior with sand as 400-433 nm showed a positive correlation while 1001-1121,1129-1344, 1493-1766, 1769-1774, 1982-1995, 2030-2039, 2044-2048, 2118-2153 and 2157-2349 nm band ranges have shown a negative correlation. Correlation analysis was performed between spectral reflectance and SOM in different soil types where significant correlations exist between 550-850 nm at 0.05 or 0.01 levels and the most sensitive bands are found at 650 -750 nm because of significant statistical correlation exists at 0.01 level (Lu et al., 2015).

4.5 Stepwise Multiple Linear Regression (SMLR)

Spectral reflectance data was modelled by using SMLR in SAS software (SAS OnDemand for Academics), considering soil properties as dependent variables while reflectance values as independent variables or explanatory variables.

Table 4.3. MLR models developed using Sentinel-2A band reflectance values.

Soil Attribute	Multiple Linear Regression Model	R^2	R^2 (Adj)
Sand (%)	$75.11 + 0.01(B) + 0.05(V7) + 0.03(NIR) - 0.09(V8A)$	0.19	0.05
Silt (%)	$-5.09 - 0.09(V7) - 0.03(NIR) + 0.11(V8A) + 0.007(S11)$	0.17	0.10
Clay (%)	$23.3 - 0.01(B) + 0.01(V5) - 0.009(V8A) - 0.003(S11)$	0.12	0.05
OM (%)	$0.59 - 0.0007(B) + 0.0007(R) - 0.0008(V5) + 0.05(S11)$	0.23	0.17
P ($mgkg^{-1}$)	$1.60 - 0.004\ (G) + 0.005(R) - 0.009(V5) + 0.006(S11)$	0.22	0.15
K ($mgkg^{-1}$)	$167.67 + 0.08(\mathrm{R}) - 0.137(\mathrm{V5}) - 0.07(\mathrm{NIR}) + 0.07(\mathrm{S11})$	0.22	0.15

Table 4.4. Results of exploratory regression

Soil Attribute	Sand (%)	Silt (%)	Clay (%)	OM (%)	P ($mgkg^{-1}$)	K ($mgkg^{-1}$)
Explan-atory Variable	Blue*, Vegetation Red Edge**	Vegetation Red Edge**	Blue*, Vegetation Red Edge***	Short-wave Infrared ****	Green*****, Short-wave Infrared******	Vegetation Red Edge***, Short-wave Infrared*******
Adj. R^2	0.67	0.29	0.32	0.51	0.25	0.18
AIC	372.68	382.40	237.60	-66.05	163.87	475.79
JB	0.00	0.06	0.14	0.88	0.83	0.27
VIF	1.84	1.00	2.09	1.00	3.50	3.04
SA	0.12	0.38	0.49	0.05	0.00	0.08

* Band-2,

** Band-8A,

*** Band-7,

**** Band-11,

***** Band-3

Table 4.5. Spatial regression equations.

Soil Attribute	Ordinary Least Squares Regression	R^2	R^2 (Adj)	AIC
Sand (%)	$63.5 + 0.03(Blue) - 0.0017(Vegetation\ Red\ Edge)$	0.11	0.07	372.6
Silt (%)	$46.7 - 0.007(Vegetation\ Red\ Edge)$	0.01	0.009	425
Clay (%)	$17.03 - 0.008(Blue) + 0.002(Vegetation\ Red\ Edge)$	0.08	0.05	237.5
OM (%)	$0.27 + 0.0001(SWIR)$	0.05	0.03	-66.0
P ($mgkg^{-1}$)	$2.24 - 0.003(Green) + 0.003(SWIR)$	0.10	0.07	163.8
K ($mgkg^{-1}$)	$177.6 - 0.09(Vegetation\ Red\ Edge) + 0.04(SWIR)$	0.19	0.15	475.7

as explanatory variables. SMLR is a statistical method for regressing many variables while deleting those that aren't important. It's an automatic technique that is used to select predictive variables (Draper et al., 1966; Efroymson, 1960; Hocking, 1976). Each step of SMLR considers a variable whether to add to or delete it from the set of explanatory variables in the form of a series of F-tests or t-tests. The developed models are shown in table 4.6.

4.6 Partial Least Squares Regression (PLSR)

PLSR is a most accepted chemometrics modeling strategy or method commonly used to interpret quantitative reflectance spectroscopic data. The PLSR is a linear multiple regression method that combines and generalizes the characteristics of multiple regression and principal component regression. PLSR is a transformer and a regressor that works similarly to PCR in that it reduces the dimensionality of the samples before applying a linear regressor to the transformed data. The PLS transformation differs from PCR in that it is supervised. PLSR was carried out in Unscrambler X (version 10.4) software CAMO, Norway after being spectra imported. The spectra were transformed by applying Savitzky-Golay smoothing with 1^{st} derivative and 7 smoothing points. Table 4.7 shows the descriptive statistics from PLSR modelling along with calibration (R^2 values and RMSE) and validation (R^2 values, RMSEP, SD, and RPD).

4.7 Random Forest Regression

It's a machine learning technique for solving classification and regression problems. It uses ensemble learning approach for solving complicated problems that integrate multiple classifiers. Many decision trees make up a random forest algorithm. The random forest regression models were evaluated based on their values of R^2 (R-

Table 4.6. SMLR modelling using hyperspectral data.

Soil Attributes	Stepwise Multiple Linear Regression	R^2	R^2 (Adj)
Sand (%)	$91.5 + 45.9(1227) - 49(1239) + 10.3(2225) - 8.15(2338)$	0.85	0.79
Silt (%)	$0.72 + 4.10(1272) - 7.98(2152) - 1.64(2284) + 7.50(2338)$	0.71	0.68
Clay (%)	$15.07 - 0.49(1084) + 2.44(2007) - 6.92(2072) + 4.84(2079)$	0.51	0.42
OM (%)	$0.48 - 0.02(588) - 0.15(937) + 0.83(938) - 0.65(940)$	0.38	0.29
P ($mgkg^{-1}$)	$2.37 - 0.37(400) - 3.45(937) + 12.4(938) - 8.84(940)$	0.28	0.19
K ($mgkg^{-1}$)	$9.76 + 23.8(561) - 454.9(588) + 427.4(590) + 3.65(1961)$	0.20	0.13

square), RMSE (Root Mean Square Error), MAE (Mean Absolute Error). Table 4.8 shows the statistical results obtained while RF regression is applied. Variable importance for soil properties is shown in section 4.11.

4.8 Geospatial Interpolation for Soil Mapping

Interpolation techniques such as IDW (Inverse Distance Weighted), kriging and spline are well-known. The general kriging equation was described in section 3.11.2. Soil physicochemical characteristics were first investigated using IDW, kriging and spline, and accuracy was assessed using RMSE (Root Mean Square Error) with known values for each sampled point based on the anticipated raster. The RMSE obtained for IDW was the lowest among all the three kriging approaches. Hence, IDW was found to be more accurate than others. The RMSE values of each character are explained in Table 4.9. IDW maps of each soil variable is shown in figure 4.2 and 4.3.

4.9 Validation of Soil Modeling

4.9.1 Soil Physicochemical Properties using MLR

Evaluation of the models was done based on Highest R^2 and lowest RMSE. Table 4.10 shows the R^2, Adjusted R^2. According to the results, the highest R^2 value is obtained for organic matter with R^2 value of 0.23 and the adjusted R^2 value of 0.17. Potassium and phosphorus both had R^2 value of 0.22 with adjusted R^2 value of 0.15. Among soil particle, the sand had R^2 value of 0.19 with adjusted R^2 value of 0.05. Silt had R^2 value of 0.17 and clay with the lowest R^2 value of 0.12. Hence, all of the soil properties didn't showed significant results with MLR using Sentinel 2 data.

4.9.2 Soil Physicochemical Properties using Exploratory Regression and OLS

Adjusted R^2, AIC (Akaike Information Criteria), JB (Jarque-Bera), VIF (Variance Inflation Factor), and SA (Spatial Autocorrelation (the Global Moran's I p-value)) are used to evaluate the model while running exploratory regression. A decent model has a high Adjusted R^2, a low AIC, (JB-nonsignificant), VIF 10 OR 7, and SA (NO pattern should exist). Results obtained through exploratory regression is shown in table 4.4 in section 4.3.2. Spatial regression models (equations) are shown in table 4.5 while model evaluation is discussed in the section 4.9.2. Results have shown that the K (potassium) had R^2 value of 0.19 with adjusted R^2 value of 0.15, P (phosphorus) had R^2 value of 0.10 with adjusted R^2 value of 0.07 and sand and R^2 value of 0.11 with adjusted R^2 value of 0.07. Among other parameters these three have somehow highest R^2 value and adjusted R^2 value.

4.9.3 Validation of SMLR using ASD Field Spec 4 Dataset

The SMLR procedure was carried out on hyperspectral data as discussed in section 4.5. The model's accuracy was validated based on R^2 and adjusted R^2. Table 4.12 shows the descriptive statistics obtained through SMLR being applied on the ASD Field Spec 4 dataset. According to the statistics, it can be seen that sand with the R^2 of 0.85 and adjusted R^2 of 0.79, silt with R^2 of 0.71 and adjusted R^2 of 0.68 and clay with R^2 of 0.51 and adjusted R^2 of 0.42 show significant results while OM (%) shows R^2 of 0.38 and adjusted R^2 of 0.29, P ($mgkg^{-1}$) R^2 of 0.28 and adjusted R^2 of 0.19 and K ($mgkg^{-1}$) R^2 of 0.20 and adjusted R^2 of 0.13 does not found to be significant.

4.10 Validation of PLSR using ASD Field Spec 4 Dataset

The PLSR model validation be done through RPD values obtained for each soil property. The residual prediction deviation (RPD) value helps predict soil properties

Table 4.7. Descriptive statistics from PLSR modeling.

Soil Attribute	Calibration Dataset		Validation Dataset			
	R^2	$RMSE_c$	R^2	$RMSE_p$	SD	RPD
Sand (%)	0.90	1.92	0.88	5.58	9.18	1.64
Silt (%)	0.89	2.38	0.79	4.70	9.81	2.08
Clay (%)	0.83	2.93	0.64	5.60	2.41	0.43
OM (%)	0.56	0.11	0.43	5.20	0.12	0.02
P ($mgkg^{-1}$)	0.68	1.02	0.78	3.22	1.18	0.3
K ($mgkg^{-1}$)	0.72	18.9	0.52	10.2	26.4	2.58

Table 4.8. Statistical results using random forest regression.

Soil Property	R^2	RMSE	MAE
Sand (%)	0.76	0.01	0.008
Silt (%)	0.71	0.06	0.03
Clay (%)	0.73	0.02	0.015
OM (%)	0.76	0.026	0.019
P ($mgkg^{-1}$)	0.79	0.03	0.02
K ($mgkg^{-1}$)	0.71	0.04	0.03

using Vis-NIR spectra. RPD values were grouped into three categories such as category A or excellent where RPD value will be greater than 2; category B or good with RPD values ranging from 1.4 to 2, and category C or unreliable with RPD less than 0.5. It is calculated by dividing standard deviation (SD) of the reference dataset to the root mean square error of prediction (RMSEP) as shown in equation

$$RPD = \frac{SD}{RMSEP} \qquad Equation\ 4.1$$

Hence, results show that sand, silt and potassium predictions are good (category B) while clay, OM and phosphorus predicted are unreliable as they lie in category C.

4.11 Variable Importance of Soil Physicochemical Properties using Random Forest Regression

Variable importance of each soil property is shown in figure 4.4 and 4.5 as obtained through random forest regression analysis. The Variable Importance chart shows each explanatory variable's importance in the regression model. As for clay Band-12 has the most significance, for sand has band-5, for silt has blue band, for potassium green band, band-12 is for organic matter, and band-7 for phosphorus is the most significant band.

Table 4.9. Evaluation of soil physicochemical characteristics using spatial interpolation.

Soil Attributes	RMSE		
	IDW(Inverse Distance Weighted Interpolation)	Kriging	Spline
Sand (%)	0.03	8.16	0.34
Silt (%)	0.03	8.61	0.33
Clay (%)	0.00	2.38	0.00
OM (%)	0.00	0.01	0.00
P ($mgkg^{-1}$)	0.00	0.83	0.03
K ($mgkg^{-1}$)	0.07	20.62	0.83

Table 4.10. Descriptive statistics of MLR.

Soil Property	R^2	Adjusted R^2
Sand (%)	0.19	0.05
Clay (%)	0.12	0.05
Silt (%)	0.17	0.10
OM (%)	0.23	0.17
P ($mgkg^{-1}$)	0.22	0.15
K ($mgkg^{-1}$)	0.22	0.15

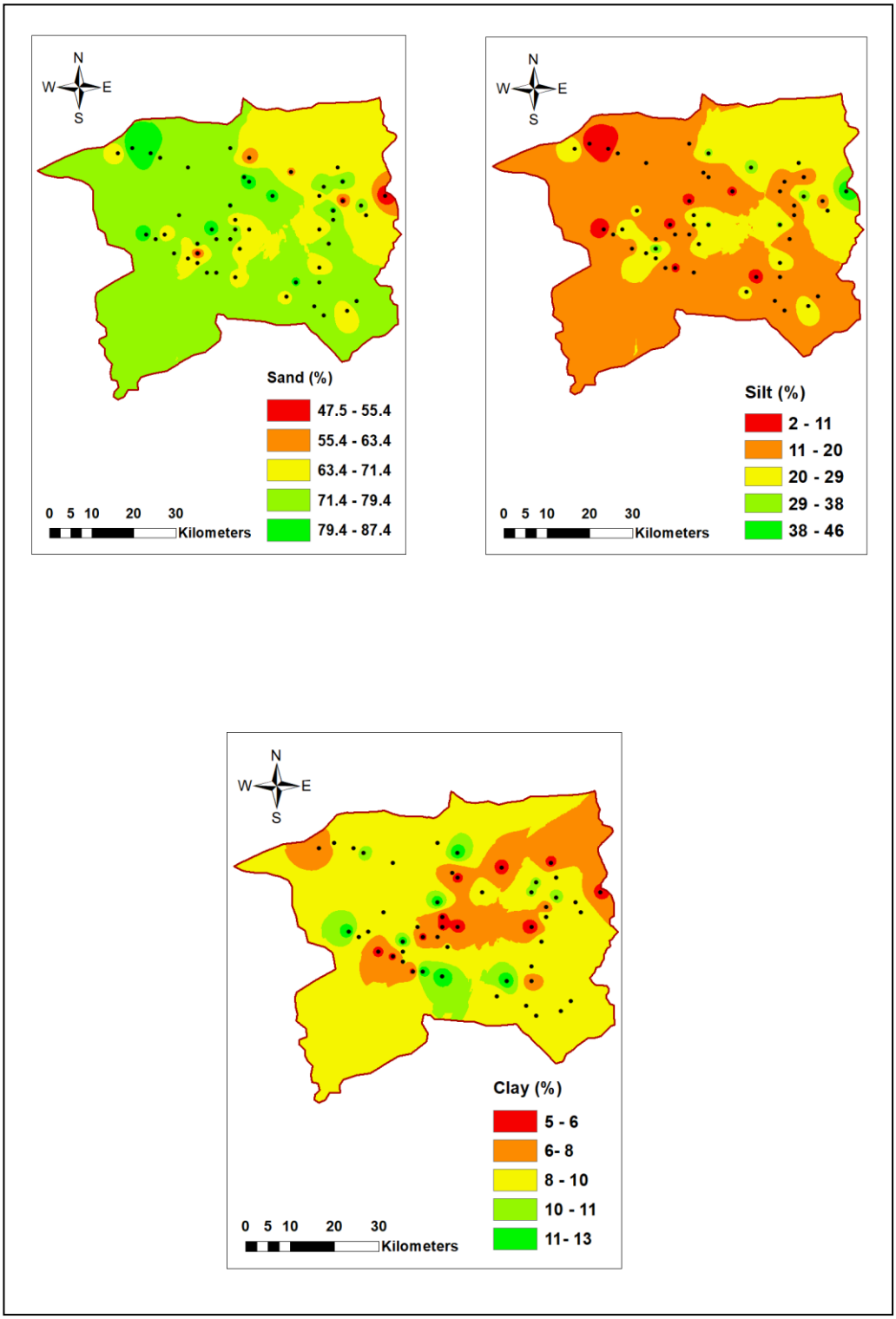

Figure 4.2. Inverse distance interpolation maps of sand, silt and clay.

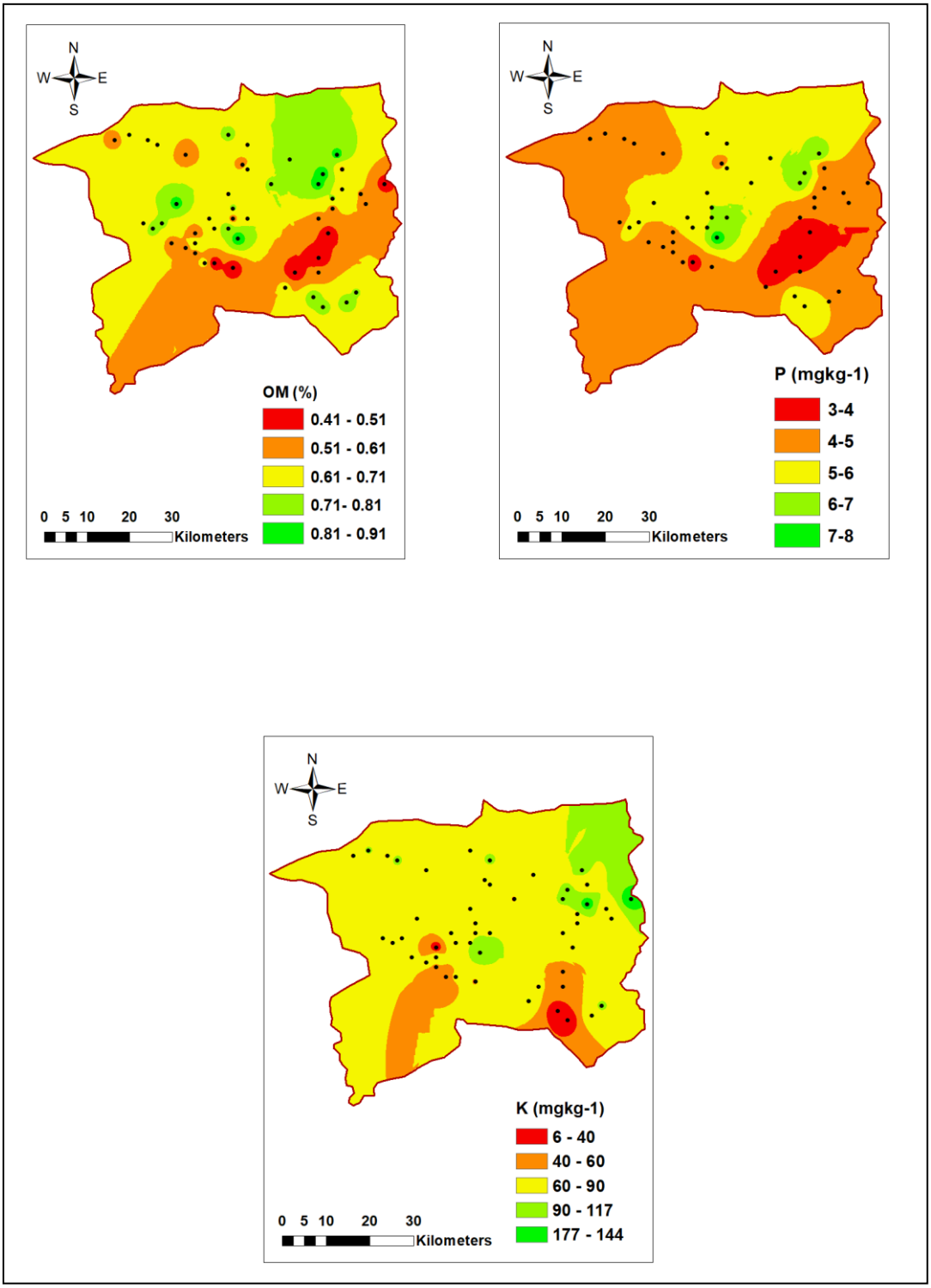

Figure 4.3. Inverse distance interpolation maps of OM, P and K.

Table 4.11. Statistics of OLS modelling.

Soil Attribute	R^2	Adjusted R^2	AIC	Histogram
Sand (%)	0.11	0.07	372.6	Normally distributed
Silt (%)	0.01	0.009	425	Normally distributed
Clay (%)	0.08	0.05	237.5	Normally distributed
OM (%)	0.05	0.03	-66.0	Normally distributed
P ($mgkg^{-1}$)	0.10	0.07	163.8	Normally distributed
K ($mgkg^{-1}$)	0.19	0.15	475.7	Normally distributed

Table 4.12. Descriptive statistics of SMLR on ASD Field Spec 4.

Soil Attribute	R^2	Adjusted R^2
Sand (%)	0.85	0.79
Silt (%)	0.71	0.68
Clay (%)	0.51	0.42
OM (%)	0.38	0.29
P ($mgkg^{-1}$)	0.28	0.19
K ($mgkg^{-1}$)	0.20	0.13

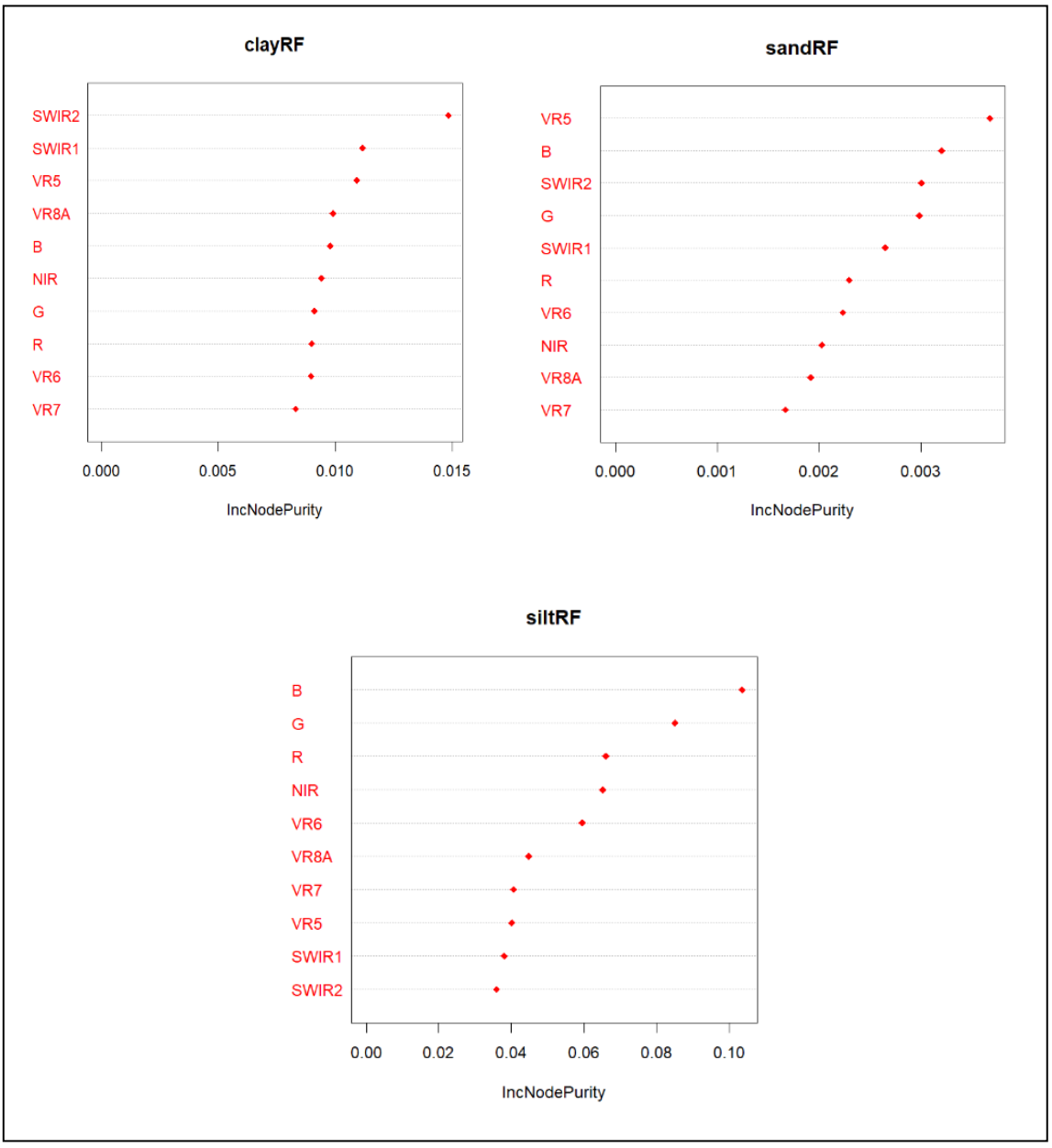

Figure 4.4. Variable importance of sand, silt and clay.

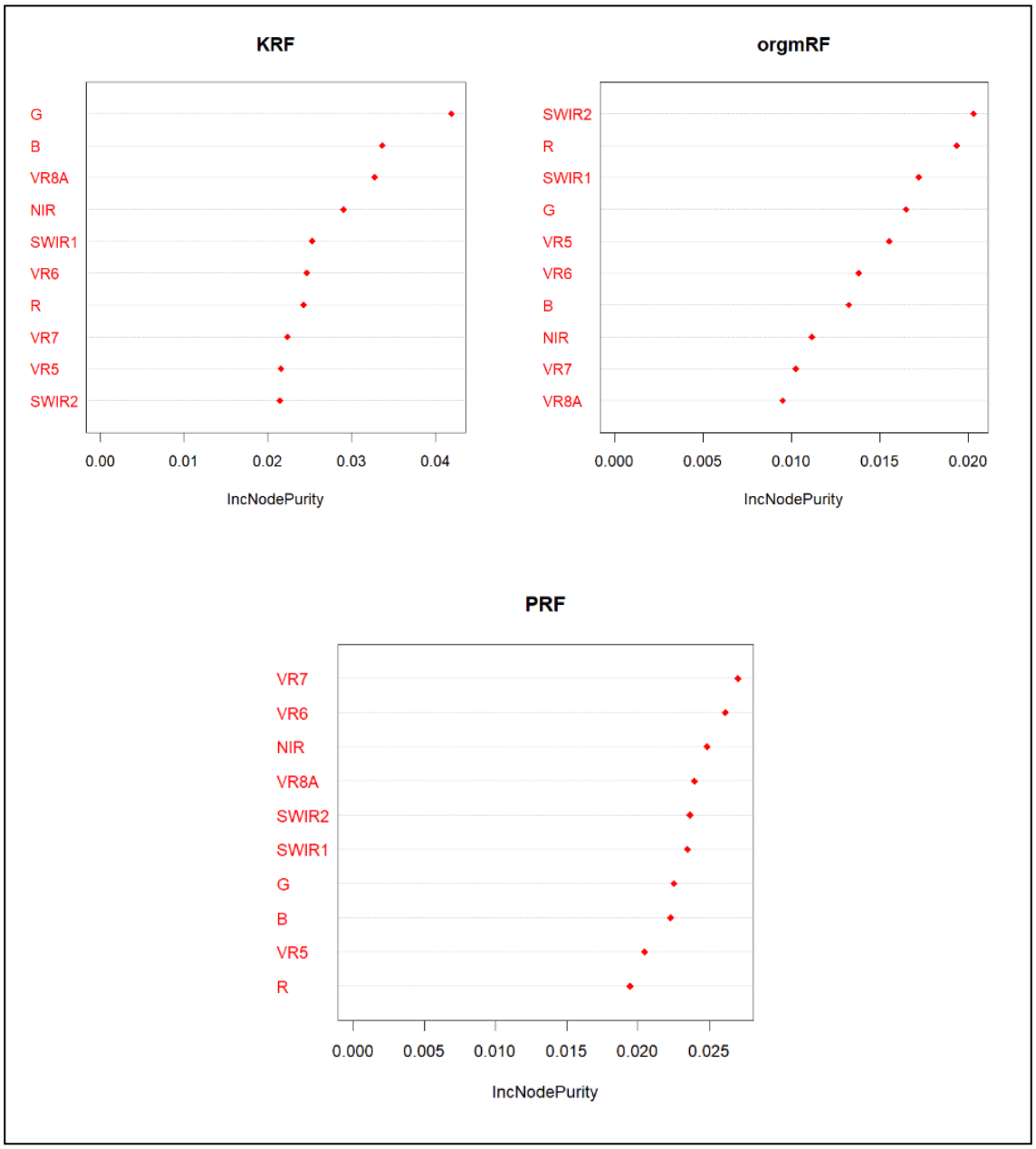

Figure 4.5. Variable importance of soil chemical properties.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

Soil physical properties were predicted through classical statistics, geospatial statistics, machine learning techniques such as 'Random Forest' while hyperspectral dataset is also used for prediction to know which performs the best. The classical statistics didn't meet the demand because it does not account for spatial variability as the spatial data has spatial dependency. Hence geospatial regression modelling is used to encounter this issue, such as exploratory regression and ordinary least squares regression and they performed better compared to the classical statistics.

The texture of study area is majorly composed of sandy loam, loamy sand and loamy soils. Soil physico-chemical properties revealed that soil pH, electrical conductivity (EC), organic matter (OM) and potassium (K) (6-145 $mgkg^{-1}$) had a normal range while soil phosphorus (3-8 $mgkg^{-1}$) was found deficient (<10 $mgkg^{-1}$) in soils of the study area. A significant relationship was found between OM and P which indicate that OM mineralization is releasing P in soil. So, we need to improve the OM content to reduce the P addition as an inorganic fertilizer. Sentinel 2A dataset didn't showed satisfactory predictions for almost all of the soil parameters using both MLR (multiple linear regression) and OLS (ordinary least squares regression). As the sand had R^2 value of 0.19 and adjusted R^2 value of 0.05, silt had R^2 value of 0.17 and adjusted R^2 value of 0.10, clay had R^2 value of 0.12 and adjusted R^2 value of 0.05, OM (organic matter) with R^2 value of 0.23 and adjusted R^2 value of 0.17, P and K both obtained R^2 value of 0.22 and adjusted R^2 value of 0.15. Similarly, non-significant results are also obtained with OLS regression as well. On the Sentinel-2A data, the somehow better and significant results are obtained using random forest regression technique. Sand with the R^2 value of 0.76, OM (%) with R^2 value of 0.76 and phosphorus with R^2 value of 0.79. ASD Field Spec 4 data had also

been modelled with soil parameters using SMLR and PLSR. SMLR has shown better predictions with soil properties but the accuracy was improved using PLSR. Sand using SMLR obtained R^2 value of 0.85 and it increased upto 0.88 (R^2 value of prediction) using PLSR, similarly, highest R^2 values were obtained for the silt, phosphorus and clay as well. Interpolation methods such as IDW, kriging and spline were used to predict different soil properties in which the IDW performed the best. It is concluded that among datasets, better results are obtained using hyperspectral datasets.

5.1 Recommendations

1. Traditional soil analysis methods are expensive, time-consuming so there is a need of inexpensive and accurate methods for soil mapping, this is where the remote sensing methods come in handy for site-specific management and soil variability mapping.
2. Diffuse Reflectance Spectroscopy can be effectively used for modelling soil properties.
3. Geospatial statistics and Artificial Neural Network models should be used for soil

   modelling.
4. High spectral and spatial resolution satellite images such as hyperspectral images be used to improve the accuracy.
5. Soil sampling size must be increased to develop a more precise and accurate soil spectral library using ASD Field Spec 4.
6. For a sustainable environment, precision agriculture is needed for which national soil national database should be created.
7. Soil texture maps should be created as their information is essential for supporting the agronomic decisions on farm management.

REFERENCES

1. Abdel-Rahman, E. M., Ahmed, F. B., & Ismail, R. (2013). Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data. *International Journal of Remote Sensing*, *34*(2), 712–728. https://doi.org/10.1080/01431161.2012.713142

2. Abrougui, K., Gabsi, K., Mercatoris, B., Khemis, C., Amami, R., & Chehaibi, S. (2019). Prediction of organic potato yield using tillage systems and soil properties by artificial neural network (ANN) and multiple linear regressions (MLR). *Soil and Tillage Research*, *190*, 202–208. https://doi.org/10.1016/j.still.2019.01.011

3. Afroj, M., Kazal, M. M. H., & Rahman, M. M. (2016). Precision agriculture in the World and its prospect in Bangladesh. *Research in Agriculture Livestock and Fisheries*, *3*(1), 1–14. https://doi.org/10.3329/ralf.v3i1.27853

4. Agbu, P. A., Fehrenbacher, D. J., & Jansen, I. J. (1990). Soil Property Relationships with SPOT Satellite Digital Data in East Central Illinois. *Soil Science Society of America Journal*, *54*(3), 807–812. https://doi.org/10.2136/sssaj1990.03615995005400030031x

5. Ahmad, S., Kalra, A., & Stephen, H. (2010). Estimating soil moisture using remote sensing data: A machine learning approach. *Advances in Water Resources*, *33*(1), 69–80. https://doi.org/10.1016/j.advwatres.2009.10.008

6. Ahmed, Z., & Iqbal, J. (2014). Evaluation of Landsat TM5 multispectral data for automated mapping of surface soil texture and organic matter in GIS. *European Journal of Remote Sensing*, *47*(1), 557–573. https://doi.org/10.5721/EuJRS20144731

7. Ajaj, Q. M., Shareef, M. A., Hassan, N. D., Hasan, S. F., & Noori, A. M. (2018). GIS based spatial modeling to mapping and estimation relative risk of different diseases using inverse distance weighting (IDW) interpolation algorithm and evidential belief function (EBF) (Case study: Minor Part of Kirkuk City, Iraq). *International Journal of Engineering and Technology(UAE)*, *7*(4), 185–191. https://doi.org/10.14419/ijet.v7i4.37.24098

8. Al-Quraishi, A. M. F., Sadiq, H. A., & Messina, J. P. (2019). Characterization and modeling surface soil physicochemical properties using Landsat images: A case study in the Iraqi Kurdistan region. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *42*(2/W16), 21–28. https://doi.org/10.5194/isprs-archives-XLII-2-W16-21-2019

9. Angelopoulou, T., Balafoutis, A., Zalidis, G., & Bochtis, D. (2020). From laboratory to proximal sensing spectroscopy for soil organic carbon estimation-A review. *Sustainability (Switzerland)*, *12*(2). https://doi.org/10.3390/su12020443

10. Angers, D. A., & Eriksen-Hamel, N. S. (2008). Full-Inversion Tillage and Organic Carbon Distribution in Soil Profiles: A Meta-Analysis. *Soil Science Society of America Journal*, *72*(5), 1370–1374. https://doi.org/10.2136/sssaj2007.0342

11. Azadi, S., & Karimi-Jashni, A. (2016). Verifying the performance of artificial neural network and multiple linear regression in predicting the mean seasonal municipal solid waste generation rate: A case study of Fars province, Iran. *Waste Management*, *48*, 14–23. https://doi.org/10.1016/j.wasman.2015.09.034

12. Bajwa, S. G., Mishra, A. R., & Norman, R. J. (2010). Canopy reflectance response to plant nitrogen accumulation in rice. *Precision Agriculture*, *11*(5), 488–506. https://doi.org/10.1007/s11119-009-9142-0

13. Balafoutis, A., Beck, B., Fountas, S., Vangeyte, J., Van Der Wal, T., Soto, I., Gómez-Barbero, M., Barnes, A., & Eory, V. (2017). Precision agriculture technologies positively contributing to ghg emissions mitigation, farm productivity and economics. *Sustainability (Switzerland)*, *9*(8), 1–28. https://doi.org/10.3390/su9081339

14. Bangelesa, F., Adam, E., Knight, J., Dhau, I., Ramudzuli, M., & Mokotjomela, T. M. (2020). Predicting Soil Organic Carbon Content Using Hyperspectral Remote Sensing in a Degraded Mountain Landscape in Lesotho. *Applied and Environmental Soil Science*, *2020*. https://doi.org/10.1155/2020/2158573

15. Baumgardner, M. F., Silva, L. F., Biehl, L. L., & Stoner, E. R. (1985). Reflectance properties of soils. *Advances in Agronomy. Vol. 38*, 1–44.

16. Behrens, T., Förster, H., Scholten, T., Steinrücken, U., Spies, E. D., & Goldschmitt, M. (2005). Digital soil mapping using artificial neural networks. *Journal of Plant Nutrition and Soil Science*, *168*(1), 21–33. https://doi.org/10.1002/jpln.200421414

17. Ben-Dor, E., & Banin, A. (1994). Visible and near-infrared (0.4-1.1 μm) analysis of arid and semiarid soils. *Remote Sensing of Environment*, *48*(3), 261–274. https://doi.org/10.1016/0034-4257(94)90001-9

18. Ben-Dor, E., & Banin, A. (1995). Near-Infrared Analysis as a Rapid Method to Simultaneously Evaluate Several Soil Properties. *Soil Science Society of America Journal*, *59*(2), 364–372. https://doi.org/10.2136/sssaj1995.03615995005900020014x

19. Beuselinck, L., Govers, G., Poesen, J., Degraer, G., & Froyen, L. (1998). Grain-size analysis by laser diffractometry: Comparison with the sieve-pipette method. *Catena*, *32*(3–4), 193–208. https://doi.org/10.1016/S0341-8162(98)00051-4

20. Bhunia, G. S., Kumar Shit, P., & Pourghasemi, H. R. (2019). Soil organic carbon mapping using remote sensing techniques and multivariate regression model. *Geocarto International*, *34*(2), 215–226. https://doi.org/10.1080/10106049.2017.1381179

21. Bouyoucos, G. J. (1927). The hydrometer as a new method for the mechanical analysis of soils. *Soil Science*, *23*(5), 343–353. https://doi.org/10.1097/00010694-192705000-00002

22. Bremmer y Mulvaney. (1982). Methods of soil analysis. Part 2. Chemical and microbiological properties (Agronomy 9). *American Society of Agronomy, Soil*

Science Society of America., 595–634.

23. Curcio, D., Ciraolo, G., D'Asaro, F., & Minacapilli, M. (2013). Prediction of Soil Texture Distributions Using VNIR-SWIR Reflectance Spectroscopy. *Procedia Environmental Sciences*, *19*, 494–503. https://doi.org/10.1016/j.proenv.2013.06.056

24. Dalal, R. C., & Henry, R. J. (1986). Simultaneous Determination of Moisture, Organic Carbon, and Total Nitrogen by Near Infrared Reflectance Spectrophotometry. *Soil Science Society of America Journal*, *50*(1), 120–123. https://doi.org/10.2136/sssaj1986.03615995005000010023x

25. Davis, E., Wang, C., & Dow, K. (2019). Comparing Sentinel-2 MSI and Landsat 8 OLI in soil salinity detection: a case study of agricultural lands in coastal North Carolina. *International Journal of Remote Sensing*, *40*(16), 6134–6153. https://doi.org/10.1080/01431161.2019.1587205

26. Denton, O. A., Aduramigba-Modupe, V. O., Ojo, A. O., Adeoyolanu, O. D., Are, K. S., Adelana, A. O., Oyedele, A. O., Adetayo, A. O., & Oke, A. O. (2017). Assessment of spatial variability and mapping of soil properties for sustainable agricultural production using geographic information system techniques (GIS). *Cogent Food and Agriculture*, *3*(1). https://doi.org/10.1080/23311932.2017.1279366

27. Draper, N. R., Smith, H., & Pownell, E. (1966). Applied regression analysis, vol 3 John Wiley & Sons. In *Inc., New York, NY [Google Scholar]*.

28. E. Birch, A. N., Begg, G. S., & Squire, G. R. (2011). How agro-ecological research helps to address food security issues under new IPM and pesticide reduction policies for global crop production systems. *Journal of Experimental Botany*, *62*(10), 3251–3261. https://doi.org/10.1093/jxb/err064

29. Efroymson, M. A. (1960). Multiple Regression Analysis: Mathematical Methods for Digital Computers. *New York: Wiley*, 191–203.

30. El-Sirafy, Z. M., El-Ghamry, A. M., Elnaggar, A. A., & El-Seedy, M. E. (2011). Spatial Distribution of Some Soil Physiochemical Properties in Kalabshow Farm Using Geostatistics. *Journal of Soil Sciences and Agricultural Engineering*, *2*(12), 1291–1301. https://doi.org/10.21608/jssae.2011.56475

31. EPA. (2002). Methods for the Determination of Total Organic Carbon (Toc) in Soils and Sediments. *Carbon*, *32*(April), 25.

32. Esfandiarpoor Borujeni, I., Mohammadi, J., Salehi, M. H., Toomanian, N., & Poch, R. M. (2010). Assessing geopedological soil mapping approach by statistical and geostatistical methods: A case study in the Borujen region, Central Iran. *Catena*, *82*(1), 1–14. https://doi.org/10.1016/j.catena.2010.03.006

33. Fabiani, S., Vanino, S., Napoli, R., Zajíček, A., Duffková, R., Evangelou, E., & Nino, P. (2020). Assessment of the economic and environmental sustainability of Variable Rate Technology (VRT) application in different wheat intensive European agricultural areas. A Water energy food nexus approach. *Environmental*

Science and Policy, *114*, 366–376. https://doi.org/10.1016/j.envsci.2020.08.019

34. Farifteh, J., Farshad, A., & George, R. J. (2006). Assessing salt-affected soils using remote sensing, solute modelling, and geophysics. *Geoderma*, *130*(3–4), 191–206. https://doi.org/10.1016/j.geoderma.2005.02.003

35. Forkuor, G., Hounkpatin, O. K. L., Welp, G., & Thiel, M. (2017). High resolution mapping of soil properties using Remote Sensing variables in south-western Burkina Faso: A comparison of machine learning and multiple linear regression models. *PLoS ONE*, *12*(1), 1–21. https://doi.org/10.1371/journal.pone.0170478

36. Franke, R. (1982). Scattered data interpolation: tests of some methods. *Mathematics of Computation*, *38*(157), 181–200. https://doi.org/10.1090/s0025-5718-1982-0637296-4

37. Ge, Y., Thomasson, J. A., & Sui, R. (2011). Remote sensing of soil properties in precision agriculture: A review. *Frontiers of Earth Science*, *5*(3), 229–238. https://doi.org/10.1007/s11707-011-0175-0

38. Gebbers, R., & Adamchuk, V. I. (2010). Precision agriculture and food security. *Science*, *327*(5967), 828–831. https://doi.org/10.1126/science.1183899

39. Gholizadeh, A., Amin, M. S. M., Borůvka, L., & Saberioon, M. M. (2014). Models for Estimating the Physical Properties of Paddy Soil Using Visible and Near Infrared Reflectance Spectroscopy. *Journal of Applied Spectroscopy*, *81*(3), 534–540. https://doi.org/10.1007/s10812-014-9966-x

40. Godinho Silva, S. H., Poggere, G. C., de Menezes, M. D., Carvalho, G. S., Guilherme, L. R. G., & Curi, N. (2016). Proximal sensing and digital terrain models applied to digital soil mapping and modeling of Brazilian Latosols (Oxisols). *Remote Sensing*, *8*(8). https://doi.org/10.3390/rs8080614

41. Gorji, T., Yildirim, A., Hamzehpour, N., Tanik, A., & Sertel, E. (2020). Soil salinity analysis of Urmia Lake Basin using Landsat-8 OLI and Sentinel-2A based spectral indices and electrical conductivity measurements. *Ecological Indicators*, *112*. https://doi.org/10.1016/j.ecolind.2020.106173

42. Grunwald, S., Vasques, G. M., & Rivero, R. G. (2015). Fusion of soil and remote sensing data to model soil properties. *Advances in Agronomy*, *131*, 1–109. https://doi.org/10.1016/bs.agron.2014.12.004

43. Hanquet, B., Sirjacobs, D., Destain, M. F., Frankinet, M., & Verbrugge, J. C. (2004). Analysis of soil variability measured with a soil strength sensor. *Precision Agriculture*, *5*(3), 227–246. https://doi.org/10.1023/B:PRAG.0000032763.54104.b4

44. Hassoun, M. H., Intrator, N., McKay, S., & Christian, W. (1996). Fundamentals of Artificial Neural Networks. *Computers in Physics*, *10*(2), 137. https://doi.org/10.1063/1.4822376

45. Hatfield, J. L. (2000). *Precision Agriculture and Environmental Quality*. *July*, 1–23.

46. Hocking, R. R. (1976). The Analysis and Selection of Variables in Linear Regression. *Biometrics*, *32*(1), 1. http://www.jstor.org/stable/2529336?origin=crossref

47. Kalota, D. (2017). Exploring relation of land surface temperature with selected variables using geographically weighted regression and ordinary least square methods in Manipur State, India. *Geocarto International*, *32*(10), 1105–1119. https://doi.org/10.1080/10106049.2016.1195883

48. Khaledian, Y., & Miller, B. A. (2020). Selecting appropriate machine learning methods for digital soil mapping. *Applied Mathematical Modelling*, *81*, 401–418. https://doi.org/10.1016/j.apm.2019.12.016

49. Khan, S. A., Mulvaney, R. L., Ellsworth, T. R., & Boast, C. W. (2007). The Myth of Nitrogen Fertilization for Soil Carbon Sequestration. *Journal of Environmental Quality*, *36*(6), 1821–1832. https://doi.org/10.2134/jeq2007.0099

50. Khanal, S., Fulton, J., & Shearer, S. (2017). An overview of current and potential applications of thermal remote sensing in precision agriculture. *Computers and Electronics in Agriculture*, *139*, 22–32. https://doi.org/10.1016/j.compag.2017.05.001

51. Khanna, A., & Kaur, S. (2019). Evolution of Internet of Things (IoT) and its significant impact in the field of Precision Agriculture. *Computers and Electronics in Agriculture*, *157*, 218–231. https://doi.org/10.1016/j.compag.2018.12.039

52. King, C., Baghdadi, N., Lecomte, V., & Cerdan, O. (2005). The application of remote-sensing data to monitoring and modelling of soil erosion. *Catena*, *62*(2–3), 79–93. https://doi.org/10.1016/j.catena.2005.05.007

53. Krishna, K. R. (2003). Soil fertility and crop production. *Choice Reviews Online*, *40*(05), 40-2794-40–2794. https://doi.org/10.5860/choice.40-2794

54. Kristof, S. J. (1971). Preliminary multispectral studies of soils. *Journal of Soil and Water Conservation*, *26*, 15–18.

55. Lazaar, A., Pradhan, B., Naiji, Z., Gourfi, A., El Hammouti, K., Andich, K., & Monir, A. (2021). The manifestation of VIS-NIRS spectroscopy data to predict and map soil texture in the Triffa plain (Morocco). *Kuwait Journal of Science*, *48*(1), 111–121. https://doi.org/10.48129/KJS.V48I1.8012

56. Lee, S., Hahn, C., Rhee, M., Oh, J. E., Song, J., Chen, Y., Lu, G., Perdana, & Fallis, A. . (2012). Relationship between Soil Physical Properties and Crop Production. *Journal of Chemical Information and Modeling*, *53*(9), 1689–1699. http://dx.doi.org/10.1016/j.tws.2012.02.007

57. Lerch, R. N., Kitchen, N. R., Kremer, R. J., Donald, W. W., Alberts, E. E., Sadler, E. J., Sudduth, K. A., Myers, D. B., & Ghidey, F. (2005). Development of a conservation-oriented precision agriculture system: Water and soil quality assessment. *Journal of Soil and Water Conservation*, *60*(6), 411–421.

58. Liao, K., Xu, S., Wu, J., & Zhu, Q. (2013). Spatial estimation of surface soil texture using remote sensing data. *Soil Science and Plant Nutrition*, *59*(4), 488–500. https://doi.org/10.1080/00380768.2013.802643

59. Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, *2*(3), 18–22.

60. Lin, C., Zhu, A. X., Wang, Z., Wang, X., & Ma, R. (2020). The refined spatiotemporal representation of soil organic matter based on remote images fusion of Sentinel-2 and Sentinel-3. *International Journal of Applied Earth Observation and Geoinformation*, *89*. https://doi.org/10.1016/j.jag.2020.102094

61. López-Granados, F., Jurado-Expósito, M., Peña-Barragán, J. M., & García-Torres, L. (2005). Using geostatistical and remote sensing approaches for mapping soil properties. *European Journal of Agronomy*, *23*(3), 279–289. https://doi.org/10.1016/j.eja.2004.12.003

62. Lu, Y. L., Bai, Y. L., Yang, L. P., Wang, L., Wang, Y. L., Ni, L., & Zhou, L. P. (2015). Hyper-spectral characteristics and classification of farmland soil in northeast of China. *Journal of Integrative Agriculture*, *14*(12), 2521–2528. https://doi.org/10.1016/S2095-3119(15)61232-1

63. Martens, H., & Naes, T. (1989). Multivariate Calibration John Wiley & Sons. *Chichester, UK*.

64. Massimo Conforti, & Gabriele Buttafuoco. (2014). Vis-NIR Spectroscopy for Determining Physical and Chemical Soil Properties: An Application to an Area of Southern Italy. *Global Journal of Agricultural Innovation, Research & Development*, *1*(1), 17–26. https://doi.org/10.15377/2409-9813.2014.01.01.3

65. Mirchooli, F., Kiani-Harchegani, M., Khaledi Darvishan, A., Falahatkar, S., & Sadeghi, S. H. (2020). Spatial distribution dependency of soil organic carbon content to important environmental variables. *Ecological Indicators*, *116*(March 2019), 106473. https://doi.org/10.1016/j.ecolind.2020.106473

66. Mohamed, E. S., Saleh, A. M., Belal, A. B., & Gad, A. A. (2018). Application of near-infrared reflectance for quantitative assessment of soil properties. *Egyptian Journal of Remote Sensing and Space Science*, *21*(1), 1–14. https://doi.org/10.1016/j.ejrs.2017.02.001

67. Montzka, S. A., Dlugokencky, E. J., & Butler, J. H. (2011). Non-CO 2 greenhouse gases and climate change. *Nature*, *476*(7358), 43–50. https://doi.org/10.1038/nature10322

68. Morellos, A., Pantazi, X. E., Moshou, D., Alexandridis, T., Whetton, R., Tziotzios, G., Wiebensohn, J., Bill, R., & Mouazen, A. M. (2016). Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. *Biosystems Engineering*, *152*, 104–116. https://doi.org/10.1016/j.biosystemseng.2016.04.018

69. Mueller, T. G., Pusuluri, N. B., Mathias, K. K., Cornelius, P. L., Barnhisel, R. I., & Shearer, S. A. (2004). Map Quality for Ordinary Kriging and Inverse Distance

Weighted Interpolation. *Soil Science Society of America Journal*, *68*(6), 2042–2047. https://doi.org/10.2136/sssaj2004.2042

70. Mzuku, M., Khosla, R., & Reich, R. (2015). Bare Soil Reflectance to Characterize Variability in Soil Properties. *Communications in Soil Science and Plant Analysis*, *46*(13), 1668–1676. https://doi.org/10.1080/00103624.2015.1043463

71. Nawar, S., Buddenbaum, H., & Hill, J. (2015). Digital mapping of soil properties using multivariate statistical analysis and ASTER data in an Arid region. *Remote Sensing*, *7*(2), 1181–1205. https://doi.org/10.3390/rs70201181

72. Oshunsanya, S. O., Oluwasemire, K. O., & Taiwo, O. J. (2017). Use of GIS to Delineate Site-Specific Management Zone for Precision Agriculture. *Communications in Soil Science and Plant Analysis*, *48*(5), 565–575. https://doi.org/10.1080/00103624.2016.1270298

73. P. Leone, A., A. Viscarra-Rossel, R., Amenta, P., & Buondonno, A. (2012). Prediction of Soil Properties with PLSR and vis-NIR Spectroscopy: Application to Mediterranean Soils from Southern Italy. *Current Analytical Chemistry*, *8*(2), 283–299. https://doi.org/10.2174/157341112800392571

74. Panday, D., Maharjan, B., Chalise, D., Kumar Shrestha, R., & Twanabasu, B. (2018). Digital soil mapping in the Bara district of Nepal using kriging tool in ArcGIS. *PLoS ONE*, *13*(10). https://doi.org/10.1371/journal.pone.0206350

75. Pellegrini, E., Rovere, N., Zaninotti, S., Franco, I., De Nobili, M., & Contin, M. (2021). Artificial neural network (ANN) modelling for the estimation of soil microbial biomass in vineyard soils. *Biology and Fertility of Soils*, *57*(1), 145–151. https://doi.org/10.1007/s00374-020-01498-1

76. Peng, J., Biswas, A., Jiang, Q., Zhao, R., Hu, J., Hu, B., & Shi, Z. (2019). Estimating soil salinity from remote sensing and terrain data in southern Xinjiang Province, China. *Geoderma*, *337*, 1309–1319. https://doi.org/10.1016/j.geoderma.2018.08.006

77. Plant, R. E., Pettygrove, G. S., & Reinert, W. R. (2000). Precision agriculture can increase profits and limit environmental impacts. *California Agriculture*, *54*(4), 66–71. https://doi.org/10.3733/ca.v054n04p66

78. Potopová, V., Trnka, M., Hamouz, P., Soukup, J., & Castraveț, T. (2020). Statistical modelling of drought-related yield losses using soil moisture-vegetation remote sensing and multiscalar indices in the south-eastern Europe. *Agricultural Water Management*, *236*. https://doi.org/10.1016/j.agwat.2020.106168

79. Preissler, H., & Loercher, G. (1995). Extraction of soil properties from laboratory and imaging spectrometry data. *International Geoscience and Remote Sensing Symposium (IGARSS)*, *3*, 1968–1970. https://doi.org/10.1109/igarss.1995.524081

80. Qiao, P., Lei, M., Yang, S., Yang, J., Guo, G., & Zhou, X. (2018). Comparing ordinary kriging and inverse distance weighting for soil as pollution in Beijing. *Environmental Science and Pollution Research*, *25*(16), 15597–15608. https://doi.org/10.1007/s11356-018-1552-y

81. Robinson, T. P., & Metternicht, G. (2006). Testing the performance of spatial interpolation techniques for mapping soil properties. *Computers and Electronics in Agriculture*, *50*(2), 97–108. https://doi.org/10.1016/j.compag.2005.07.003

82. Sadeghi, M., Jones, S. B., & Philpot, W. D. (2015). A linear physically-based model for remote sensing of soil moisture using short wave infrared bands. *Remote Sensing of Environment*, *164*, 66–76. https://doi.org/10.1016/j.rse.2015.04.007

83. Schimmelpfennig, D., & Ebel, R. (2016). Sequential adoption and cost savings from precision agriculture. *Journal of Agricultural and Resource Economics*, *41*(1), 97–115.

84. Schloeder, C. A., Zimmerman, N. E., & Jacobs, M. J. (2001). Comparison of Methods for Interpolating Soil Properties Using Limited Data. *Soil Science Society of America Journal*, *65*(2), 470–479. https://doi.org/10.2136/sssaj2001.652470x

85. Setianto, A., & Triandini, T. (2015). Comparison of Kriging and Inverse Distance Weighted (Idw) Interpolation Methods in Lineament Extraction and Analysis. *Journal of Applied Geology*, *5*(1), 21–29. https://doi.org/10.22146/jag.7204

86. Sheng, J., Yu, P., Zhang, H., & Wang, Z. (2021). Spatial variability of soil Cd content based on IDW and RBF in Fujiang River, Mianyang, China. *Journal of Soils and Sediments*, *21*(1), 419–429. https://doi.org/10.1007/s11368-020-02758-1

87. Short, N. M. (1982). *The Landsat tutorial workbook: basics of satellite remote sensing*. https://doi.org/10.2307/635039

88. Silvero, N. E. Q., Demattê, J. A. M., Amorim, M. T. A., Santos, N. V. dos, Rizzo, R., Safanelli, J. L., Poppiel, R. R., Mendes, W. de S., & Bonfatti, B. R. (2021). Soil variability and quantification based on Sentinel-2 and Landsat-8 bare soil images: A comparison. *Remote Sensing of Environment*, *252*. https://doi.org/10.1016/j.rse.2020.112117

89. Simmonds, S. E., Kinlan, B. P., White, C., Paradis, G. L., Warner, R. R., & Zacherl, D. C. (2014). Geospatial statistics strengthen the ability of natural geochemical tags to estimate range-wide population connectivity in marine species. *Marine Ecology Progress Series*, *508*, 33–51. https://doi.org/10.3354/meps10871

90. Sims, J. T. (2000). Soil test phosphorus: Olsen P. *Methods of Phosphorus Analysis for Soils , Sediments , Residuals , and Waters*, 20–21.

91. Skoog, D. A., Holler, F. J., & Crouch, S. R. (2007). *Principles of Instrumental Analysis. Thomson Brooks. Cole, Canada.*

92. Smith, J. L., & Doran, J. W. (2015). Measurement and Use of pH and Electrical Conductivity for Soil Quality Analysis. In *Methods for Assessing Soil Quality* (pp. 169–185). https://doi.org/10.2136/sssaspecpub49.c10

93. Srisomkiew, S., Kawahigashi, M., & Limtong, P. (2021). Digital mapping of soil

chemical properties with limited data in the Thung Kula Ronghai region, Thailand. *Geoderma*, *389*. https://doi.org/10.1016/j.geoderma.2021.114942

94. Stenberg, B., Viscarra Rossel, R. A., Mouazen, A. M., & Wetterlind, J. (2010). Visible and Near Infrared Spectroscopy in Soil Science. *Advances in Agronomy*, *107*(C), 163–215. https://doi.org/10.1016/S0065-2113(10)07005-7

95. Su, S., Jiang, Z., Zhang, Q., & Zhang, Y. (2011). Transformation of agricultural landscapes under rapid urbanization: A threat to sustainability in Hang-Jia-Hu region, China. *Applied Geography*, *31*(2), 439–449. https://doi.org/10.1016/j.apgeog.2010.10.008

96. Sulyman, M., Noori, A., & Al-Attar, A. (2020). *Study and GIS-Based Mapping of Soil Chemical Properties in Kirkuk City, Iraq*. *January*. https://doi.org/10.4108/eai.28-6-2020.2298163

97. Tokekar, P., Hook, J. Vander, Mulla, D., & Isler, V. (2016). Sensor Planning for a Symbiotic UAV and UGV System for Precision Agriculture. *IEEE Transactions on Robotics*, *32*(6), 1498–1511. https://doi.org/10.1109/TRO.2016.2603528

98. Tsai, F., & Philpot, W. (1998). Derivative analysis of hyperspectral data. *Remote Sensing of Environment*, *66*(1), 41–51. https://doi.org/10.1016/S0034-4257(98)00032-7

99. Vågen, T. G., Winowiecki, L. A., Tondoh, J. E., Desta, L. T., & Gumbricht, T. (2014). Mapping of soil properties and land degradation risk in Africa using MODIS reflectance. *Geoderma*, *263*, 216–225. https://doi.org/10.1016/j.geoderma.2015.06.023

100. Vasileios C. Drosos, & Chrisvaladis Malesios. (2012). Measuring the Accuracy and Precision of the Garmin GPS Positioning in Forested Areas: A Case Study in Taxiarchis-Vrastama University Forest. *Journal of Environmental Science and Engineering B*, *1*(4). https://doi.org/10.17265/2162-5263/2012.04.015

101. Vaudour, E., Gomez, C., Fouad, Y., & Lagacherie, P. (2019). Sentinel-2 image capacities to predict common topsoil properties of temperate and Mediterranean agroecosystems. *Remote Sensing of Environment*, *223*, 21–33. https://doi.org/10.1016/j.rse.2019.01.006

102. Ver Hoef, J. M., & Cressie, N. (1993). Multivariable spatial prediction. *Mathematical Geology*, *25*(2), 219–240. https://doi.org/10.1007/BF00893273

103. Vibhute, A. D., Kale, K. V., Mehrotra, S. C., Dhumal, R. K., & Nagne, A. D. (2018). Determination of soil physicochemical attributes in farming sites through visible, near-infrared diffuse reflectance spectroscopy and PLSR modeling. *Ecological Processes*, *7*(1). https://doi.org/10.1186/s13717-018-0138-4

104. Vohland, M., Besold, J., Hill, J., & Fründ, H. C. (2011). Comparing different multivariate calibration methods for the determination of soil organic carbon pools with visible to near infrared spectroscopy. *Geoderma*, *166*(1), 198–205. https://doi.org/10.1016/j.geoderma.2011.08.001

105. Volkan Bilgili, A., van Es, H. M., Akbas, F., Durak, A., & Hively, W. D. (2010). Visible-near infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey. *Journal of Arid Environments*, *74*(2), 229–238. https://doi.org/10.1016/j.jaridenv.2009.08.011

106. W. Siesler, Y. Ozaki, S. K. and H. M. H. (2002). Near-infrared spectroscopy. Principles, instruments, applications. In *Journal of Chemometrics*. https://books.google.be/books?id=U7vqrf2YqmcC&pg=PA122&lpg=PA122&dq=sample+cells+for+nir+spectroscopy&source=bl&ots=R0o36S_erH&sig=ACfU3U2ImtGblrIDYgY0j2tSXgTMSiyrKQ&hl=nl&sa=X&ved=2ahUKEwjW5fuBiOXoAhWnsaQKHZX-AlU4ChDoATAFegQICxA1#v=onepage&q=sample cells

107. Waldrop, M. P., Zak, D. R., Sinsabaugh, R. L., Gallo, M., & Lauber, C. (2004). Nitrogen deposition modifies soil carbon storage through changes in microbial enzymatic activity. *Ecological Applications*, *14*(4), 1172–1177. https://doi.org/10.1890/03-5120

108. Wang, B., Waters, C., Orgill, S., Gray, J., Cowie, A., Clark, A., & Liu, D. L. (2018). High resolution mapping of soil organic carbon stocks using remote sensing variables in the semi-arid rangelands of eastern Australia. *Science of the Total Environment*, *630*, 367–378. https://doi.org/10.1016/j.scitotenv.2018.02.204

109. Wang, J., Ding, J., Yu, D., Ma, X., Zhang, Z., Ge, X., Teng, D., Li, X., Liang, J., Lizaga, I., Chen, X., Yuan, L., & Guo, Y. (2019). Capability of Sentinel-2 MSI data for monitoring and mapping of soil salinity in dry and wet seasons in the Ebinur Lake region, Xinjiang, China. *Geoderma*, *353*, 172–187. https://doi.org/10.1016/j.geoderma.2019.06.040

110. Watson, D. F. (1992). Contouring: a guide to the analysis and display of spatial data. *Contouring: A Guide to the Analysis and Display of Spatial Data*. https://doi.org/10.1016/0098-3004(93)90069-h

111. Zhang, Y., Guo, L., Chen, Y., Shi, T., Luo, M., Ju, Q. L., Zhang, H., & Wang, S. (2019). Prediction of soil organic carbon based on Landsat 8 monthly NDVI data for the Jianghan Plain in Hubei Province, China. *Remote Sensing*, *11*(14). https://doi.org/10.3390/rs11141683

112. Zhou, T., Geng, Y., Chen, J., Pan, J., Haase, D., & Lausch, A. (2020). High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms. *Science of the Total Environment*, *729*. https://doi.org/10.1016/j.scitotenv.2020.138244

113. Ziegel, E. R., Deutsch, C. V., & Journel, A. G. (1998). Geostatistical Software Library and User's Guide. *Technometrics*, *40*(4), 357. https://doi.org/10.2307/1270548

APPENDICES

Appendix-1: Soil Physical Properties

S-No.	Texture	Name	Latitude	Longitude	Sand(%)	Silt(%)	Clay(%)
1	***Sandy Loam***	Dhoke Patwala	32.78	72.32	81	5.5	13.5
2	***Sandy Loam***	Shah Muhammadi	33.05	71.95	65	28	7
3	***Sandy Loam***	Dhoke Sherjangal	32.75	72.30	70	21.5	8.5
4	***Loamy Sand***	Dhoke Tali	33.06	71.97	87.5	4	8.5
5	***Loamy Sand***	Pichnand	32.88	72.00	85	1.5	13.5
6	***Loamy Sand***	Multan khurd	33.05	72.01	87.5	3	9.5
7	***Sandy Loam***	Khuiiaan	33.04	72.03	72.5	17	10.5
8	***Sandy Loam***	Tamman	33.02	72.09	79	11.5	9.5
9	***Loamy Sand***	Bhudiyaal	33.00	72.21	80	13	7
10	***Sandy Loam***	Kot Shams	32.87	72.21	71.5	20	8.5
11	***Sandy Loam***	Sukka	32.88	72.04	69	22.5	8.5
12	***Sandy Loam***	Dhoke Chokera	32.96	72.37	70	19.5	10.5
13	***Sandy Loam***	Dhakli Jasiyaal	32.98	72.38	75	14.5	10.5
14	***Sandy Loam***	Leti	32.92	72.07	71.5	20	8.5
15	***Sandy Loam***	Kot Sarang	33.02	72.41	65	29	6
16	***Loamy Sand***	Dhrada	32.84	72.06	76.5	17.5	6
17	***Loamy Sand***	Mustafa Abad	32.99	72.42	81	9.5	9.5

Cont.

S-No.	Texture	Name	Latitude	Longitude	Sand(%)	Silt(%)	Clay(%)
18	***Sandy Loam***	Balwaal	32.83	72.09	67.5	26.5	6
19	***Sandy Loam***	Naka Kahut	32.94	72.46	75	16.5	8.5
20	***Sandy Loam***	Dhoke Mangiyaal	32.92	72.46	71	19.5	9.5
21	***Sandy Loam***	Dhurnaal	32.82	72.11	71.5	20	8.5
22	***Sandy Loam***	Malakwaal	32.91	72.40	71.5	19	9.5
23	***Sandy Loam***	Pira Fatiyaal	32.89	72.37	65	30	5
24	***Loamy Sand***	Dhoke Fatehkhel	32.80	72.13	80	14	6
25	***Sandy Loam***	Jhatla	32.86	72.39	79	11.5	9.5
26	***Sandy Loam***	Dhoke Jamaal	32.80	72.15	79	7.5	13.5
27	***Sandy Loam***	Thoyaa	32.81	72.37	65	25.5	9.5
28	***Sandy Loam***	Dhoke Hawapur	32.79	72.19	70	16.5	13.5
29	***Loamy Sand***	Dhoke Agraal	32.84	72.11	49	42.5	8.5
30	***Sandy Loam***	Khichiaan	32.78	72.37	76.5	16.5	7
31	***Sandy Loam***	Dhoke Miaal	32.86	72.11	75	11.5	13.5
32	***Sandy Loam***	Chingii	32.73	72.36	75	15.5	9.5
33	***Sandy Loam***	Kutehraa	32.71	72.38	72.5	18	9.5
34	***Sandy Loam***	Bhulumaar	32.72	72.43	65	25.5	9.5
35	***Loamy Sand***	Qadarpur	32.74	72.45	80	11.5	8.5

Cont.

S-No.	Texture	Name	Latitude	Longitude	Sand(%)	Silt(%)	Clay(%)
36	***Loamy Sand***	Muhammad Abad	32.93	72.40	82.5	10.5	7
37	***Loamy Sand***	Dhoke Sherdastal	32.87	72.15	77.5	16.5	6
38	***Loam***	Tae	32.95	72.42	51.5	37	11.5
39	***Sandy Loam***	Dhraabi	32.96	72.51	47.5	46.5	6
40	***Loamy Sand***	Dhoke Musahib	32.89	72.14	86.5	5	8.5
41	***Loamy Sand***	Gojwaal	32.87	72.18	80	11.5	8.5
42	***Sandy Loam***	Dhulii	32.85	72.20	66.5	25	8.5
43	***Sandy Loam***	Wanharr	32.89	72.19	74	20	6
44	***Sandy Loam***	Dhoke Fakeeran	32.89	72.22	64	30	6
45	***Sandy Loam***	Darrot	32.91	72.19	65	29	6
46	***Sandy Loam***	Dhermond	32.94	71.18	80	6.5	13.5
47	***Sandy Loam***	Moglay	33.01	72.31	62.5	31.5	6
48	***Sandy Loam***	Misriyaal	33.04	72.22	56	30.5	13.5
49	***Sandy Loam***	Patwaal	33.06	72.18	74	17.5	8.5
50	***Loamy Sand***	Sanghwala	32.99	72.22	82.5	11.5	6
51	***Loamy Sand***	Saghar	32.96	72.27	81.5	9	9.5

Appendix-2: Soil Chemical Properties

S-No.	Name	pH	EC	OM (%)	P (mg/Kg)	K(mg/Kg)
1	Dhoke Patwala	7.74	0.76	0.41	3	63
2	Shah Muhammadi	7.75	0.84	0.60	4	63
3	Dhoke Sherjangal	7.86	0.76	0.64	4	91
4	Dhoke Tali	7.84	0.84	0.63	4	91
5	Pichnand	7.87	0.76	0.64	4	91
6	Multan khurd	7.85	0.76	0.64	5	87
7	Khuiiaan	7.77	0.76	0.68	5	91
8	Tamman	7.72	0.76	0.55	4	63
9	Bhudiyaal	7.74	0.76	0.54	4	63
10	Kot Shams	7.81	0.76	0.76	6	63
11	Sukka	7.83	0.76	0.73	5	67
12	Dhoke Chokera	7.79	1.01	0.87	7	91
13	Dhakli Jasiyaal	7.77	0.93	0.87	7	91
14	Leti	7.79	0.93	0.86	6	87
15	Kot Sarang	7.78	1.10	0.85	7	95
16	Dhrada	7.92	0.76	0.57	4	63
17	Mustafa Abad	7.74	0.76	0.60	4	63
18	Balwaal	7.76	0.76	0.58	4	63
19	Naka Kahut	7.73	0.84	0.61	4	63
20	Dhoke Mangiyaal	7.74	0.76	0.59	4	67
21	Dhurnaal	7.75	0.76	0.60	4	59
22	Malakwaal	7.73	0.84	0.58	4	63
23	Pira Fatiyaal	7.75	0.76	0.60	4	63
24	Dhoke Fatehkhel	7.79	0.84	0.64	5	35
25	Jhatla	7.63	0.76	0.44	3	63
26	Dhoke Jamaal	7.65	0.76	0.43	3	67
27	Thoyaa	7.64	0.76	0.43	3	61
28	Dhoke Hawapur	7.64	0.76	0.44	4	61

Cont.

S-No.	Name	pH	EC	OM (%)	P (mg/Kg)	K(mg/Kg)
29	Dhoke Agraal	7.81	0.84	0.72	5	89
30	Khichiaan	8.15	0.93	0.53	4	61
31	Dhoke Miaal	8.13	0.93	0.51	5	7
32	Chingii	7.63	0.93	0.78	6	6
33	Kutehraa	7.61	1.44	0.73	6	9
34	Bhulumaar	7.63	1.44	0.74	5	85
35	Qadarpur	7.62	1.52	0.72	5	93
36	Muhammad Abad	7.62	1.35	0.73	5	89
37	Dhoke Sherdastal	7.73	1.44	0.71	5	85
38	Tae	7.72	0.84	0.67	4	145
39	Dhraabi	7.70	0.93	0.46	4	145
40	Dhoke Musahib	7.81	1.10	0.69	6	89
41	Gojwaal	7.83	0.84	0.78	6	89
42	Dhulii	7.69	1.18	0.92	8	117
43	Wanharr	7.85	0.84	0.55	6	89
44	Dhoke Fakeeran	7.84	0.93	0.71	6	89
45	Darrot	7.84	0.93	0.73	6	89
46	Dhermond	7.85	0.93	0.71	6	89
47	Moglay	7.85	0.76	0.73	6	85
48	Misriyaal	7.83	0.84	0.71	6	93
49	Patwaal	7.86	0.93	0.73	6	89
50	Sanghwala	7.83	0.84	0.71	6	61
51	Saghar	7.83	0.93	0.72	6	61