

# **ASSESSMENT OF SPATIAL PATTERN OF SOIL SALINITY IN COASTAL AGRICULTURAL AREAS USING MULTI-SENSOR APPROACH**



Sarah Joey P. Salgado  
MS Spatial Engineering

June 2021

# **ASSESSMENT OF SPATIAL PATTERN OF SOIL SALINITY IN COASTAL AGRICULTURAL AREAS USING MULTI-SENSOR APPROACH**

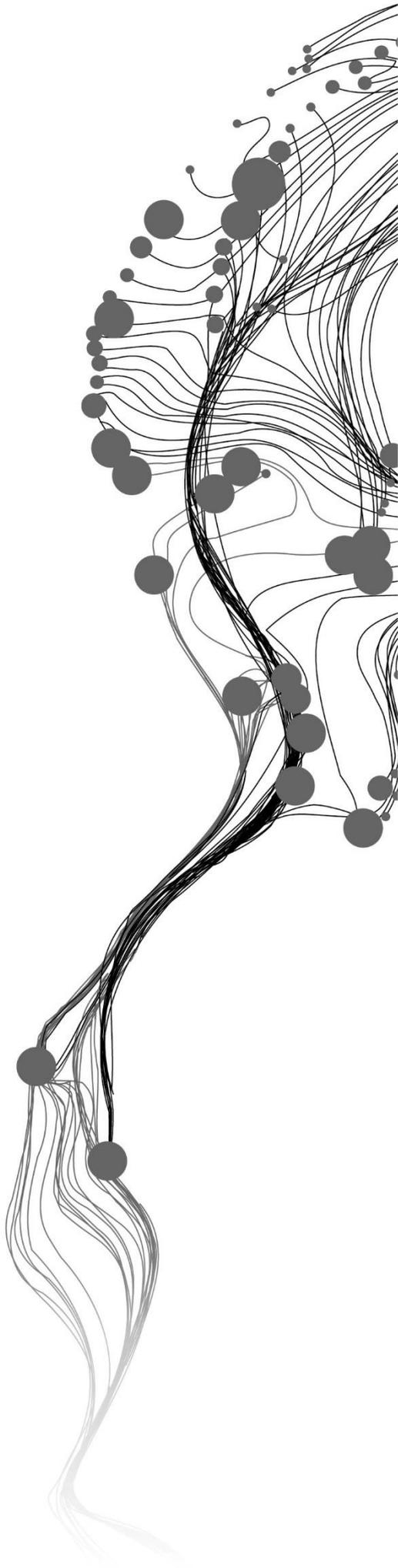
SARAH JOEY P. SALGADO

June, 2021

SUPERVISORS:

prof. dr. A. D. Nelson

dr. Y. Zeng



# **ASSESSMENT OF SPATIAL PATTERN OF SOIL SALINITY IN COASTAL AGRICULTURAL AREAS USING MULTI-SENSOR APPROACH**

SARAH JOEY P. SALGADO

Enschede, The Netherlands, June, 2021

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Spatial Engineering Specialization: FORAGES

## **SUPERVISORS:**

prof. dr. A.D. Nelson

dr. Y. Zeng

## **THESIS ASSESSMENT BOARD:**

dr. ir. C.A.J.M de Bie, Chair

dr. Renaud Mathieu, External Examiner

drs. R.G. Nijmeijer, Procedural Advisor

#### DISCLAIMER

This document describes work undertaken as part of a programme of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

## ABSTRACT

Soil salinization is the process where water-soluble salts accumulate in the soil. It is considered as one of the most expensive soil degradation problems due to its high spatial and temporal variability. In the Philippines, there is no updated map available at the time of this study about the location, extent, and severity of areas affected by soil salinity. This research utilizes remote sensing data and ground-truth data in a machine learning algorithm (Random Forests regression) to detect and retrieve soil salinity in the rice areas of the Province of Ilocos Sur, Philippines. The province was selected as the study area due to increasing levels of soil pH that lead to salt accumulation. In addition, over-pumping of groundwater for irrigation and household use, salt-making, and extension of agricultural areas to non-suitable areas are the human activities that exacerbate this phenomenon. Moreover, the province is also one of the target areas that the Department of Agriculture – Bureau of Soils and Water Management plans to update soil salinity.

The main objective of this study is to develop a method to detect the spatial pattern of soil salinity effectively. Furthermore, this study is geared up to recommend strategies for the stakeholder to combat and alleviate the adverse effects of soil salinization in the coastal rice areas of the Province of Ilocos Sur, Philippines. To achieve the research objective, satellite data from Sentinel-1, Sentinel-2, and Landsat-8 were obtained. Twenty features from Sentinel-1 were generated comprising Gamma-nought and Gray Level Co-occurrence Matrix (GLCM) bands in VV and VH polarization. Seventeen bands of vegetation and salinity indices were calculated from Sentinel-2, and the land surface temperature from the thermal bands of Landsat-8 was derived. Ancillary data composed of soil properties, climate, and geographical coordinates were also used as input for the Random Forests regression. Furthermore, two different approaches were adopted: optimized multi-sensor predictors, where the predictor variables are the collection of those variables within the importance threshold from the three different sensors individually, and multi-sensor predictors in which all variables from three different sensors are used to select the predictor variables based on the importance threshold.

Accuracy assessment shows that the multi-sensor predictors method was better in detecting and retrieving soil salinity. The RMSE of this method is 0.15,  $R^2$  of 0.82, and Pearson correlation coefficient of 0.91, which indicates the excellent performance of this model. Spatial variability of predicted soil salinity shows that the coastal rice areas have higher soil salinity levels than those near mountains. Multiyear variability of soil salinity was predicted in 2017, 2018, and 2020, showing an increase of soil salinity in the area ranging from 0.04 to 0.14 decisiemens per meter (dS/m) in four years. It was found out that the integration of the three sensors is more efficient and more accurate in detecting and retrieving soil salinity than using a single sensor.

## ACKNOWLEDGEMENT

This scientific research is hard work and made possible with the support of the following persons. First of all, I would like to express my greatest appreciation to my supervisors, prof. dr. Andy Nelson, Chairman of the Department of Natural Resources, and dr. Yijian Zeng from the Department of Water Resources. Without their valuable advice, enthusiasm, and dynamism throughout the study period, this thesis would not have been completed and written successfully;

I would also like to express my gratitude to Willem Nieuwenhuis, ITC Developer from the Department of Natural Resources, for his help and guidance in the system settings, setting up the Python environment, and modifying my script for this study. Also to ir. Bas Restios, ITC Developer from the Department of Water Resources, who shared his experience and information on processing satellite data; to our course coordinator, drs. Tiny Luiten, MBA, for her advice, never-ending kindness, and encouragement in my study, especially during the milestones of my research;

Many thanks for the Thesis Assessment Board – chaired by dr. ir. Kees de Bie for their thorough criticism, valuable advice, and suggestions that helped in the conduct of this study;

In times of pandemic and homesickness, friends are always there with me. I wish to express my thankfulness to my friends and classmates in Spatial Engineering, primarily to Ranju Pote, Bisrat Araya, Fauzan Muzakki, and Kisanet Molla, whom I spent good, challenging days and made the research fun and interesting—extending my deepest gratitude to Kars van der Linden, for his love, patience, support, and understanding.

To my colleagues in the Department of Agriculture – Bureau of Soils and Water Management: former OIC-Director Angel C. Enriquez; OIC-Director Sonia M. Salguero; Mr. Dominciano D. Ramos, Jr., Division Chief; and all of the staff in Soils Survey Division for their assistance and encouragement throughout the research period;

I want to thank my family for their constant sustenance and encouragement and my late grandfather, who has been with me with every step of this memorable journey.

Lastly and above all, I am very indebted to our Lord God for His grace and blessings. In Him, I offer this incredible work.

# TABLE OF CONTENTS

---

<b>ABSTRACT .....</b>	<b>i</b>
<b>ACKNOWLEDGEMENT .....</b>	<b>ii</b>
<b>1. INTRODUCTION.....</b>	<b>2</b>
1.1 Soil Salinization and its impacts .....	2
1.2 Global issue of soil salinity.....	4
1.3 Soil salinity in the Philippines.....	4
1.4 Remote Sensing of Soil Salinity.....	5
1.5 Machine Learning Approaches .....	6
1.6 Random Forests Regression .....	9
1.7 Problem Formulation.....	9
1.7.1 Wicked problem .....	9
1.7.2 Research Objectives.....	10
1.7.3 Research Questions .....	10
1.7.4 Research Hypothesis.....	10
<b>2. DATA AND RESEARCH METHODS .....</b>	<b>11</b>
2.1 The Study Area .....	11
2.2 Materials, Data, and Software Used .....	15
2.3 Data Collection .....	15
2.3.1 Existing Data .....	15
2.3.2 Satellite Data .....	16
2.3.3 Ancillary Data .....	16
2.4 Processing of Satellite Data .....	17
2.4.1 Sentinel-1 SAR Data.....	17
2.4.2 Sentinel-2 MSI Data.....	18
2.4.3 Landsat-8 OLI Data .....	20
2.5 Processing of Ancillary Data .....	22
2.5.1 Soil properties, Climate and DEM.....	22
2.5.2 Geographical Location.....	22
2.5.3 Bare soil areas mask.....	22
2.5.4 Random sample points.....	22
2.5.5 Extracting pixel features.....	23
2.6 Modelling Soil Salinity .....	23
2.7 Accuracy Assessment.....	23
2.8 Soil Salinity Variability Analysis .....	24
2.8.1 Spatial Variability.....	24
2.8.2 Multiyear Variability.....	24
<b>3. RESULTS AND DISCUSSION .....</b>	<b>27</b>
3.1 Random Forests Regression .....	27
3.1.1 Individual Sensors .....	27
3.1.2 Optimized Multi-Sensor Predictors .....	30
3.1.3 Multi-sensor Predictors .....	33
3.2 Predicted Soil Salinity.....	36
3.3 Spatial Variability .....	40
3.4 Multiyear Variability .....	40

3.5 Soil Salinity Change .....	44
3.6 Management of salt-affected soils .....	44
3.6.1 For policy-making bodies .....	44
3.6.2 Farmers and farmer associations .....	47
3.7 Limitations of the Research and Future Work.....	47
<b>4. CONCLUSION AND RECOMMENDATIONS.....</b>	<b>49</b>
4.1 <i>What is the spatial distribution of soil salinity predicted from the variables generated from Sentinel-1, Sentinel-2, and Landsat-8?</i>	
49	
4.2 <i>What is the accuracy of using the multi-sensor approach in retrieving soil salinity?.....</i>	49
4.3 <i>To what extent and pattern do soil salinity change over multiple years based on the multi-sensor approach?.....</i>	50
<b>REFERENCES.....</b>	<b>51</b>
<b>APPENDICES .....</b>	<b>56</b>

## LIST OF FIGURES

---

Figure 1. Process of salt salinization (Hanson, 2011) .....	2
Figure 2. Schematic diagram of the factors affecting the soil salinity in agricultural lands .....	3
Figure 3. The sensitivity of rice in soil salinity (DA-PhilRice, 2011).....	3
Figure 4. Excess salts from Harmonized World Soil Database (Fischer et al., 2018) .....	4
Figure 5. Random Forests Regression.....	9
Figure 6. Study Area.....	12
Figure 7. General landscape cross-section of the study area .....	13
Figure 8. Field-measured soil salinity .....	14
Figure 9. Workflow for processing Sentinel-1 Product.....	18
Figure 10. Workflow for processing Sentinel-2 products .....	19
Figure 11. Workflow for deriving LST from Landsat-8.....	21
Figure 12. General workflow of the methods used in the study.....	26
Figure 13. List of feature importance from Sentinel-1 .....	27
Figure 14. Scatterplot of field-measured vs. predicted soil salinity from Sentinel-1.....	28
Figure 15. List of feature importance from Sentinel-2.....	28
Figure 16. Scatterplot of field-measured vs. predicted soil salinity from Sentinel-2.....	29
Figure 17. List of feature importance from Landsat-8 .....	29
Figure 18. Scatterplot of field-measured vs. predicted soil salinity from Landsat-8.....	30
Figure 19. List of feature importance for OMSP .....	30
Figure 20. Scatterplot of field-measured vs. predicted soil salinity from OMSP .....	33
Figure 21. List of input features and their importance values from MSP .....	34
Figure 22. List of Feature importance from MSP .....	35
Figure 23. Scatterplot of field-measured vs. predicted soil salinity for MSP .....	36
Figure 24. Predicted soil salinity map based on Optimized Multi-Sensor Predictors Approach .....	38
Figure 25. Predicted soil salinity map based on Multi-Sensor Predictors approach.....	39
Figure 26. Predicted soil salinity for years 2017 to 2020.....	41
Figure 27. Soil salinity change in most affected areas.....	43
Figure 28. Soil salinity change.....	46

## LIST OF TABLES

---

Table 1. Soil salinity classification for rice production (DA-BSWM, 2021 & FAO, 1976).....	4
Table 2. Studies related to soil salinity mapping.....	7
Table 3. Summary of data, types, and their sources.....	15
Table 4. Footprint of the study area to download the images.....	16
Table 5. General description of the ancillary data.....	16
Table 6. List of Sentinel-1 images.....	17
Table 7. List of predictors generated for Sentinel-1.....	17
Table 8. List of Sentinel-2 images.....	18
Table 9. Band names of Sentinel-2.....	18
Table 10. List of salinity indices generated from Sentinel-2 Data.....	19
Table 11. List of Landsat-8 images.....	20
Table 12. Metadata of thermal bands from Landsat-8 OLI Data.....	20
Table 13. Soil salinity levels for the study.....	24
Table 14. List of images used in multiyear analysis of soil salinity.....	24
Table 15. Statistical Metrics of OMSP Model.....	32
Table 16. Statistical Metrics of MSP Model.....	35

## LIST OF APPENDICES

---

Appendix A. Map of Rice Areas.....	56
Appendix B. Soils of the Area .....	57
Appendix C. Preprocessing of Sentinel-1 .....	59
Appendix D. Preprocessing of Sentinel-2 .....	61
Appendix E. Preprocessing of Landsat-8.....	62
Appendix F. Protocol on Soil Sampling .....	64
Appendix G. Python Script for Random Forests regression .....	66
Appendix H. Predicted Soil Salinity in Different Municipalities Using OMSP Model .....	73
Appendix I. Predicted Soil Salinity in Different Municipalities Using MSP Model.....	74

# 1. INTRODUCTION

## 1.1 Soil Salinization and its impacts

Soil salinization is a process where the water-soluble salt accumulates in the soil surface or subsurface (USDA Natural Resources Conservation Service, 1998) that reduces soil fertility (Panagos et al., 2012). The simplest way to describe how soil salinity occurs is when irrigation water that contains salt particles is applied to the agricultural area. A portion of water is infiltrated down the soil profile. The plants absorb some water through transpiration. Some water is evaporated into the atmosphere, leaving the salt particles in the soil, as shown in Figure 1. Soil salinization also occurs in areas with dry climates and low precipitation, and excessive salts are not flushed from the soil, removing deep-rooted vegetation and raising of the water table as the consequence, sea-level rise, and inappropriate use of fertilizers (FAO, 1986).

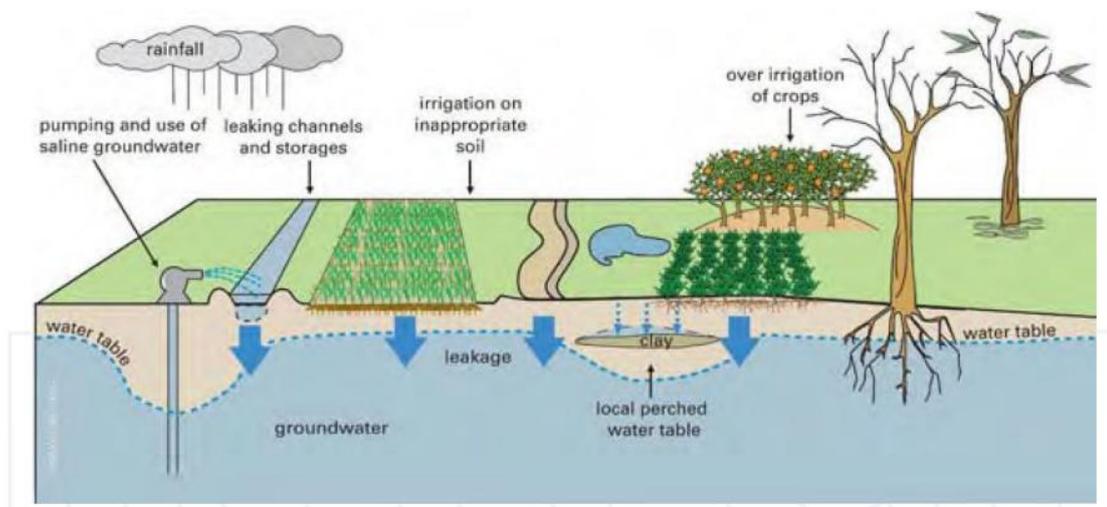


Figure 1. Process of salt salinization (Hanson, 2011)

Salinization affects the soil's physicochemical properties and ecological balance in the area (Shrivastava & Kumar, 2015) and is a major factor of land deterioration (Szabolcs, 1998), affecting crop production globally (FAO, 2019). Soil salinity is considered as one of the most expensive environmental hazards because of its high spatial and temporal variability (Zhang et al., 2015). Furthermore, around 20% of the cultivated land and 33% of the irrigated land is salt-affected and degraded in the world mainly because of human activities (FAO and ITPS, 2015).

Many anthropogenic factors contribute to soil salinization that affects the soil's water balance and energy flow, leading to severe salinization (Figure 2). These include improper irrigation methods, lack of drainage, agricultural intensification, deforestation, overgrazing, change in land use and cultivation patterns, depletion of freshwater layers, and chemical contamination brought by excessive application of mineral fertilizers (Szabolcs, 1998). Additionally, the consequences of climate change, such as rising sea levels, exacerbate this process, affecting farmers and local communities (FAO, 2019) due to saline intrusion. Moreover, soil salinity has become a severe problem for irrigated areas where irrigation often results in secondary salinization (Glick et al., 2007).

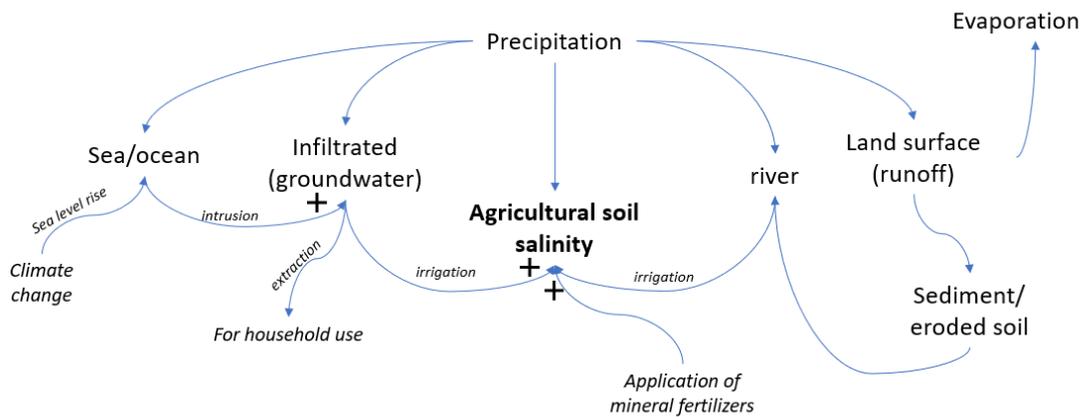


Figure 2. Schematic diagram of the factors affecting the soil salinity in agricultural lands

The presence of salt affects the crops' respiration and photosynthesis that is essential for crop growth by decreasing the biological  $N_2$  fixation and soil nitrogen mineralization (Dobermann & Fairhurst, 2000). For rice, 2.7 estimated million hectares of rice-growing areas have been constrained by soil salinity (Haefele et al., 2014). The rice plant is affected by salinity in each growth stage, though to different degrees, as shown in Figure 3. During germination, rice is tolerant, becomes very sensitive during the early seeding stage, becomes tolerant again during vegetative growth, returns to being very sensitive again during pollination and fertilization, and finally becomes tolerant at maturity to the harvesting stage (DA-PhilRice, 2011). Table 1 shows the soil salinity classification and its effects on the response of rice crops.

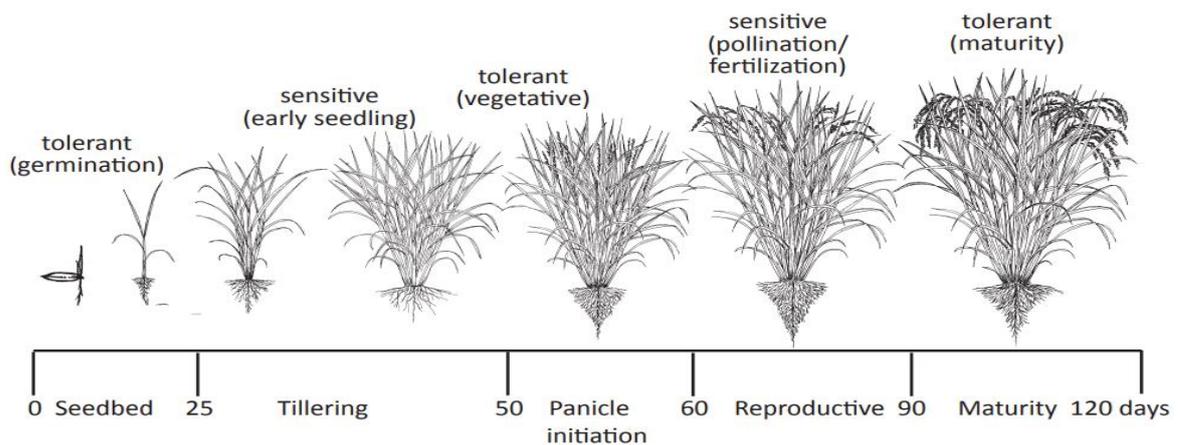


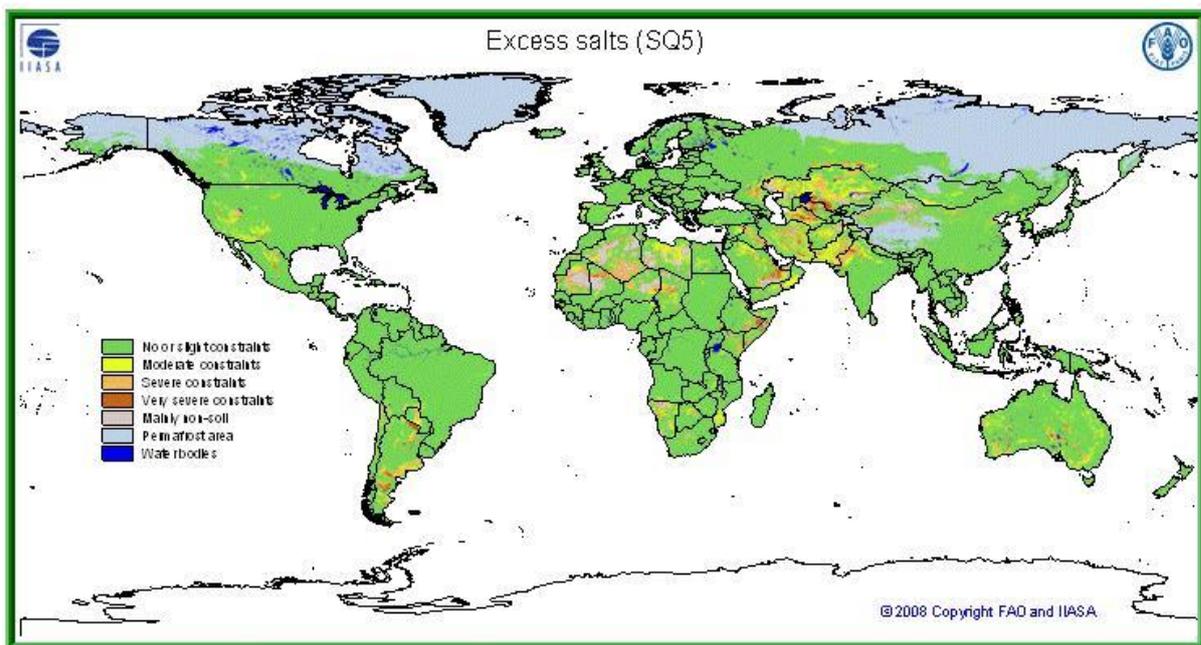
Figure 3. The sensitivity of rice in soil salinity (DA-PhilRice, 2011)

Table 1. Soil salinity classification for rice production (DA-BSWM, 2021 &amp; FAO, 1976)

Electrical conductivity, dS/m	Soil Salinity Class	Hazard for Rice Growth	Rice Response
0 - 2	Non-saline	Very low	Negligible
2.1 - 4	Slightly saline	Low	Restricted yield of sensitive crops
4.1 - 8	Moderately saline	Moderate	Restricted yield of many crops
8 - 16	Severely saline	High	Only a few tolerant crops yield satisfactorily
>16	Very severely saline	Very high	Only a few tolerant crops grow satisfactorily

## 1.2 Global issue of soil salinity

Soil salinity can occur in all climatic conditions and dynamically spreads globally in more than 100 countries, making no continent is free from this phenomenon. It is expected to increase due to rising sea levels, rising temperatures, and lower precipitation brought by climate change in the future (Shahid et al., 2018). It was assumed that salinization has expanded on the areas primarily reported as salt-affected areas during the 1970s and 1980s (Zamann et al., 2018). The Harmonized World Soil Database of the FAO Soils Portal provides an overview of the global salinity levels on a 1:5,000,000 scale (Fischer et al., 2018) is shown in Figure 4.



Further information on the supplementary datasets can be obtained from the Land Use Change and Agriculture Program, International Institute for Applied Systems Analysis (IIASA) • Schlossplatz 1 • A-2361 Laxenburg, Austria  
Phone: (+43 2236) 807 0 • Fax: (+43 2236) 71 313 • Web: <http://www.iiasa.ac.at>

Figure 4. Excess salts from Harmonized World Soil Database (Fischer et al., 2018)

## 1.3 Soil salinity in the Philippines

Assessing soil salinization is crucial for agricultural areas, especially in the Philippines, where recent countrywide information on soil salinity is dated 1986 (DA-PhilRice, 2011). Soil salinization in the country was observed in low-lying coastal areas due to salt-water intrusion and the use of water with salt particles to irrigate agricultural fields. An estimated

0.4 million hectares were saline-prone, and 0.2 million hectares were severely affected (Asio et al., 2009). This number is expected to increase due to overexploitation and uncontrolled groundwater extraction in the coastal areas for household and agriculture practices. The Department of Agriculture - Philippine Rice Research Institute (DA-PhilRice) reported that salinization had become a widespread constraint to rice production in most parts of the country (DA-PhilRice, 2011). Unfortunately, there is little opportunity to expand the rice areas as most of the lands surrounding it are being intended into another land use type such as built-up, industrial, and commercial areas expanding into agricultural areas.

Although saline areas in the Philippines are smaller than the other countries in Southeast Asia, they are still important as a potential production area for rice and other staple crops (DA-PhilRice, 2011). Additionally, according to the Department of Agriculture - Bureau of Soils and Water Management (DA-BSWM), based on the reconnaissance survey conducted in the 1950s, 45 provinces were affected by soil salinization. However, the extent, severity, and variability of salinization are still unknown and need an update (DA-BSWM, 2019). The DA-BSWM's Land Degradation and Assessment (Project LADA) conducted in 2011 to 2013 shows that an estimate of 758 hectares of irrigated agricultural lands experiences severe soil salinity in Regions I (Ilocos Region) and III (Central Luzon) (Carating, 2015).

#### **1.4 Remote Sensing of Soil Salinity**

Various studies and literature show different approaches in mapping and assessing soil salinity. Traditionally, conventional methods such as field-based measurements and soil analysis in laboratories are commonly used. However, these approaches are expensive, time-consuming, laborious, and unsuitable for temporal analysis (Allbed et al., 2014; Allbed & Kumar, 2013). Therefore, it is essential to recourse to effective mapping and monitoring through remote sensing technologies to keep track of soil salinity changes and anticipate further degradation. This section discusses how salt-affected soils are detected using remote sensing, which can map and monitor this phenomenon more efficiently and economically feasible.

In the 1990s to early 2000s, satellite data for assessing surface soil salinity ranges from multispectral to hyperspectral and microwave sensor data. Landsat TM, Landsat MSS, and SPOT XS multispectral sensor data have been used for soil salinity mapping in Western Nile Delta (Goosens et al., 1994). The addition of thermal bands added to visible-NIR bands was useful in extracting saline soils when there is complex spectral confusion (Singh et al., 2017). Landsat ETM+ has also been used (Band 1, 3, 4, and 7) but was ineffective in discriminating saline to non-saline areas due to spectral confusions (Verma et al., 1994). Moreover, small patches of soil surface where salinity was evident, implying that satellite images with low to medium resolution were not useful (Singh et al., 2017). IKONOS, Quickbird, and Worldview – 2 with a high spatial resolution (less than 5 meters) have been used for small patches of saline areas and were useful in differentiating various salinity classes at farm levels (El-Haddad & Garcia, 2006). The results from their studies suggest that higher resolution data has a great potential in studying soil salinity at farm levels. However, this has a higher cost of imagery data. Hyperspectral sensors, such as the Hyperion imagery, have also been used in examining the surface soil salinity in more detail than multispectral data. Results showed that Hyperion could not map soil salinity at low soil salinity levels but could be used for areas with severe soil salinity levels (Dutkiewicz et al., 2009).

During the 2010s, multispectral optical sensors and hyperspectral data were successfully used to map soil salinity based on the correlation between several indices derived from spectral bands and soil reflectance data (Allbed & Kumar, 2013). The spectral reflectance of the salt present at the soil surface was initially used as a direct indicator of salinity. Several studies have used Landsat-8 OLI, which has been useful in detecting, mapping, and monitoring soil salinity (Allbed & Kumar, 2013). However, the salinity can be

detected directly if the area is dry (Mougenot & Pouget, 1993) and is not covered by vegetation for most of the year (Metternicht & Zinck, 2003).

Aside from surface soil salinity, land use or land cover has also been used as an indirect indicator to predict and map soil salinity. The soil salinity has been predicted and mapped by assessing the crop condition using the spectral vegetation indices. Vegetation indices derived from high temporal resolution satellites such as the Moderate Resolution Imaging Spectroradiometer (MODIS) were found to have the potential to detect soil salinity efficiently (Paliwal et al., 2019). Among the spectral vegetation indices used were the Normalized Difference Vegetation Index (NDVI), Soil-adjusted Vegetation Index (SAVI), Ratio Vegetation Index (RVI), Brightness Index (BI), and Green Vegetation Index (GVI) (Aldakheel, 2011; Eldeiry & Garcia, 2008; Scudiero et al., 2014; Zhang et al., 2011). Hick and Russell (1990) stated that soil salinity is better detected and identified through the combinations and band ratios among visible and near-infrared bands rather than by individual bands. In tropical areas where optical sensors are limited due to cloud coverage, the Synthetic Aperture Radar (SAR) data has been used to detect soil salinity. Hoa et al., 2019, attempted to detect soil salinity in tropical areas focusing on Vietnam using Sentinel-1 SAR data. The intensity and phase images were related to field-measured salinity.

### **1.5 Machine Learning Approaches**

Machine Learning (ML) algorithms have recently been used to model and predict soil salinity from remotely sensed data. The most commonly used MLs are Neural Networks (Hoa et al., 2019), Random Forests (Hoa et al., 2019; Ivushkin et al., 2019), Support Vector Regression (Taghadosi et al., 2019b), and Gaussian Process (Hoa et al., 2019). These ML algorithms are defined as data-driven, technique-based, and dependent on methods due to inputs retrieved from the remote sensing data and field data (Hoa et al., 2019). Table 2 on the next page summarizes some fundamental studies that assess soil salinity using satellite data, remote sensing, and machine learning methods.

Table 2. Studies related to soil salinity mapping

Study No	Title	Remote sensing data	Method	Results	Reference
1	Salinity stress detection in rice crops using time-series MODIS VI data	MODIS VI	Changes in crop phenology from MODIS time-series data were correlated with measured soil salinity	Seasonal integral (SI) as the best indicator of soil salinity; negative correlation with $r$ coefficient values of -0.76 and -0.84 for SI and mean Electric conductivity (EC); both are found to be statistically significant ( $P > 0.001$ ); SI and EC have a strong negative correlation ( $r = -0.92$ ) in homogenous rice pixels and $r = -0.63$ for mixed pixels	(Paliwal et al., 2019)
2	Detecting soil salinity with MODIS time series VI data	MODIS VI	Vegetation indices and crop phenology were correlated with measured soil salinity in 3 different sites	The strong negative correlation of NDVI and EC ( $r = -0.57$ ) and EVI and EC ( $r = -0.59$ ); biomass decreased linearly with the increased soil salinity in cropland ( $R^2 = 0.85$ ). SI as the best indicator of soil salinity	(Zhang et al., 2015)
3	Soil Salinity Mapping Using SAR Sentinel-1 Data and Advanced Machine Learning Algorithms: A Case Study at Ben Tre Province of the Mekong River Delta (Vietnam)	Sentinel 1 SAR Data	Backscatter values and GLCM from Sentinel-1 were correlated with measured salinity in 5 different machine learning algorithms	Performance of the five models was assessed through RMSE = 2.885, MAE = 1.897, $r = 0.808$ for GP and outperformed the other models; advanced machine learning models can be used for mapping soil salinity in Delta areas; a useful tool for assisting farmers and the policymakers	(Hoa et al., 2019)
4	Quantitative Estimation of Soil Salinity Using UAV-Borne Hyperspectral and Satellite Multispectral Images	A hyperspectral camera installed in an Unmanned aerial vehicle (UAV)	Reflectance factors from UAV and satellite data were correlated with field measured data to predict soil salinity in the Random Forests Regression method	Bare land exhibited the most severe salinity, followed by vegetation area and then sparse vegetation. RMSE = 2.98, CC=0.94 and RPD values=1.40 dS m <sup>-1</sup> . Concluded that a UAV-borne hyperspectral imager is a useful tool for field-based soil salinity mapping and monitoring	(Hu et al., 2019)
5	Retrieval of soil salinity from Sentinel-2	Sentinel-2 in combination	Satellite-derived soil features from Sentinel-2 and Landsat-8 were correlated	Kernel-based regression showed the most accuracy for modeling soil salinity $R^2 = 87.42\%$ and RMSE=5.1962	(Taghadosi et al., 2019a)

	multispectral imagery	with Landsat 8 (thermal band)	with measured salinity using MLR and SVR regression methods		
6	Assessing soil salinity using soil salinity and vegetation indices derived from IKONOS high-spatial-resolution imageries: Applications in a date palm dominated region	IKONOS satellite image	Vegetation and soil salinity indices derived from the IKONOS satellite were correlated with measured salinity	NDSI and SI-T indices in arid areas with low vegetation cover assess soil salinity; the SAVI index would yield better results for evaluating soil salinity in densely vegetated areas.	(Allbed et al., 2014)
7	Studying Vegetation Salinity: From the Field View to a Satellite-Based Perspective	Sentinel-2	Different spectral slopes in the VIR, NIR, and SWIR of Sentinel-2 were correlated with laboratory-measured halites	Generated Sentinel-2 based vegetation salinity index (SVSI) using $(\text{band } 4 - \text{band } 2) / (\text{band } 5 + \text{band } 11)$	(Lugassi et al., 2017)
8	Soil salinity prediction and mapping by machine learning regression in Central Mesopotamia, Iraq	Landsat 5 TM and ALOS L-band radar data	Biophysical indicators from TM and soil component from ALOS were correlated with measured salinity in SVR, RFR, and MLR algorithms	Random Forests performed better than SVR with higher accuracy (93.4–94.2% vs. 85.2–89.4%), and less normalized root mean square error (NRMSE; 6.10–7.69% vs. 10.29–10.52%)	(Wu et al., 2018)
9	Soil salinity mapping using dual-polarized SAR Sentinel-1 imagery	Sentinel 1 SAR Data	Radar intensities and GLCM from Sentinel-1 were correlated with measured salinity in the SVR method	Radial Basis Function had the most accuracy of the coefficient of determination $R^2=0.97$ , RMSE=0.3561, $G_{FO, VV}$ , and $R_{VH}$ had the best performance in salinity detection	(Taghadosi et al., 2019b)
10	Quantitative assessment of soil salinity using remote sensing data based on the artificial neural network, case study: Sharif Abad Plain, Central Iran	Landsat-8	Spectral parameters from Landsat-8 were correlated with measured salinity in the Neural Network model	GFF algorithm is the best method of preparing a soil salinity map; values from the ANN model are lesser than actual values	(Habibi et al., 2021)

## 1.6 Random Forests Regression

As mentioned in section 1.5, Random Forests is one of the machine learning algorithms that has been used in predicting soil salinity. Random Forests is a non-parametric, supervised learning algorithm proposed by Breiman, 2001 which uses ensemble learning methods for classification and regression. Random Forests does not make strong assumptions about the form of the mapping function. The model is free to learn any functional form from the training data by not making any assumptions. It operates by constructing many decision trees at training time and outputting the class that is the mean prediction (for regressor) of the individual trees. In addition, it is a meta-estimator because it combines the result of multiple predictions, which aggregates the decision trees (Chakure, 2019).

With the use of a training dataset, for example, dataset X, each tree is created from different dataset rows. At each node, a different set of sample features are selected for splitting. Then, each tree will make its prediction. Furthermore, the predictions are averaged to produce a single result. This averaging improves the accuracy of the model and reduces overfitting (Mwiti, 2021). An illustration of how Random Forests regression (RFR) works is shown in Figure 5.

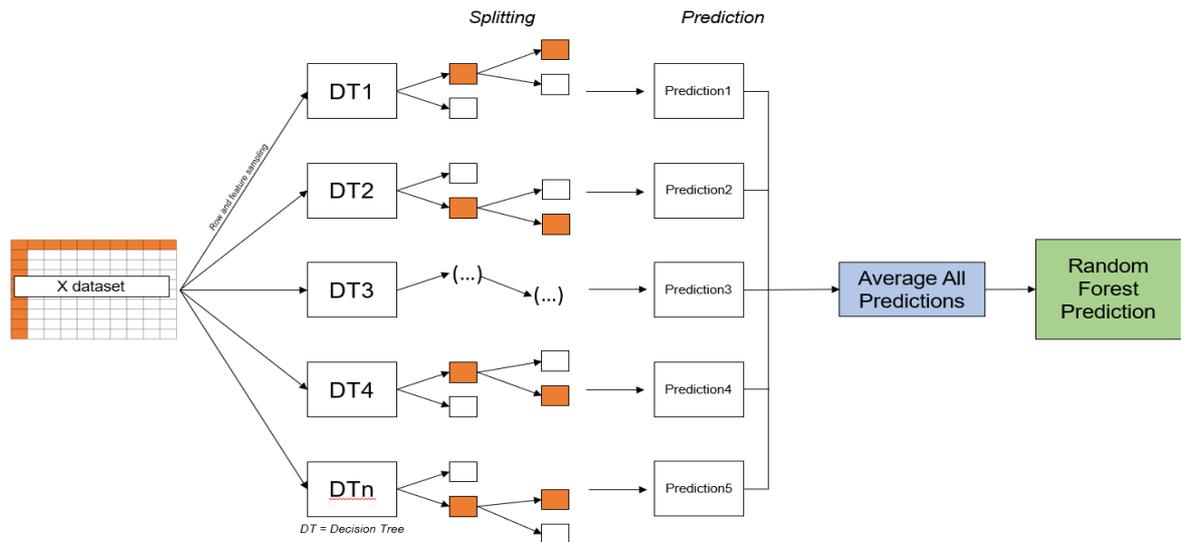


Figure 5. Random Forests Regression

## 1.7 Problem Formulation

### 1.7.1 Wicked problem

Soil salinization in agricultural areas affects crop growth and production. In the Philippines, the regions affected by salinization and severity levels are still unknown due to a lack of updated information and database. Also, communicating the negative impacts of soil salinization in agricultural areas to the stakeholders such as the policymakers, local farmers, and other interested parties remains a challenge, thus creates a wicked problem.

This study proposed a new method for retrieving and detecting soil salinization using a multi-sensor approach with field-measured data with the use of free high-resolution satellite data, remote sensing techniques, and machine learning methods.

### 1.7.2 Research Objectives

#### *General Objective*

The general objective of this study is to develop a method to effectively detect the spatial pattern of soil salinity using a multi-sensor approach. Furthermore, this study is geared up to recommend strategies for the stakeholder to combat and alleviate the adverse effects of soil salinization in the coastal rice areas of the Province of Ilocos Sur, Philippines.

#### *Specific Objectives*

Specifically, this research aims to:

- determine the spatial extent of surface soil salinity from the predictor variables generated from Sentinel-1, Sentinel-2, and Landsat 8;
- evaluate the performance of detecting soil salinity from a multi-sensor approach using the Random Forests Regression algorithm;
- map and analyze the multi-year spatial pattern of soil salinity

### 1.7.3 Research Questions

- What is the spatial distribution of soil salinization predicted from the variables generated from Sentinel-1, Sentinel-2, and Landsat 8?
- What is the accuracy of using the multi-sensor approach in retrieving soil salinity?
- To what extent does soil salinity change over multiple years based on the multi-sensor approach?

### 1.7.4 Research Hypothesis

1. Combining data from multiple sensors will result in a more accurate soil salinity map than single sensor-based approaches.

## 2. DATA AND RESEARCH METHODS

### 2.1 The Study Area

The Province of Ilocos Sur is located in the central part of the Ilocos Region in the Northwest of Luzon Island, Philippines. The province lies within 16°40' to 17°54'N and 120°20' to 120°48'E, and the location is shown in Figure 6. It is bounded by the provinces of Ilocos Norte in the north, Abra and Mountain Province in the east, and La Union in the south. On the west is the West Philippine Sea. Vigan is the capital city of the province that is 276 miles or 443 kilometers from Manila. The province's total land area is approximately 2,596 square kilometers (National Statistics Office, 2012).

According to the 2019 soil mapping of the DA-BSWM, the total surveyed area for rice production is 13,872 hectares (DA-BSWM, personal communication, 23 July 2020). The soil sampling and testing results show that the rice areas have low soil fertility due to low macronutrient content and high soil pH level. A map of the rice areas is shown in Appendix A and shows that most rice areas are located primarily in coastal areas.

The selection of this study area was due to various reasons: first, there is no updated map on soil salinity countrywide because the only information about soil salinity was dated 1986; second, the results of the 2019 soil survey by the DA-BSWM shows that soil salinity is an issue in some rice-growing areas; and lastly, the province is one of the target areas that the DA-BSWM is planning to have information about soil salinity.

The province falls into Type I of Modified Coronas' Climate Classification, generally. According to Philippine Atmospheric, Geophysical, and Astronomical Services Administration (PAGASA), this is defined as having two pronounced seasons, dry from November to April and wet during the year. However, the Hernandez type of classification, which has higher spatial resolution than Coronas, under PAGASA, defines that the province's climate is generally arid (Type E). This type has more dry months than wet months. At most, there are only 4 ½ wet months in a year. However, the southernmost portion of the province (near the municipality of Cervantes) was observed to be humid (Type B), where rain is evenly distributed throughout the year with at most three dry months and the eastern part of Sugpon is dry (Type D) with rain not sufficiently distributed with at most six dry months (Ilocosur.gov.ph, n.d.).

The DA-BSWM had conducted a reconnaissance soil survey of the province of Ilocos Sur in 1954. The soil survey report showed that the province comprises the middle portion of a distinct physiographic unit known as the Ilocos Coast strip, which extends along the western side of Luzon. It is very irregular, with the broadest portion hardly exceeding 20 kilometers. The province is relatively uneven in physical features. The relief ranges from level to hilly and mountainous, as shown in Figure 7. The narrow coastal plain is level to undulating, with several low hills scattered at random throughout the entire length of the province. The eastern portion, which borders Abra and the Mountain Province, is hilly to mountainous. The highest mountains in the province are the peaks of the Malaya Range southwest of the town of Cervantes, the two highest peaks of which are the Malaya mountains and Mount Libo (Mariano et al., 1954). The information on the soils of the province is shown in Appendix B.



Figure 6. Study Area

In general, the province drains to the west. Due to the narrowness of the plain and the nearness of the mountains and hills to the sea, the rivers and streams are generally short and rapid with few or no meanders. Most of the low-lying lands are subject to annual floods, and due to the high velocity of the streams, they are usually destructive. The Abra River, which rises from the slopes of Mount Dana in the Mountain Province, debouches onto the plain near the town of Santa, south of Vigan. Some of the larger rivers in the province are the Chico, a branch of the Amburayan River, Buaya river, near the towns of Sta. Cruz, Candon, Sta. Maria, Narvacan, Parsua, Cabugao, and Sinait river. None of these are navigable except the Abra River, where raft and dugouts can be used as far inland as Bangued, the capital of Abra Province (Mariano et al., 1954).

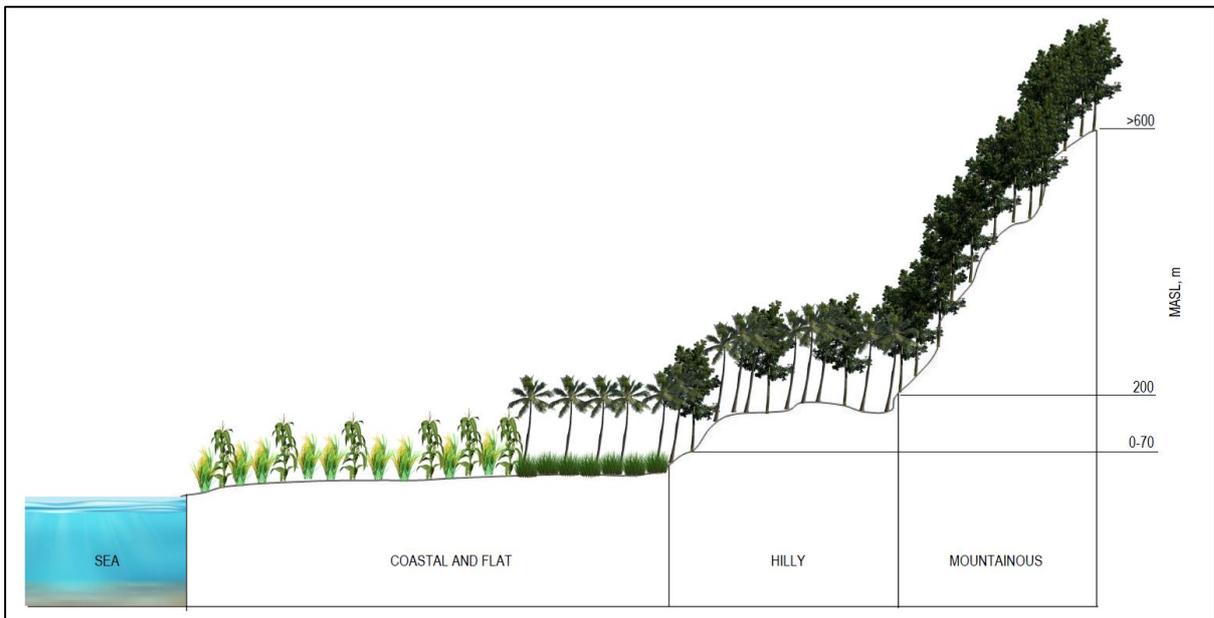


Figure 7. General landscape cross-section of the study area

The sources of soil salinity in coastal areas of the province are increasing sea level rise, over-pumping of the aquifers, outflow along the river that occurs when seawater moves upstream into the river during high tide but low river flow (DA-BSWM, 2021). Climate is a factor for the increasing sea-level rise. The tropical monsoon climate causes freshwater to accumulate on the soil during the wet season and wash away the saline water. However, there's not much fresh water supply during the dry season, and outflow along the river carries salt water. This salt will be washed out again in the next wet season. However, human activities accelerated soil salinization. One is the over-pumping of aquifers in agricultural areas, which is then associated with poor farming practices. In other areas, local people use shallow-tube well or pumping irrigation water from rivers or creeks. In the database of DA-BSWM, there are currently 59 shallow-tube well facilities throughout the province servicing 1,700 hectares (DA-BSWM, personal communication, 23 July 2020). The province is also well-known for salt-making. Salt beds were established in the coastal areas in the different municipalities of the province (Fenix, 2020).

Figure 8 below shows the general overview of soil salinity using sample points collected and tested by the DA-BSWM for 2019 ranging from 0.02 to 4.26 dS/m with an average of 0.25 dS/m. The soil samples were collected from 0 to 30 centimeters deep (DA-BSWM, personal communication, 23 July 2020). From the figure, higher salt concentrations are found in the province's coastal areas with the highest values in Santa Catalina, the City of Vigan, and Santa. However, as the sample points went further from the coastal areas, the salt concentrations are decreasing.

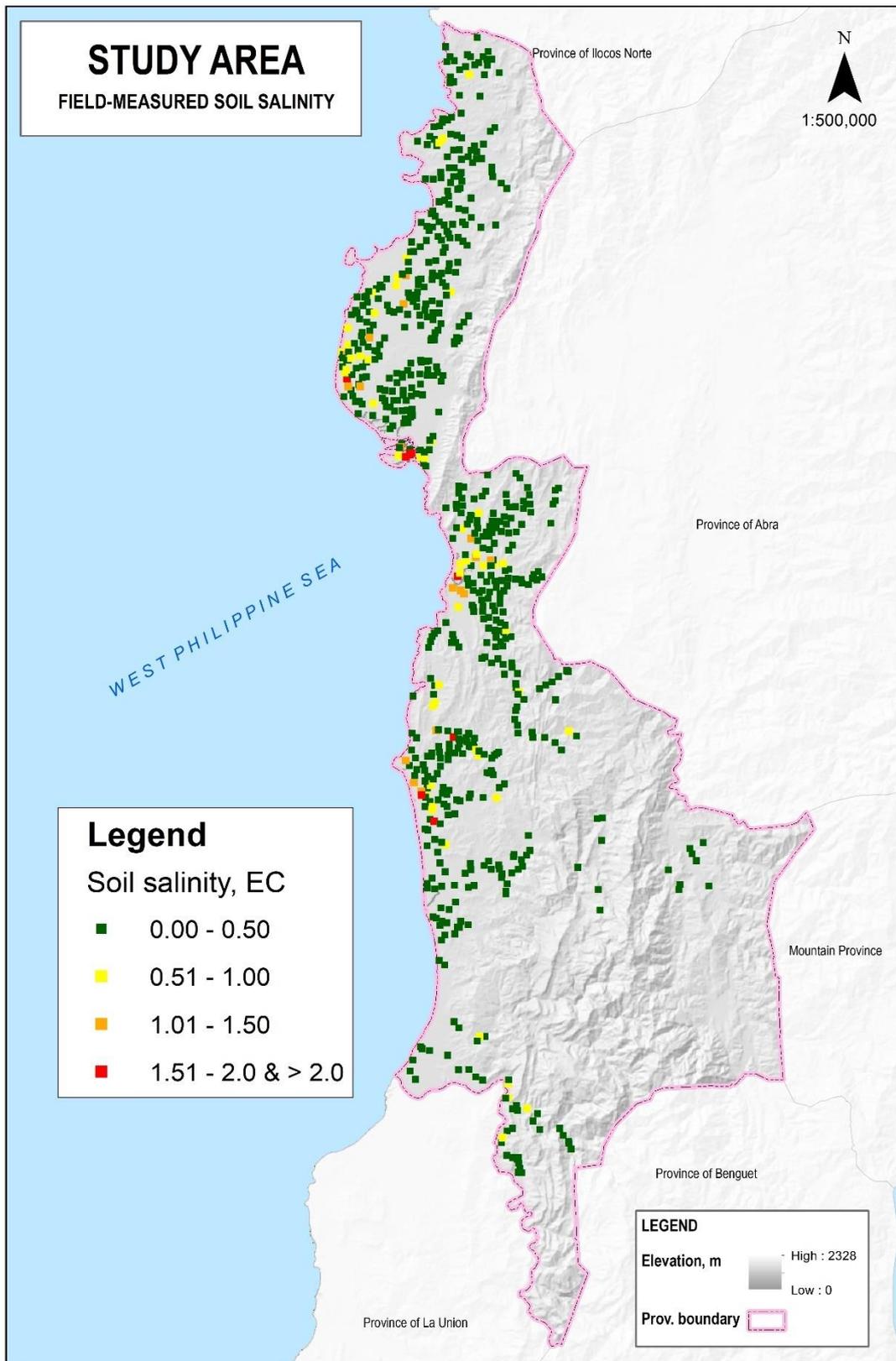


Figure 8. Field-measured soil salinity

## 2.2 Materials, Data, and Software Used

The list of materials and data used in the study included field data, satellite data, and ancillary data comprised of soil properties, geographic location, and climate data. Table 3 summarizes data used, their types, format, sources, and collection method.

Table 3. Summary of data, types, and their sources

No.	Data/Information	Type/Format	Source and Collection Method
1	Ground-truth data	Excel	Provided by DA-BSWM
2	Rice area	Shapefile	Provided by DA-BSWM
3	Satellite data (Sentinel1)	Compressed zip file, SAFE file	Retrieved from Sentinel sci-hub website
4	Satellite data (Sentinel2)	Compressed zip file, SAFE file	Retrieved from USGS website
5	Satellite data (Landsat8)	Compressed zip file, TIFF file	Retrieved from USGS website
6	Clay, silt, loam, soil pH, bulk density	TIFF file	Retrieved from SoilGrids website
7	Digital Elevation Model	TIFF file	Provided by DA-BSWM
8	Soil moisture and climate data (precipitation and temperature)	.nc file	Retrieved from the TerraClimate website
9	Geographical coordinates	TIFF file	User-generated

The software programs used were Sentinel Application Platform (SNAP) version 8 for processing images of Sentinel-1 and Sentinel-2; ArcGIS 10.8.1 was used to process images from Landsat-8, process images generated from SNAP, and creating visualizations. Spyder 4, a part of Anaconda 3, a python software used to develop and run the Random Forests Regression model.

## 2.3 Data Collection

### 2.3.1 Existing Data

Existing data was composed of field or ground truth data from DA-BSWM last March to April 2019 (DA-BSWM, personal communication, 23 July 2020). There were 706 soil samples collected from rice areas, as shown in Figure 8. The ground truth data is an excel file that contains the locations where the soil samples were taken and the results of soil analysis in rice areas of Ilocos Sur. The file contained the ID or the identification number of each soil sample, date when the soil samples were collected, pre-defined sampling point, sample number, barangay, latitude, longitude, municipality, farmer's name, laboratory code, organic matter content, phosphorous, potassium, zinc, manganese, copper, iron, soil moisture, soil pH, electric conductivity (EC) and soil texture. For this research, only the ID, municipality, geographical coordinates, and EC were used.

DA-BSWM also provided a delineation of rice areas (Appendix A). According to them, this delineation was done manually in Google Earth, supported by the sampling points taken in the ground (DA-BSWM, personal communication, 23 July 2020). This delineation also includes other vegetation such as corn and vegetables present in the area during soil sampling. The primary purpose of the land is for rice production. This delineation file was available as a shapefile in Universal Transverse Mercator (UTM) Zone 51 North (EPSG: 32651).

### 2.3.2 Satellite Data

Three different satellite data were used in this study: Sentinel-1 Synthetic Aperture Radar (SAR) data, Sentinel-2 Multispectral Instrument (MSI), and Landsat-8 Operational Land Imager (OLI). These were collected through Sentinel SciHub for Sentinel-1 and the NASA USGS Earth Explorer for Sentinel-2 and Landsat-8.

For Sentinel-1, the images were retrieved based on sensing date, in descending order, the polarization of VV+VH, the product type is Single Look Complex (SLC), and the sensor mode was interferometric swath (IW) with a 250-kilometer swath width. The time window was March 15 to April 10, 2019. This time window was selected to coincide with the dates of soil sampling.

For Sentinel-2 and Landsat-8, since both are optical satellite data, the cloud cover was limited to 10%, and the date range of March 1, 2019, to April 30, 2019. The extent of the images was based on the coordinates in Table 4 below.

Table 4. Footprint of the study area to download the images

Point	Coordinates
1	17°57'09" N, 121°00'51" E
2	17°57'28" N, 121°12'24" E
3	16°38'00" N, 120°08'57" E
4	16°37'13" N, 120°57'53" E

### 2.3.3 Ancillary Data

The ancillary data used in the study are soil properties like bulk density, soil pH level, soil texture (sand, silt, and clay), digital elevation model (DEM), soil moisture, climate data (average temperature and annual precipitation), and geographical coordinates. Soil properties such as bulk density, soil pH level, and soil texture were retrieved from SoilGrids (Poggio et al., 2021) through ISRIC – WDC Soils in WGS 1984 (EPSG: 4326). The DA-BSWM provided the DEM projected in UTM Zone 51 N (EPSG: 32651), containing information on the elevation from mean sea level.

Soil moisture and climate data are downloaded from TerraClimate (Abatzoglou et al., 2018) in WGS 84 (EPSG: 4326). TerraClimate is a monthly climate and climatic water balance dataset for global terrestrial surfaces from 1958-2020. These data provide essential inputs for ecological and hydrological studies at global scales that require high spatial resolution and time-varying data. All data have monthly temporal resolution and a ~4-km (1/24th degree) spatial resolution (Abatzoglou et al., 2018). Table 5 below shows the description of the ancillary data used in the study.

Table 5. General description of the ancillary data

No.	Soil Property	Depth	Unit	Spatial Resolution
1	Bulk density	0-5 cm	cg/cm <sup>3</sup>	250 m
2	Clay content	0-5 cm	g/kg	250 m
3	Sand content	0-5 cm	g/kg	250 m
4	Silt content	0-5 cm	g/kg	250 m
5	Soil pH level	0-5 cm	pH*10	250 m
6	Digital Elevation	Not applicable	m	100 m

	Model			
7	Soil moisture	Undefined	m <sup>3</sup> water/m <sup>3</sup> soil	~4km
8	Average Temperature	Not applicable	°C	~4km
9	Precipitation	Not applicable	cm	~4km

## 2.4 Processing of Satellite Data

### 2.4.1 Sentinel-1 SAR Data

The Sentinel-1 data were downloaded on The Copernicus Open Access Hub website (<https://scihub.copernicus.eu/dhus/#/home>). Level-1 Single Look Complex (SLC) was the product type used in dual (VV and VH) polarization, and Interferometric Swath (IW) was selected for Sensor Mode. The image dates were retrieved based on their sensing period in descending order. The images retrieved are listed in Table 6:

Table 6. List of Sentinel-1 images

No.	Date	Identifier
1	March 17, 2019	S1B_IW_SLC__1SDV_20190317T100658_20190317T100726_015393_01CD2B_CECC
2	March 17, 2019	S1B_IW_SLC__1SDV_20190317T100632_20190317T100700_015393_01CD2B_E2ED
3	March 27, 2019	S1A_IW_SLC__1SDV_20190327T214555_20190327T214623_026529_02F906_7027
4	April 10, 2019	S1B_IW_SLC__1SDV_20190410T100658_20190410T100726_015743_01D8B3_758E
5	April 10, 2019	S1B_IW_SLC__1SDV_20190410T100633_20190410T100701_015743_01D8B3_7D75

The images with the same date (1&2, and 4&5) are image pairs that cover the study area. However, the image had not fully covered the study area. Thus, image number 3 was also used. An overview of the step-by-step process of pre-processing Sentinel-1 images is shown in Figure 9, and the detailed explanation of each step can be found in Appendix C. There were 20 variables, of which two were from the Gamma-nought bands, and 18 were from the Gray-Level Co-occurrence Matrix (GLCM) generated from Sentinel-1 as listed in Table 7.

Table 7. List of predictors generated for Sentinel-1

B1	VH Gamma0	B11	VV Gamma0
B2	VH contrast	B12	VV contrast
B3	VH dissimilarity	B13	VV dissimilarity
B4	VH homogeneity	B14	VV homogeneity
B5	VH energy	B15	VV energy
B6	VH max	B16	VV max
B7	VH entropy	B17	VV entropy
B8	VH GLCM mean	B18	VV GLCM mean
B9	VH GLCM variance	B19	VV GLCM variance
B10	VH GLCM correlation	B20	VV GLCM correlation

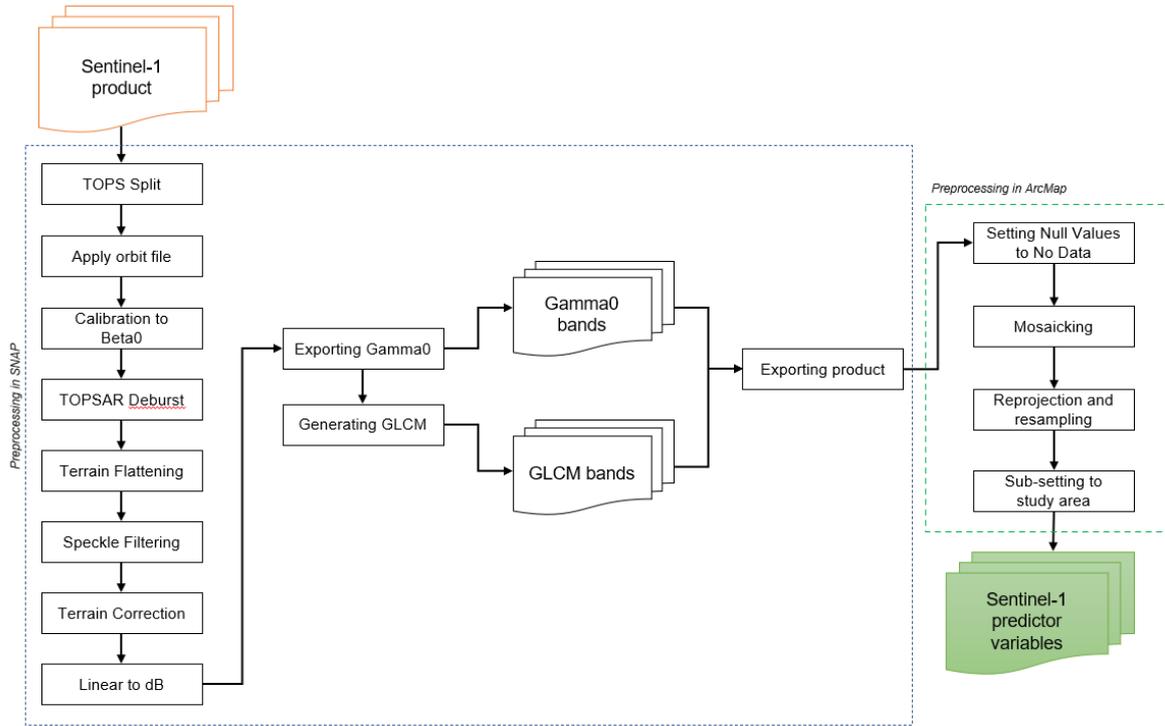


Figure 9. Workflow for processing Sentinel-1 Product

**2.4.2 Sentinel-2 MSI Data**

The Sentinel-2 images were retrieved from USGS Earth Explorer (<https://earthexplorer.usgs.gov/>). The data retrieved are in Level-1C Top-of-Atmosphere. The cloud cover in the search criteria was limited to 10%, with the following spatial extent stated in Table 4. The list of Sentinel-2 images retrieved is listed in Table 8.

Table 8. List of Sentinel-2 images

No.	Date	Identifier
1	March 7, 2019	L1C_T50QRE_A019348_20190307T023810
2	March 7, 2019	L1C_T50QRD_A019348_20190307T023810
3	March 27, 2019	L1C_T51QIV_A019634_20190327T023815
4	March 27, 2019	L1C_T51QTU_A019634_20190327T023815
5	April 11, 2019	L1C_T51QIV_A010940_20190411T023827
6	April 11, 2019	L1C_T50QRD_A010940_20190411T023827

Sentinel-2 has a total of 13 bands in different wavelengths of the electromagnetic spectrum. This satellite data is also available in three different spatial resolutions: 10, 20, and 60 meters. Table 9 shows the bands of Sentinel-2 as described by European Space Agency (ESA, n.d.).

Table 9. Band names of Sentinel-2

Band No.	Band name	Spatial Resolution	Wavelength, $\mu\text{m}$
Band 1	Coastal aerosol	60	0.443
Band 2	Blue	10	0.490
Band 3	Green	10	0.560
Band 4	Red	10	0.665

Band 5	Vegetation Red Edge	20	0.705
Band 6	Vegetation Red Edge	20	0.740
Band 7	Vegetation Red Edge	20	0.783
Band 8	NIR	10	0.842
Band 8A	Vegetation Red Edge	20	0.865
Band 9	Water Vapor	60	0.945
Band 10	SWIR – Cirrus	60	1.375
Band 11	SWIR	20	1.610
Band 12	SWIR	20	2.190

Figure 10 shows the overview of the pre-processing steps conducted for Sentinel-2 images. The detailed information can be found in Appendix D. A total of 17 vegetation and soil salinity indices have been generated from Sentinel-2 data. The list is shown in Table 10.

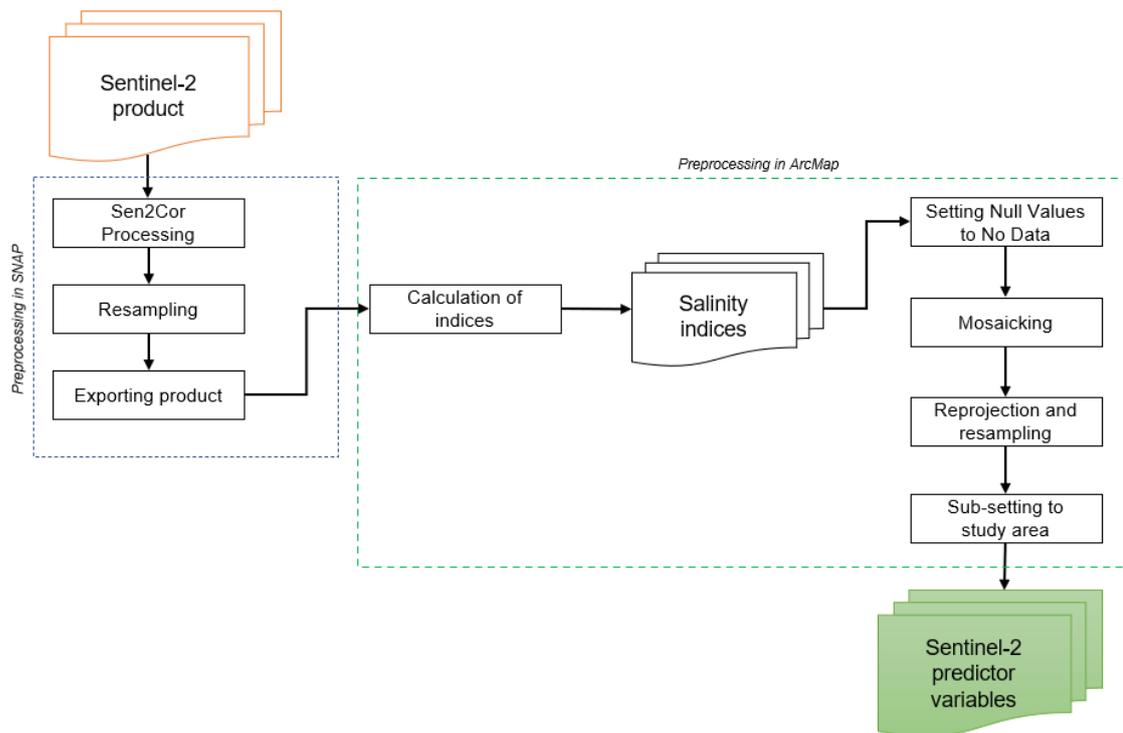


Figure 10. Workflow for processing Sentinel-2 products

Table 10. List of salinity indices generated from Sentinel-2 Data

No.	Index	Formula	Reference
1	Normalized Difference Vegetation Index	$\frac{(NIR - R)}{(NIR + R)}$	(Rouse et al., 1973)
2	Normalized Difference Salinity Index	$\frac{(R - NIR)}{(NIR + R)}$	(M. Khan & Sato, 2001)
3	Normalized Difference Water Index	$\frac{(G - NIR)}{(G + NIR)}$	(Gao, 1996)
4	Canopy Response Salinity Index	$\frac{(R * NIR) - (B * G)}{(R * NIR) + (B * G)}$	(Valley et al., 2014)

5	Combined Spectral Response index	$\frac{(B + G)}{(R + NIR)} * NDVI$	(Fernández-Buces et al., 2006)
6	Enhanced Vegetation Index	$\frac{2.5(NIR - R)}{(NIR + 6 * G - 7.5 * B + 1)}$	(Liu & Huete, 1995)
7	Brightness Index	$\sqrt{R^2 + NIR^2}$	(N. Khan et al., 2005)
8	Generalized Difference Vegetation Index	$\frac{(NIR^2 - R^2)}{(NIR^2 + R^2)}$	(N. Khan et al., 2005)
9	Simple Ratio Vegetation Index	$\frac{NIR}{R}$	(Pearson et al., 1972)
10	Salinity Index	$\sqrt{B * R}$	(M. Khan et al., 2001)Khan 2001
11	Salinity Index - 1	$\frac{(R * NIR)}{G}$	(Abbas et al., 2013; M. Khan et al., 2001)
12	Salinity Index - 2	$\frac{(B * R)}{R}$	(Abbas et al., 2013; M. Khan et al., 2001)
13	Salinity Index - 3	$\frac{(B - R)}{(B + R)}$	(M. Khan & Sato, 2001)
14	Salinity Index - 4	$\sqrt{R^2 + G^2}$	(Douaoui et al., 2006)
15	Salinity Index - 5	$\sqrt{B^2 + R^2 + NIR^2}$	(Douaoui et al., 2006)
16	Salinity Index - 7	$\frac{B}{R}$	(Abbas et al., 2013; M. Khan et al., 2001)
17	Soil Adjusted Vegetation Index	$1 + L \frac{(NIR - R)}{(NIR + R + L)}$	(Huete, 1988)

### 2.4.3 Landsat-8 OLI Data

The Landsat-8 images were downloaded from USGS Earth Explorer (<https://earthexplorer.usgs.gov/>). The cloud cover was limited to 10%, and the extent of the polygon to retrieve the images was the same as for Sentinel-2. The list of Landsat-8 images is listed in Table 11.

Table 11. List of Landsat-8 images

No.	Date	Identifier
1	March 16, 2019	LC08_L1TP_117048_20190316_20190325_01_T1
2	April 1, 2019	LC08_L1TP_117048_20190401_20190421_01_T1

The metadata details of Thermal Infrared Sensor (TIRS) for Landsat-8 are listed in table 12. The metadata was used as input values for deriving land surface temperature (LST).

Table 12. Metadata of thermal bands from Landsat-8 OLI Data

Information	Image 1	Image 2
Sun Elevation	57.44314117	61.88528253
Radiance-Mult-Band 10		3.3420E-04
Radiance-Mult-Band 11		3.3420E-04
Radiance-Add-Band 10		0.10000
Radiance-Add-Band 11		0.10000

$K_1$ for Band 10	774.8853
$K_1$ for Band 11	480.8883
$K_2$ for Band 10	1321.0789
$K_2$ for Band 11	1201.1442

Landsat-8 OLI Data was used for deriving LST based on the following formula:

$$LST = \frac{BT}{1 + W * \left(\frac{BT}{\rho}\right) \ln \epsilon \lambda}$$

Where:

- LST is the Land Surface Temperature in Celsius
- BT is the brightness temperature at satellite
- W is the wavelength of emitted radiance
- $\rho$  is a constant equal to  $1.438 \times 10^{-2}$  m K derived from

$$\rho = h \frac{c}{\sigma}$$

Where  $h$  is Planck's constant ( $6.626 \times 10^{-34}$  JS)

$c$  is the velocity of light ( $2.998 \times 10^8$  m/s)

$\sigma$  is the Boltzmann constant ( $1.38 \times 10^{-23}$  J/K)

$\epsilon \lambda$  is the emissivity calculated

Figure 11 shows the overview of the processes conducted in deriving land surface temperature. The detailed information can be found in Appendix E.

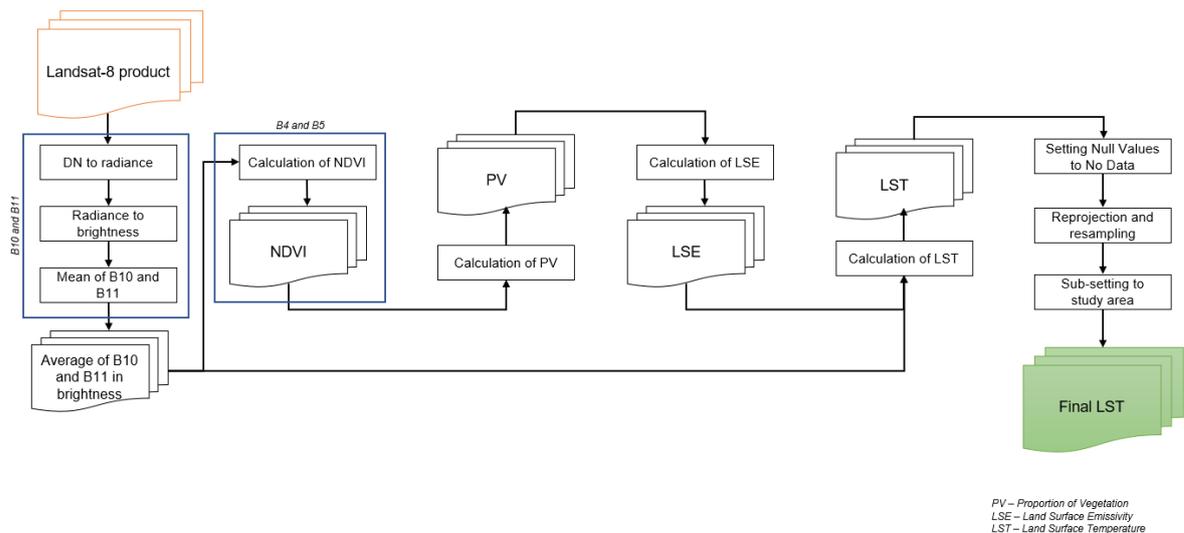


Figure 11. Workflow for deriving LST from Landsat-8

## **2.5 Processing of Ancillary Data**

### **2.5.1 Soil properties, Climate and DEM**

Ancillary data were processed using ArcMap 10.8. Processing includes setting no data values, reprojection, resampling, and clipping into the study area. Further processing of the images included averaging maximum and minimum temperature using the raster calculator tool and recalculating pixel values for soil pH. SoilGrids website reports that the soil pH level is multiplied by 10. Thus, it was recalculated to get the original range of soil pH (0-14) using a raster calculator tool.

### **2.5.2 Geographical Location**

Geographic location (x and y coordinates) was also considered as an auxiliary (explanatory variables). A raster file was created containing the information on latitude and longitude in ArcMap. First, the boundary of the study area was converted into a raster file having  $10 \times 10$  meter spatial resolution. Then, the rasterized study area was converted into a point in a vector format. Two fields were then added into the point file, containing the latitude and longitude information in meters. Lastly, the point file was converted into raster again for both latitude and longitude fields.

### **2.5.3 Bare soil areas mask**

According to the soil sampling protocol of the DA-BSWM, the soil samples must be taken in areas that have no presence of vegetation, not fertilized, and not been disturbed, or have undergone land preparation. However, upon making the delineation of rice areas, DA-BSWM confirmed that they also include the rice areas with standing crop on the date of soil sampling (DA-BSWM, personal communication, 12 December 2020). These had to be removed with a masking process. The protocol of soil sampling used by the DA-BSWM can be found in Appendix F.

The generated NDVI image from Sentinel-2 was used to extract the bare soil areas where NDVI was less than 0.35 values. This image was clipped to the DA-BSWM's delineation of rice area to retrieve the bare soil in rice areas. The areas that were less than 0.2 hectares were removed because they did not represent the original rice area delineation. There were 1,614 rice areas generated and used for creating random points.

### **2.5.4 Random sample points**

The geographical coordinates of the sampling points from DA-BSWM referred to the location where the "fieldman" was standing in the field when soil samples were taken in the rice areas. It does not represent the exact location of each of the ten soil samples taken for that rice area, nor does it necessarily represent the centroid of those ten sample locations. Therefore, it is unclear where the sub-samples were collected inside the rice area. Random points were created to compensate for this issue.

Using the create random points tool in ArcMap, ten random sample points were created within the newly created rice areas having a maximum distance of 50 meters. This process resulted in 6,587 new random points. Note that this number was expected to be 16,140 points (ten times of the newly created rice areas); however, since not all the polygons have the same area and extent, not all polygons had 10 points in them.

### 2.5.5 Extracting pixel features

Although this is not considered ancillary data, this method is essential to extract the pixel values of generated predictor variables. The pixel values were extracted using the Extract Multi Values to Points tool in ArcMap. This table was exported into an excel file and used as an input for machine learning.

## 2.6 Modelling Soil Salinity

The code for Random Forests (RF) regression was run in the Python Anaconda environment. The code contains the following ten parts that determine the flow of predicting soil salinity:

1. General description of the code
2. Importing plugins
3. Specifying work path
4. Reading the excel data and transforming it into a pandas data frame
5. Training and testing of excel data
6. Saving the trained dataset and drawing the figure of results
7. Reading and loading the satellite data
8. Reformatting the satellite data into a pandas data frame
9. Predicting soil salinity
10. Creating soil salinity map

The number of estimators or number of trees was set to 300, and the random state was set to 42 to maintain reproducibility. Two multi-sensor approaches were used in training and testing the data. These are Optimized Multi-Sensor Predictors and Multi-Sensor Predictors.

The Optimized Multi-Sensor Predictor (OMSP) is the process of combining the predictors that are derived from different sensors that were within the set threshold. Mainly, this method is merging the three different sensors into a union. The training and testing of data from Random Forests regression were run separately for predictors generated from Sentinel-1, Sentinel-2, Landsat-8, and all ancillary data. Every predictor from the three sensors within the significance threshold was used as input to establish the model.

On the other hand, the Multi-Sensor Predictors (MSP) is a process where all predictor variables of all sensors were used to train the data and then select the predictors within the significance threshold to establish the model. This approach also means that a group of sensor data is associated with a common purpose. All of the 20 features generated from Sentinel-1, 17 salinity indices generated from Sentinel-2, one from Landsat-8, and 11 ancillary data were used to run in one training to see which remote sensing data is associated with the other.

The set threshold for this study used the 0.05 feature importance value. This value is the typical threshold in selecting predictors that could reduce the noise, data redundancy, and uncertainty in establishing the RF model (Paul et al., 2013). Both of these approaches resulted in two different models were used in predicting soil salinity.

## 2.7 Accuracy Assessment

The two approaches were assessed using statistical metrics. The metrics used in measuring the accuracy of soil salinity modeling are root-mean-squared error (RMSE), r-squared ( $R^2$ ), and Pearson correlation

coefficient (PCC). These metrics were the basis of selecting which of the two approaches are better in predicting soil salinity. Furthermore, the predicted soil salinity maps in rice areas were prepared and layout in ArcMap.

## 2.8 Soil Salinity Variability Analysis

### 2.8.1 Spatial Variability

The variability in soil salinity was analyzed per municipality. This variability was then related to the input data used in the model. Soil salinity levels guided the computation of the spatial extent of soil salinity in hectares in Table 13 below. Although using the USDA's criteria for soil salinity levels, the predicted soil salinity range falls into the low to slight category (Table 1). The range is projected into low to high values for this study, as shown in Table 13.

Table 13. Soil salinity levels for the study

Predicted EC level, dS/m	Description
0.0-0.5	Low
0.5-1.0	Moderately Low
1.0-1.5	Moderately High
1.5-2.0	High
>2.0	

### 2.8.2 Multiyear Variability

The better-performing model from 2019 was used to predict the soil salinity for 2017, 2018, and 2020. The remote sensing data was retrieved on the same footprint used in establishing the model, as shown in Table 4 of Section 2.3.2. Table 14 below shows the list of images from the three different sensors. Furthermore, the climate data and soil moisture were also treated as varying explanatory variables.

Table 14. List of images used in multiyear analysis of soil salinity

Sensor	Image date	Identifier
Sentinel-1	30 March 2017	S1A_IW_SLC__1SDV_20170330T215340_20170330T215410_015927_01 A420_0C53
		S1A_IW_SLC__1SDV_20170330T215408_20170330T215435_015927_01 A420_146C
	6 April 2018	S1A_IW_SLC__1SDV_20180406T215347_20180406T215416_021352_02 4BFF_5204
		S1A_IW_SLC__1SDV_20180406T215414_20180406T215442_021352_02 4BFF_6A1D
	26 March 2020	S1A_IW_SLC__1SDV_20200326T215359_20200326T215429_031852_03 AD13_6E9D
		S1A_IW_SLC__1SDV_20200326T215427_20200326T215455_031852_03 AD13_BD08
Sentinel-2	7 March 2017	L1C_T50QRE_A008909_20170307T023516
		L1C_T51QTU_A008909_20170307T023516
	17 March 2018	L1C_T50QRD_A005363_20180317T023117
		L1C_T50QRE_A005363_20180317T023117

---

	21 March	L1C_T50QRD_A024782_20200321T023819
	2020	L1C_T51QTV_A024782_20200321T023819
Landsat-8	10 March	LC08_L2SP_117048_20170310_20200904_02_T1
	2017	
	29 March	LC08_L2SP_117048_20180329_20200901_02_T1
	2018	
	3 April	LC08_L2SP_117048_20180329_20200901_02_T1
	2020	
Precipitation		TerraClimate_ppt_2017.nc
		TerraClimate_ppt_2018.nc
		TerraClimate_ppt_2020.nc
Soil moisture		TerraClimate_soil_2017.nc
		TerraClimate_soil_2018.nc
		TerraClimate_soil_2020.nc
Temperature, max		TerraClimate_tmax_2017.nc
		TerraClimate_tmax_2018.nc
		TerraClimate_tmax_2020.nc
Temperature, min		TerraClimate_tmin_2017.nc
		TerraClimate_tmin_2018.nc
		TerraClimate_tmin_2020.nc

---

The multiyear variability of soil salinity from 2017 to 2020 was analyzed, and the hotspot areas were determined. The hotspot areas are the areas with increased soil salinity. The image processing was done in ArcMap using the raster calculator tool in ArcMap, calculating the difference of predicted soil salinity in 2017 and 2020. Areas with negative differences describe the areas with decrease soil salinity, and the positive difference shows areas with increasing soil salinity levels. Both spatial and multiyear variability were further assessed in this method. The general workflow of the methods used in this study is shown in Figure 12 below.

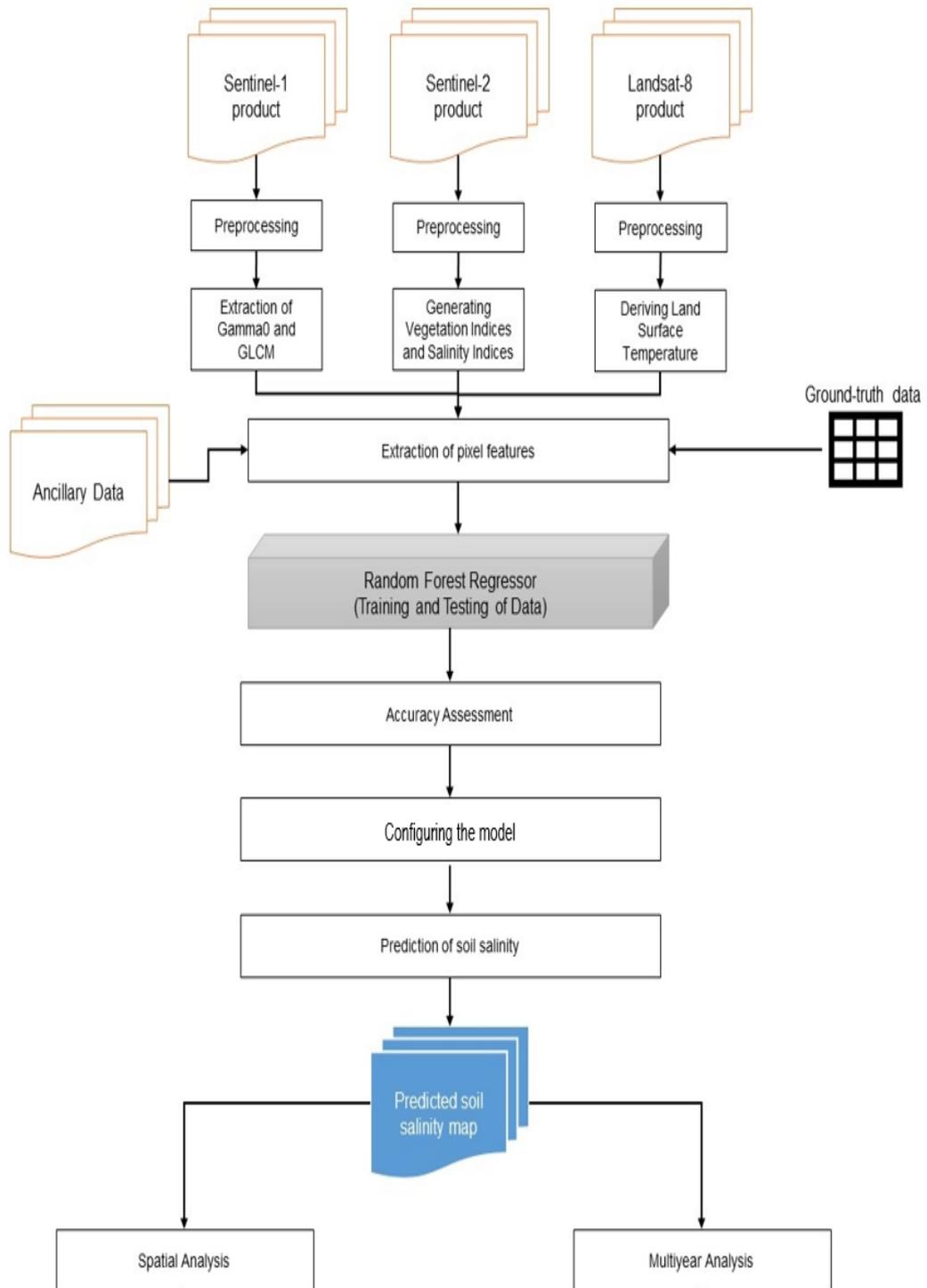


Figure 12. General workflow of the methods used in the study

## 3. RESULTS AND DISCUSSION

### 3.1 Random Forests Regression

Five Random Forests regression (RFR) models were developed in Python: three RFRs all individually for Sentinel-1, Sentinel-2 and Landsat-8, one for a combination of the important predictors from all these sensors (Optimized Multi-Sensor Predictors), and the fifth for running a combination of all predictors from all three sensors (Multi-sensor Predictors). After running the RFR for individual sensors, the remote-sensing-based predictor variables higher than the 0.05 variable importance threshold were selected based on Section 2.6. The training and testing of data took an average of 40 seconds. Ancillary data layers were included in all five models regardless of their variable importance score. An example of Python script is in Appendix G.

#### 3.1.1 Individual Sensors

##### A. Sentinel-1 SAR Data

Thirty-one features were used as an input variable for the Random Forests regression. The list of feature importance values is shown in Figure 13. The GLCM Variance in VH polarization was the only Sentinel-1 variable above the 0.05 threshold. Thus, it was the only Sentinel-1 layer that was used in predicting soil salinity. The training and testing of this model resulted in an RMSE of 0.19,  $R^2$  of 0.59, and a Pearson correlation coefficient of 0.77. Their correlation between predicted salinity and field data is shown in Figure 14.

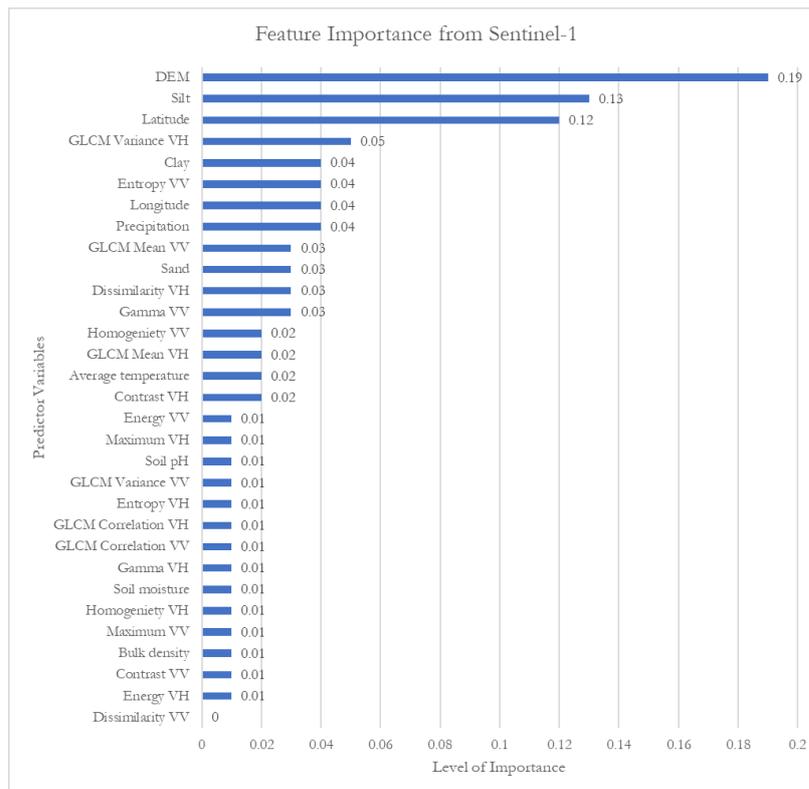


Figure 13. List of feature importance from Sentinel-1

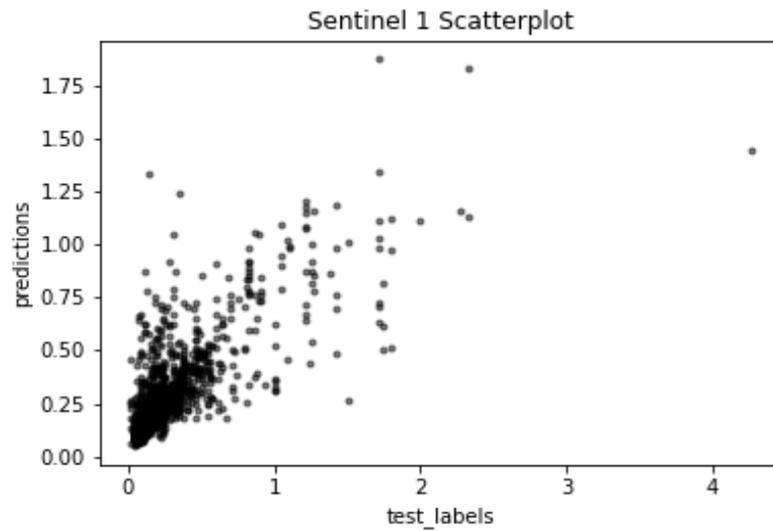


Figure 14. Scatterplot of field-measured vs. predicted soil salinity from Sentinel-1

#### B. Sentinel-2 Multispectral Optical Data

There were also a total of 31 features that were used in the training and testing of data from Sentinel-2. The list of feature importance values from Sentinel-2 was shown in Figure 15. Two of the seventeen Sentinel-2 predictor variables, the Simple Ratio Vegetation Index (RVI) and Soil Adjusted Vegetation Index (SAVI), were above the 0.05 threshold. This model's training and testing resulted in an RMSE of 0.16,  $R^2$  of 0.69, and a Pearson correlation coefficient of 0.83. The scatterplot of soil salinity predictions versus field data is in Figure 16.

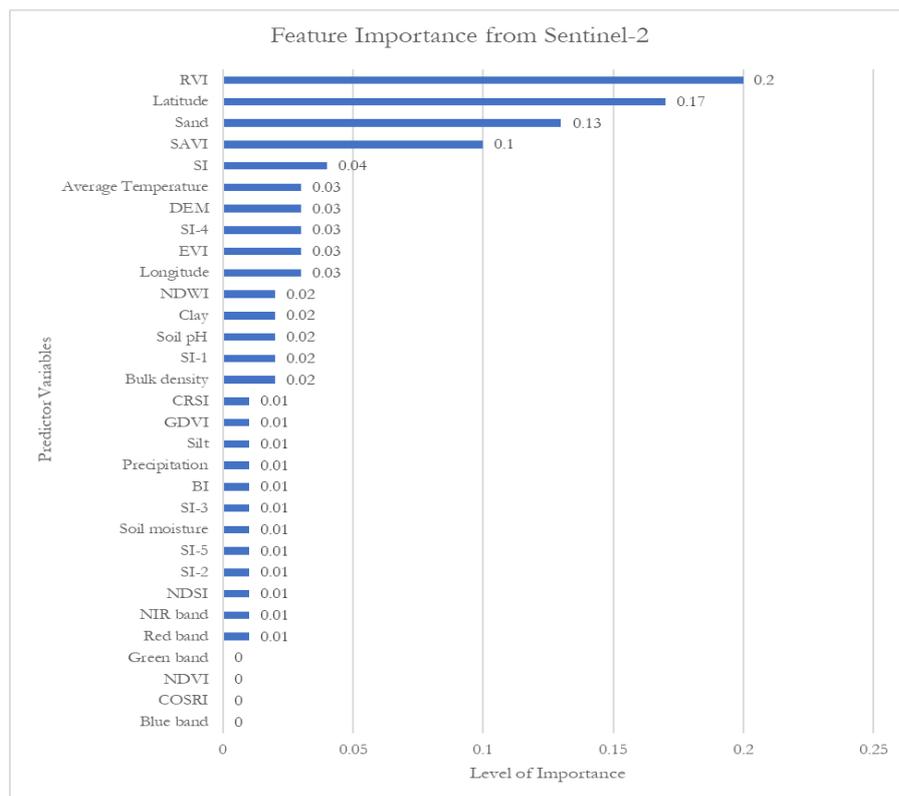


Figure 15. List of feature importance from Sentinel-2

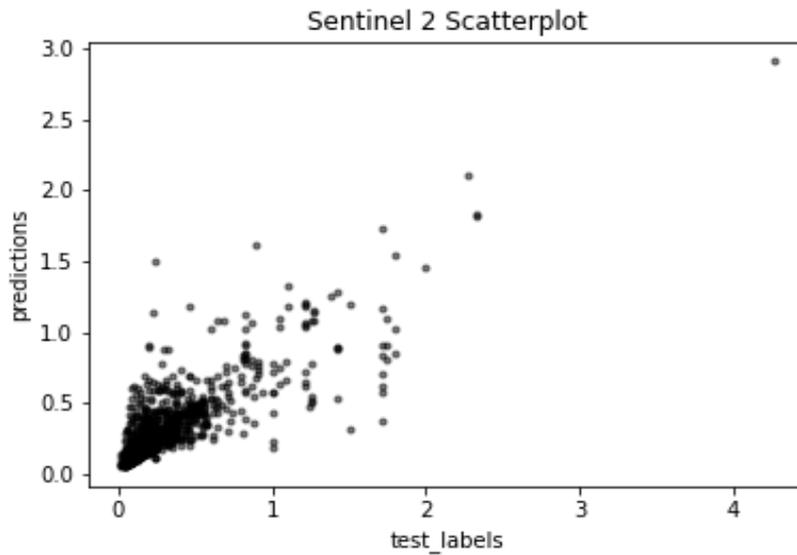


Figure 16. Scatterplot of field-measured vs. predicted soil salinity from Sentinel-2

### C. Landsat-8 Thermal Data

The only feature that was generated in Landsat-8 is the land surface temperature. The list of feature importance values is shown in Figure 17. Since the LST is within the threshold, it was used as an input for the OMSP in predicting soil salinity. The LST alone with ancillary data resulted in RMSE of 0.14,  $R^2$  of 0.77, and a Pearson correlation coefficient of 0.88, which is the best performing among the three individual sensors and indicating that the thermal bands of Landsat-8 are a potential source of detecting and predicting soil salinity. The scatterplot of field-measured and predicted soil salinity is shown in Figure 18.

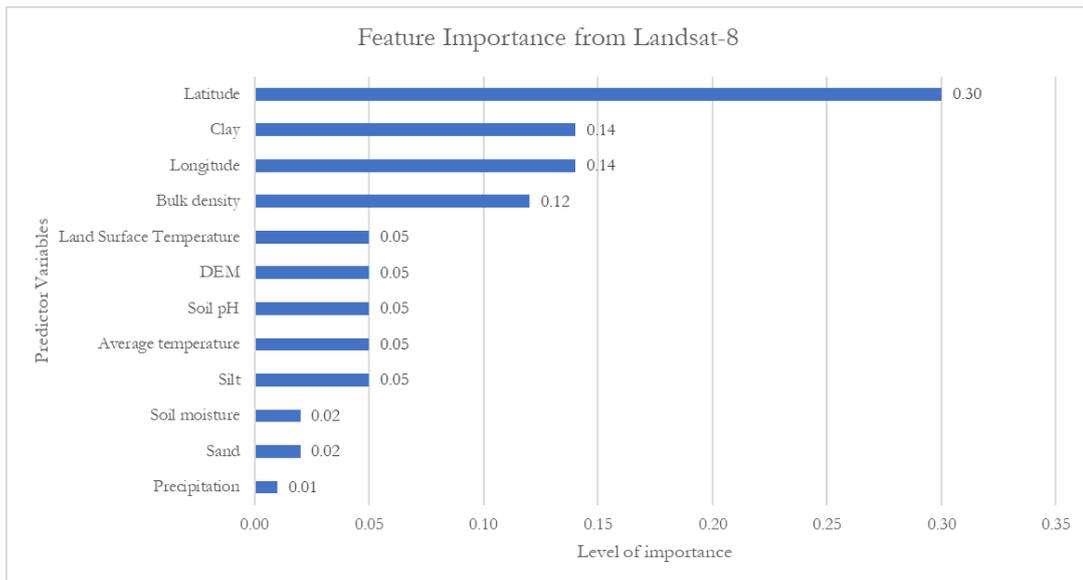


Figure 17. List of feature importance from Landsat-8

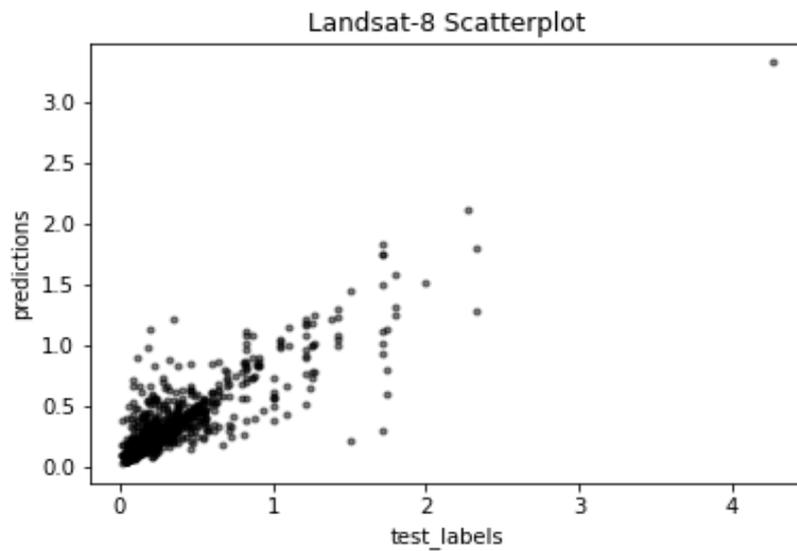


Figure 18. Scatterplot of field-measured vs. predicted soil salinity from Landsat-8

### 3.1.2 Optimized Multi-Sensor Predictors

The important features of individual sensors from sections 3.2.1-A, B, and C were used in this Random Forests regression and ancillary data. The list of important features is shown in Figure 19 below.

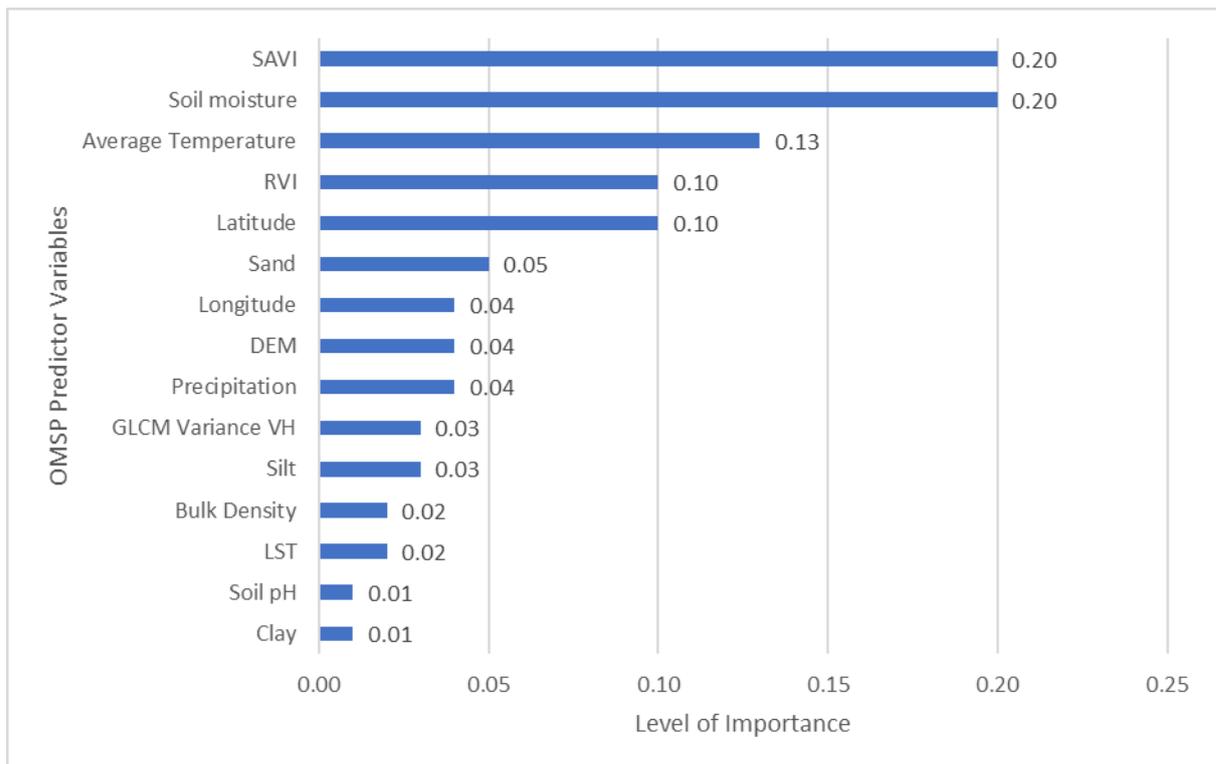


Figure 19. List of feature importance for OMSP

Figure 19 shows the importance of both sensor data and ancillary data in the Random Forests regression model. Remote sensing data from SAVI contributes the most weight (0.20 importance) from Sentinel-2.

SAVI is used when vegetation is low, and the soil surface is exposed. It functions the same as NDVI, but this index has corrected the soil reflectance factor. The reflectance of light in the red and near-infrared spectra influences the vegetation index value (USGS, n.d.). In addition, SAVI is usually calculated to detect and map the extent of a healthy vegetation cover, enhance the differentiation of saline areas, suppressing the vegetation (Mokarram et al., 2015).

Soil moisture was also found to have the same importance level as SAVI (0.20), although the data used has a very coarse spatial resolution (~4km). Soil moisture is one of the key factors in assessing salt-affected soils because salt concentration depends on soil moisture. This predictor is followed by the average temperature, which can be interrelated to soil moisture. High temperatures will lead to less soil moisture due to evaporation. Latitude was also found to be highly important. Due to the geographic location and layout of the study area, differences in latitude results in differences in the soil salinity. Also, latitude affects the differences in local climates, such as the temperature and precipitation, as mentioned in section 3.1.2.

Another remote sensing layer generated from Sentinel-2 is the Simple Ratio Vegetation Index (RVI). This index, simply the soil reflectance in the NIR band divided by the reflectance in the red band, is the quickest way to distinguish the green leaves from other objects and estimate the relative biomass present in the image (Hiphien, 2021). A high value of RVI usually indicates healthy vegetation, and lower values indicate soil, water, or ice areas (Humboldt State University, 2014). These values of RVI could be related to assessing soil salinity in bare soil areas and remove the effect of vegetation.

Precipitation, DEM, and longitude have the same level of importance (0.04). Precipitation influences the average temperature and soil moisture. Precipitation, rainfall, or irrigation to salt-affected soils may wash away the salt particles and avoid salt accumulation. The DEM, which measures the ground surface elevation from the mean sea level, was also important. DEM has been used in several studies to map landforms and delineate areas affected by soil salinity (Samra & Ali, 2018), especially in low plains (Yahiaoui et al., 2015). Furthermore, Ali et al., 2016 stated the importance of DEM that land surface is highly considered in assessing and managing salt-affected areas. It was also mentioned in the geology of the study area and shown in Figure 6 that the water is being drained to the west (towards the West Philippine Sea). Longitude has lower importance than latitude due to the layout of the study area, that there's more variation in latitude than longitude. While latitude is related to climate, longitude is associated with the distance of areas affected by soil salinity to the coast or sea.

The next features were about the soil's physical properties, even at a low importance level. Soil texture determines how much water will be able to pass through the soil. The presence of water affects the salt content in the soil. If there is a good soil texture, the salt particles will be washed away and will not accumulate on the soil. Silt and clay, which have 0.03 and 0.01 importance, respectively, have a smaller particle size, and smaller particle size can pack closely together. Small particle size blocks the particles' spaces and prevents the water and salt particles from passing through. The sand has a larger particle size, resulting in less surface area, and thus salt particles will not accumulate in these areas. Bulk density affects the water flow between soil particles due to soil porosity, and soil moisture determines the amount of water stored in the soil. Furthermore, bulk density is dependent on soil texture. Sandy soils have relatively high bulk density due to particle size.

Land surface temperature from Landsat-8 that utilizes the thermal bands of this sensor is a newly added information in soil salinity mapping. This feature alone, together with the ancillary data, garnered the

highest correlation value in predicting soil salinity. However, no correlation can be seen in comparing the LST and field-measured EC. This explains the importance and effects of using ancillary data. This could be related to the climate data and soil moisture in the study area.

The last remote sensing data that performs well in the model is the GLCM Variance in VH polarization generated from Sentinel-1. The GLCM Variance performs the same tasks as the common descriptive statistic called “variance.” It relies on the mean and the dispersion around the mean of cell values within the GLCM images. Furthermore, it is mainly associated with the visual edges of land cover patches, which explains the relationship of DEM and GLCM Variance. DEM differentiates the physical land characteristics in terms of elevation. GLCM Variance is used to extract information on land cover and bare soil within the same elevation group. It was also found out that this feature works better in VH polarization. Using C-band, the bare soil areas are clearly visible on this feature and strongly correlated in VH polarization (CEOS.org, 2018). The absence of standing crops increases this correlation. It separates it from VV polarization because plant growth, especially in newly planted areas, can cause confusion where soil roughness is likely to be the leading cause for scattering. This feature was also the most important feature in the soil salinity mapping conducted by Hoa et al., 2019 in Vietnam, utilizing Sentinel-1 data alone to model soil salinity in a machine learning environment.

The last important feature in the model is the soil pH level. We can recall that the study area has been selected due to its high soil pH values. Agricultural soils are turning into alkaline soils, and that alkalinity is related to soil salinity. Soil pH has a low importance value because the main effect of soil pH on soil is the mobility of macro and micronutrients present in the soil (California Envirothon, 2017). In addition, it influences the rate of biochemical breakdown or mineral weathering in the soil; thus, the remote sensing and ancillary data for this study cannot be detected. With all the predictor variables identified, the statistical measures of RMSE,  $R^2$  and PCC are shown in Table 15 below.

Table 15. Statistical Metrics of OMSP Model

Sensor	Important Predictor/s	Metrics		
		RMSE	$R^2$	PCC
Sentinel-1	GLCM Variance VH	0.13	0.80	0.90
Sentinel-2	Soil Adjusted Vegetation Index and Simple Ratio Vegetation Index			
Landsat-8	Land Surface Temperature			
Ancillary	Soil property, climate, and geographical location			

The metrics from Table 15 above show that the selected features yielded an RMSE of 0.13, which indicates that the model is performing well in predicting the soil salinity. The  $R^2$  value of 0.80 shows that the relationship of the selected predictors in predicting soil salinity is high and that 80% of the data fits the model. The PCC value of 0.90 shows a very high positive relationship. In addition, a scatterplot of the field-measured vs. predicted soil salinity is shown in Figure 20.

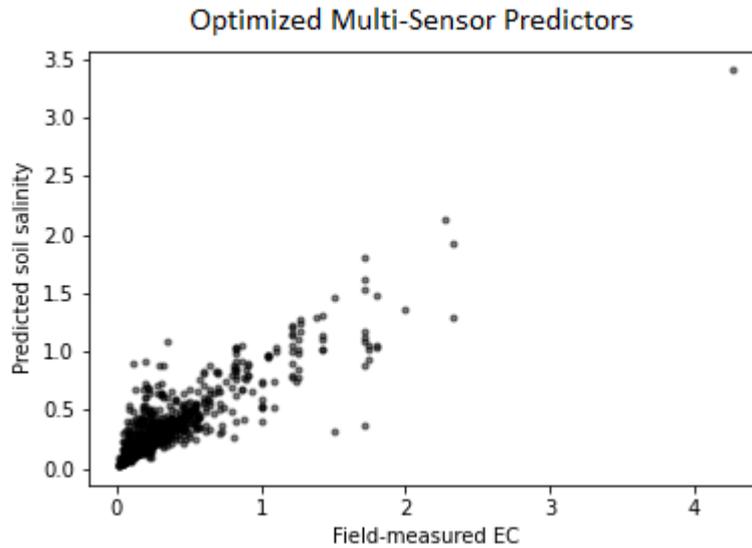


Figure 20. Scatterplot of field-measured vs. predicted soil salinity from OMSP

### 3.1.3 Multi-sensor Predictors

The list of feature importance values from MSP is shown in Figure 21 below. The satellite-based predictor variables within the 0.05 significance threshold are Energy from Sentinel-1 and RVI and SAVI from Sentinel-2. These important predictor variables were used in the model. Figure 22 shows the list of important features of the final MSP model.

The RVI is the most essential feature of the MSP model. This result is surprising because RVI has been widely used for areas with vegetation while the study focuses on bare soils. However, if we look again at the formula of RVI, it utilizes the NIR and red bands. The reflectance of soil in the electromagnetic spectrum is best on these bands. SAVI is also of high importance. Similar to RVI, this index also utilizes the NIR and red bands.

The importance of geographical location in this method is similar to the OMSP method. Bulk density, which has low significance from the OMSP method, became the 3<sup>rd</sup> most important feature in this approach. This feature is followed by the soil properties, DEM, and soil pH. The last important features were climate data and soil moisture.

The Energy in VH polarization generated from Sentinel-1 is the least important feature in this method. Although considered least important, it is still performing in the model by affecting the input features. The texture is one of the essential spatial features of an image (Kupidura, 2019). Energy is one of the image textures from GLCM, is the orderliness of pixels that share similar backscatter characteristics that is beneficial in land-use and land-cover classification. The use of the Energy feature in the model is it helped in classifying the land cover in the image. Similar to OMSP, this predictor variable in Sentinel-1 also gave information in VH polarization. The results of metrics for MSP are shown in Table 16.

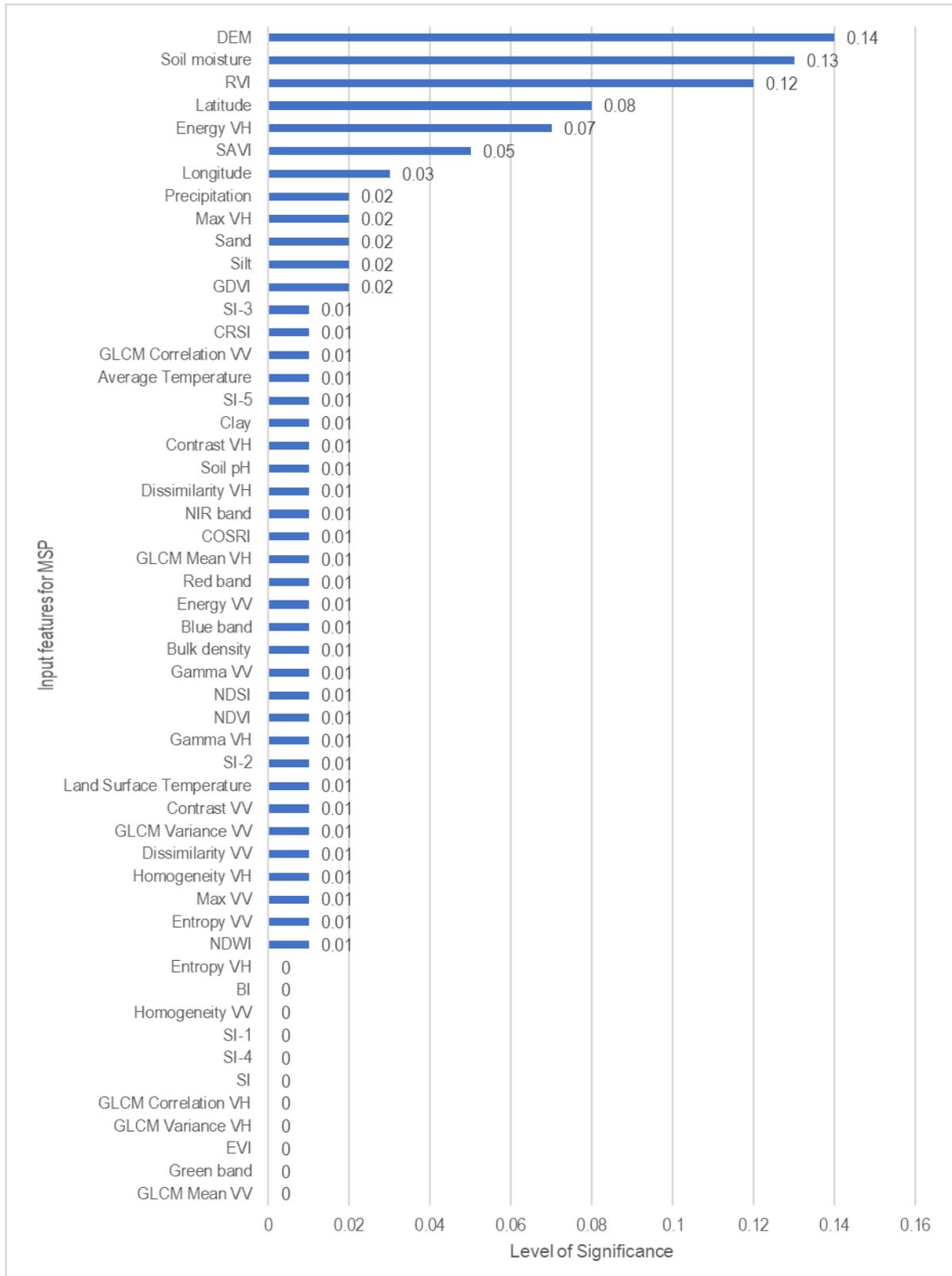


Figure 21. List of input features and their importance values from MSP

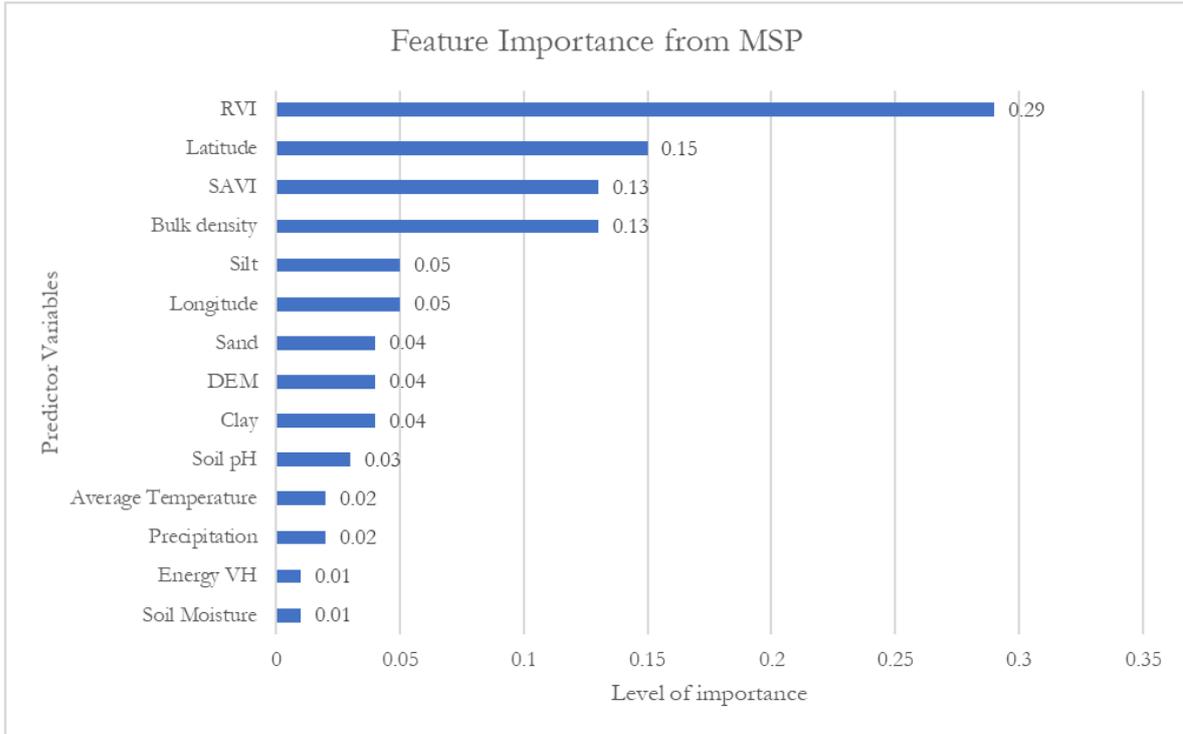


Figure 22. List of Feature importance from MSP

Table 16. Statistical Metrics of MSP Model

Sensor	Important Predictor/s	Metrics		
		RMSE	R <sup>2</sup>	PCC
Sentinel-1	Energy VH	0.13	0.82	0.91
Sentinel-2	Soil Adjusted Vegetation Index and Simple Ratio Vegetation Index			
Ancillary	Soil property, climate, and geographical location			

The Multi-Sensor Predictors' metrics for predicting soil salinity showed slightly higher accuracy than the OMSP approach. The RMSE is 0.13, indicating a good fit, R<sup>2</sup> value of 0.82, and a PCC of 0.91, indicating a very high correlation. Thus, the variables used in this model have a positive strong relationship. This relationship can also be seen on the scatterplot of field-measured soil salinity and predicted soil salinity in Figure 23. This result is not that different from the first method; however, there is a change in the remote sensing data that is important from Sentinel-1. Upon training and testing the data, the LST had an importance value of 0.01; thus, it was not included in the final model.

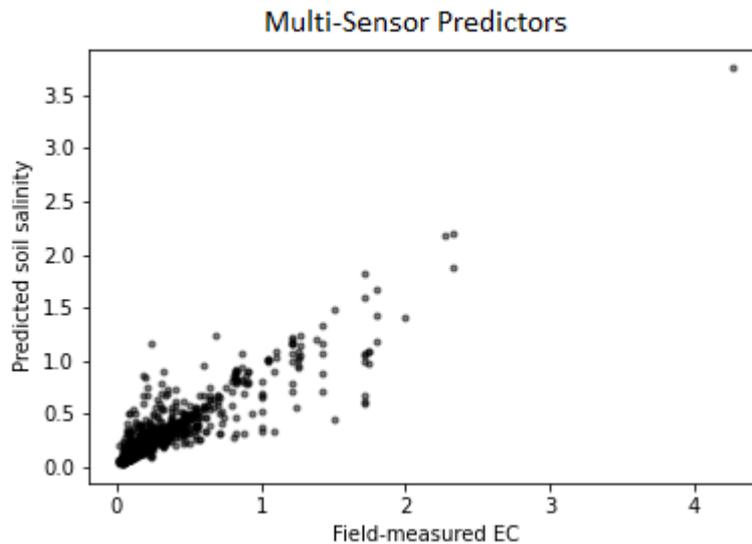


Figure 23. Scatterplot of field-measured vs. predicted soil salinity for MSP

### 3.2 Predicted Soil Salinity

Two soil salinity maps were produced using the two different approaches. Figures 24 and 25 show the predicted soil salinity based on these approaches. The ranges of the predicted EC were 0.25 to 1.20 dS/m and 0.3 to 1.9 dS/m for OMSP and MSP, respectively. Generally, this range is different from the field-measured EC with a range of 0.02 to 4.26 dS/m. We can recall that the scatterplot of field-measured vs. predicted soil salinity from Figures 20 and 23 that both approaches can predict soil salinity up to 3.5 EC. The model then used this range in predicting soil salinity based on input maps; however, most of the points are concentrated on the lower soil salinity values. The Random Forests regression, by its function, averages the predictions within the range that is mostly between 0 to 2.0 EC. This feature of Random Forests explains why the predicted soil salinity in both methods resulted in even lower values in the resulting maps. In addition, the areas with predicted soil salinity are limited to the rice areas. During the processing (masking) of predicted soil salinity images in ArcMap, the original values of soil salinity are higher than the EC ranges mentioned above but decreased when rice areas mask was applied.

Both of the methods show that the predicted soil salinity is evident in the coastal areas. However, the MSP approach detected more areas with higher soil salinity values than the OMSP method. In addition, the spatial distribution of the predicted soil salinity in MSP is similar to field data visually. In the OMSP approach, the predicted soil salinity showed that the soil salinity is evident in small patches located in coastal parts of Sinit, Cabugao, City of Vigan, Caoayan, Santa, Santa Maria and City of Candon. Moderately high EC (yellow-green) were also observed near the river systems near the coastal area. Low EC was found on areas alongside the river systems, however, in between mountain valleys. The areas with low EC values (0-0.3 dS/m) have 93.36 hectares or 1.19% of the rice areas, whereas moderately low EC (0.3-0.6 dS/m) have 7,580.19 hectares 96.81% of the rice areas. Lastly, the high EC values (0.9 to 1.2 dS/m) have 156.52 hectares or 2.00% of the rice areas.

In MSP, high soil salinity values on this range are located on coastal parts of Sinit, San Juan, San Vicente, Santa Catalina, City of Vigan, Caoayan, Santa, Santa Maria, San Esteban, City of Candon, Santa Lucia, and Santa Cruz. Moderately high EC is also located in flat or low-lying areas but is farther to the coast. Low EC values are located in the in-field valleys. The EC values that this method has predicted are higher than

the individual sensor method. The areas with low EC (0.0-0.5 dS/m) has a total area of 6,461.63 hectares or 82.52% of the rice areas; moderately low EC (0.5-1.0 dS/m) has an extent of 106.23 hectares or 1.36%; moderately high EC (1.0-1.5 dS/m) has a total area of 1,260.26 hectares or 16.10% of the rice areas, and the high EC (1.5-2.0 dS/m) has a total area of 2.01 hectares or 0.03% of the rice areas. Appendices H and I show the variability of soil salinity per municipality using OMSP and MSP, respectively.



Figure 24. Predicted soil salinity map based on Optimized Multi-Sensor Predictors Approach

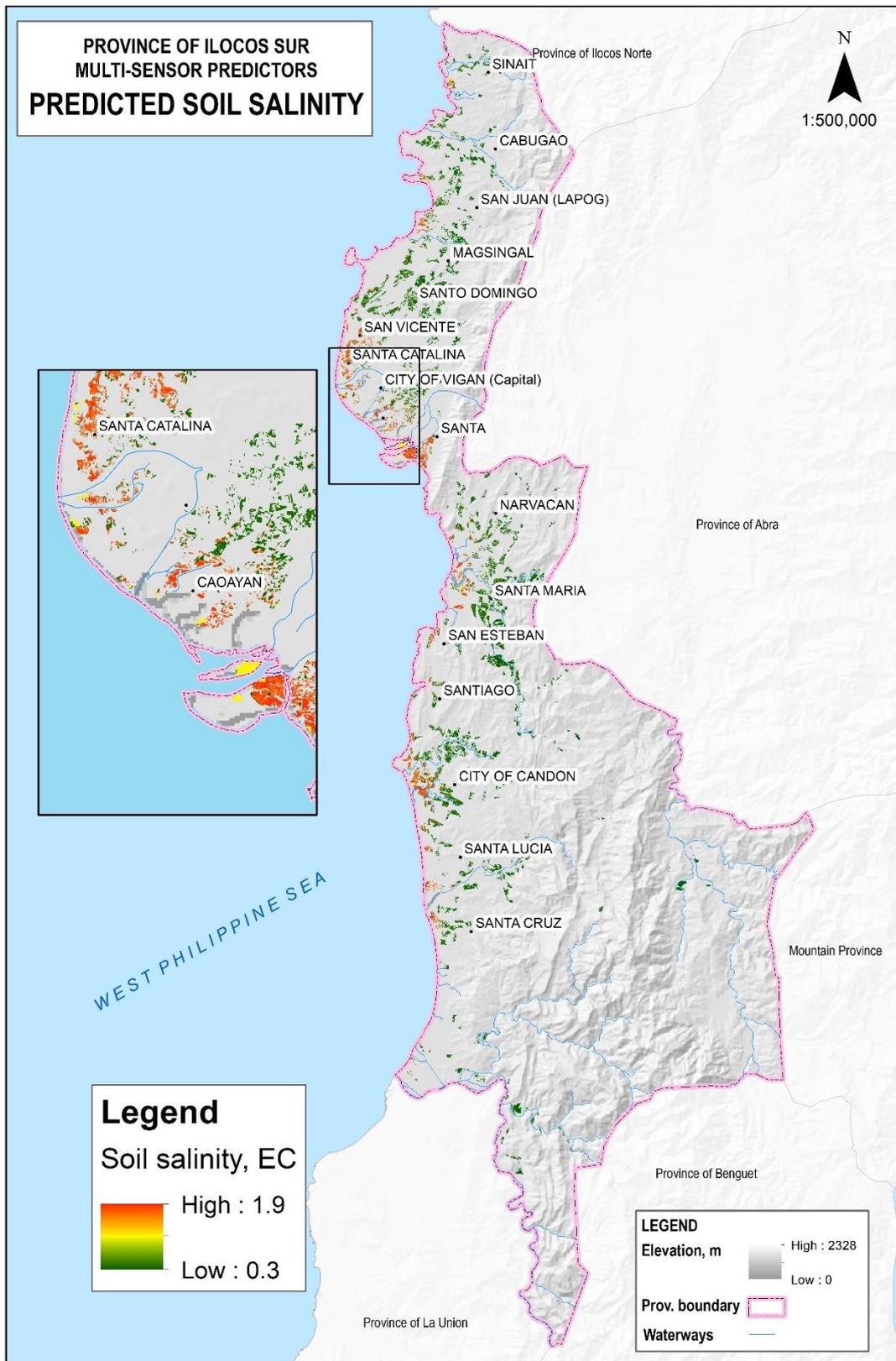


Figure 25. Predicted soil salinity map based on Multi-Sensor Predictors approach

### 3.3 Spatial Variability

Generally, high levels of EC are observed in the coastal areas of the study area. These are located in the coastal municipalities with soils of clay, silt loam, and sandy textures. Among these are the soils of Bantog clay found at Sinait, Cabugao, Sta. Maria, City of Candon, Narvacan, and Sta. Lucia; San Manuel silt loam soils located in Sinait, San Juan, San Vicente, Santa Catalina, Sta. Maria, Magsingal, City of Candon, Sta. Lucia, and Sta. Cruz; Umingan sandy loam soils in Santa.

Some of the rice areas are expanding on the beach sandy soils in the City of Vigan, Sta. Catalina and Cabugao and also in the river wash soils on Santa with predicted soil salinity of moderately high. This result is surprising because these types of soils are not suitable for rice production.

Area-wise, the MSP approach predicted more areas with moderately high soil salinity than the OMSP approach. Using the soil salinity levels described in Table 14 of section 2.8.1, only the municipality of Santa has the highest level of soil salinity (1.5 to 2.0 dS/m). Moreover, it can be seen from the produced maps and spatial extent of predicted soil salinity that the municipality of Santa and the City of Candon are the locations that experience high salt concentration in their rice areas.

### 3.4 Multiyear Variability

The Multi-sensor Predictors model was used to predict soil salinity for 2017, 2018, and 2020. Figures 25 shows the multiyear change of soil salinity in the Province of Ilocos Sur.

Visually, changes in the soil salinity in the years 2017 to 2020 can be seen easily on coastal rice areas. From 2017, the soil salinity in the rice areas in San Vicente, Santa Catalina, City of Vigan, Santa, Narvacan, and City of Candon started with EC values ranging from 1.0 to 1.5 dS/m. However, in 2018 these areas had increased in soil salinity ranging from 1.5 to 2.0 dS/m. In addition, low soil salinity values in some rice areas of Narvacan, Santa Maria, and San Esteban became moderately low in the year 2018. Recalling from the MSP approach that the predicted soil salinity is for 2019, the areas affected by soil salinity are also in coastal areas but more evident in fewer coastal municipalities. This variability can be related to the drought in Ilocos Sur last 2017, and extended drought has been of concern for the province (Toldo, 2019). In September 2018, a strong typhoon with the local name of *Ompong* had caused up to six-meter storm surges that hit the province's coastal areas (ABS-CBN News, 2018). Another typhoon named Rosita had also made landfall in the province last October 2018 (Rappler, 2018). Although no storm surges have been reported, this typhoon might have pushed the salt particles into the sea, causing the soil salinity for the summer of 2019 low.

In March 2019, droughts and El Niño were reported to affect the province in the same year. This phenomenon brought less rainfall and warmer temperatures (Flores, 2019). Some typhoons passed by the province in November 2019; however, this has not landed and affected the province, but more on the Province of Ilocos Norte (Arceo, 2019), approximately 200 kilometers northward away from the City of Vigan to the City of Laoag. Come 2020, and the rice areas experienced an increase in soil salinity again. The presence of drought and less rainfall during the previous year might have caused this increase in soil salinity in coastal rice areas. Since there is no field-measured soil salinity available, there is no data validation of whether droughts and typhoons have caused multiyear changes in soil salinity. However, with the general process of soil salinity shown in Figure 1, climate conditions can still be related to soil salinity dynamics.

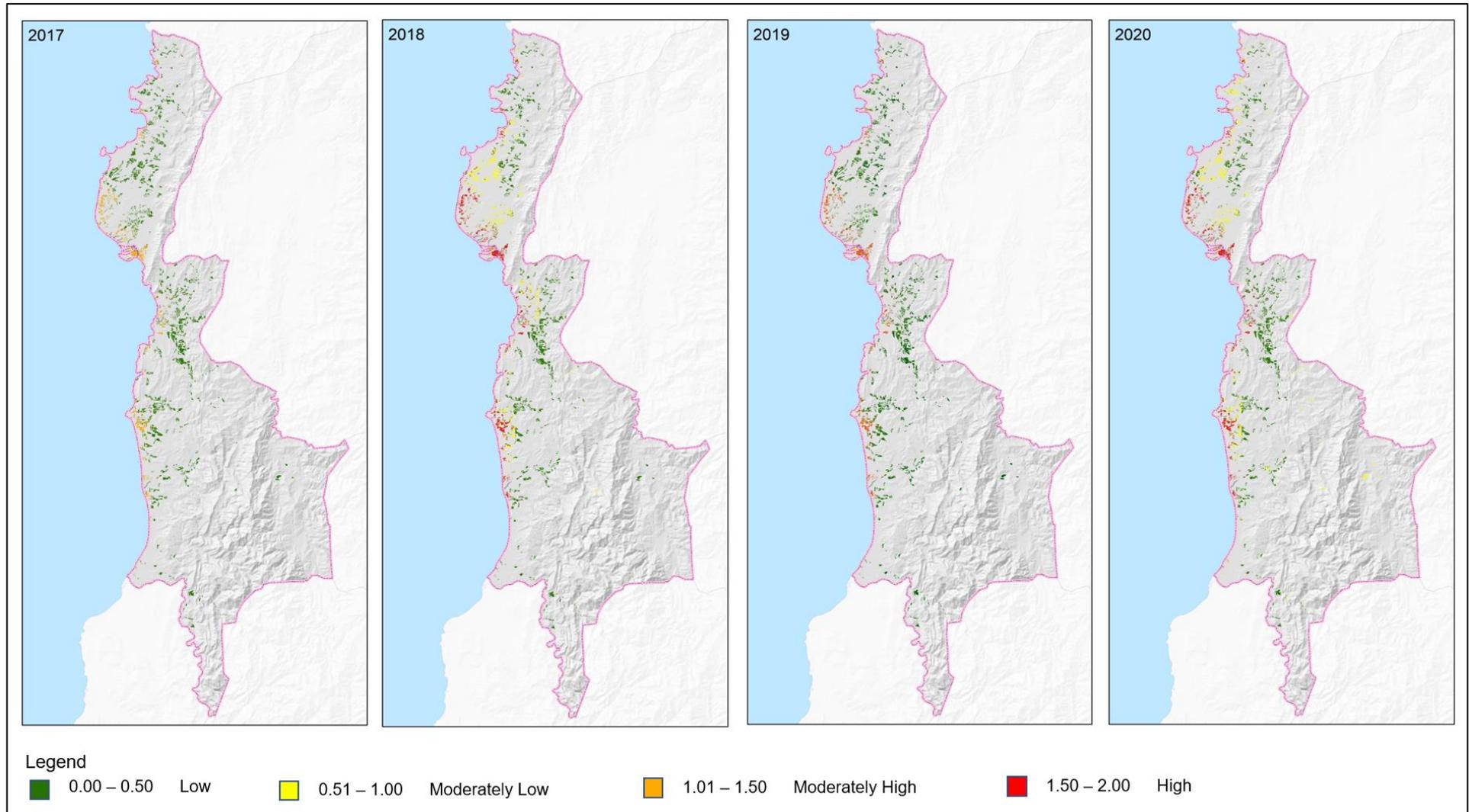


Figure 26. Predicted soil salinity for years 2017 to 2020

Most of the areas with high soil salinity values were identified in the northwest part of the province, around the City of Vigan. Figure 26 shows the soil salinity change in these municipalities. From 2017, the coastal rice areas had moderately high soil salinity in San Vicente, Santa Catalina, City of Vigan, Caoayan, and Santa. However, in 2018, these areas became high in soil salinity. As a result, most rice areas in Santo Domingo and the City of Vigan went from low to moderately low in soil salinity. In 2019, the areas with high soil salinity 2018 remained high, and the municipalities of Santo Domingo and the City of Vigan reached a low level of soil salinity again. The pattern and distribution of soil salinity in 2020 are not that different in the year 2018. High soil salinity values are still evident in San Vicente, Santa Catalina, Caoayan, and Santa. However, a small portion of the coastal rice area in Santo Domingo had a low salinity level in 2020. In the municipality of Santa, fewer patches of rice areas remained at low salinity levels in 2020 than in 2018.

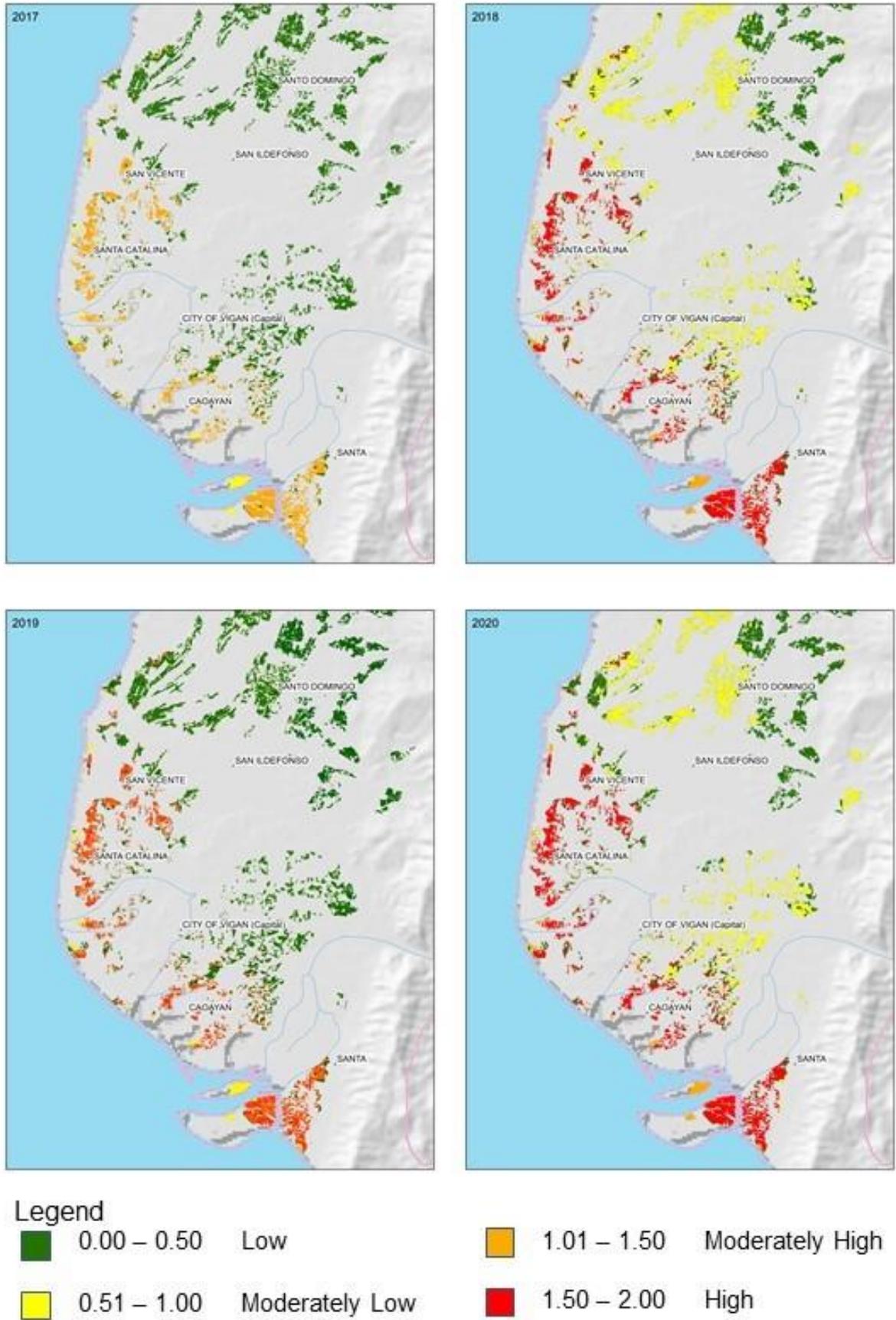


Figure 27. Soil salinity change in most affected areas

### 3.5 Soil Salinity Change

Soil salinity change was generated by computing the difference of predicted soil salinity of 2017 and 2020. The change in soil salinity is shown in Figure 27; generally, there is an increase in soil salinity in all rice areas. The areas under green show an increase of soil salinity by a minimum of 0.04 dS/m for the whole four years. This change increases and is most evident in areas under red colors. These rice areas had increased soil salinity by 0.15 dS/m, maximum from 2017 to 2020. The soil salinity change also indicates that significant changes in soil salinity were observed in coastal areas and areas that low changes in soil salinity were found nearer to undulating to hilly areas. The general surface cross-section of the study area from Figure 7 explains this spatial variability and change in soil salinity. Due to hilly and undulating areas, salt particles were pushed to the west, coastal and low-lying areas.

### 3.6 Management of salt-affected soils

This section outlines the possible utilization of data generated in this thesis.

#### 3.6.1 For policy-making bodies

Soil salinity data is evident in the province of Ilocos Sur and mainly detected on coastal rice areas. Areas high in soil salinity significantly changed over time and were identified and needed to be taken care of, especially with better and sustainable soil management. Therefore, policy-making bodies could include soil salinity information in developing time-efficient and budget-friendly projects that maintain agricultural soils' health. Furthermore, since areas high in soil salinity are also provided, they can use the information for project prioritization.

For the national government level, DA-BSWM, this research serves as a basis and benchmark of soil salinity in coastal agricultural areas. Because they are mandated to map and provide information about agricultural soils, they can use the data to add to their soil information database. They can also recommend strategies to the local government down to farmer level on the sustainable use of land and water resources. For example, proper fertilizer formulation that could not increase the soil salinity can be suggested that contains the amount, timing of application, and the number of applications. There are also rice areas that have been detected in soils that are not suitable for rice production: beach sand and river wash. They can recommend other crops that are more suitable to produce in these areas because rice production might exacerbate the soil salinity. For water management, since they have a database on the areas that use pumps to extract groundwater for irrigation, they could recommend other irrigation sources. Some of these can be rainwater harvesting facilities or small-scale dams, which are also part of their functions. To monitor the local changes in soil salinity, they can use soil moisture data generated aerial images from UAVs with higher spatial resolution.

Another national-level government that can be of concern is the Philippine Rice Research Institute (PhilRice). PhilRice aims to develop and promote technologies that are ecosystem-based, location, and problem-based. For example, the introduction of salt-resistant rice seeds can be distributed to the farmers severely affected by soil salinity. Their publication, *Management of Salt-Affected Soils for Rice Production*, is a good manual for managing soil salinity that includes cultural management of salt-tolerant seeds.

Local government units such as the provincial and municipal levels could disseminate this information to the farmers during their weekly or monthly Municipal Agriculture Officer's (MAO's) meeting. Through

this, they can work with farmers on rehabilitation, conservation, and management strategies together with two national agencies mentioned above. They could also collect new soil samples in the rice areas to check and monitor the soil salinity and track the changes with the method used in this study.

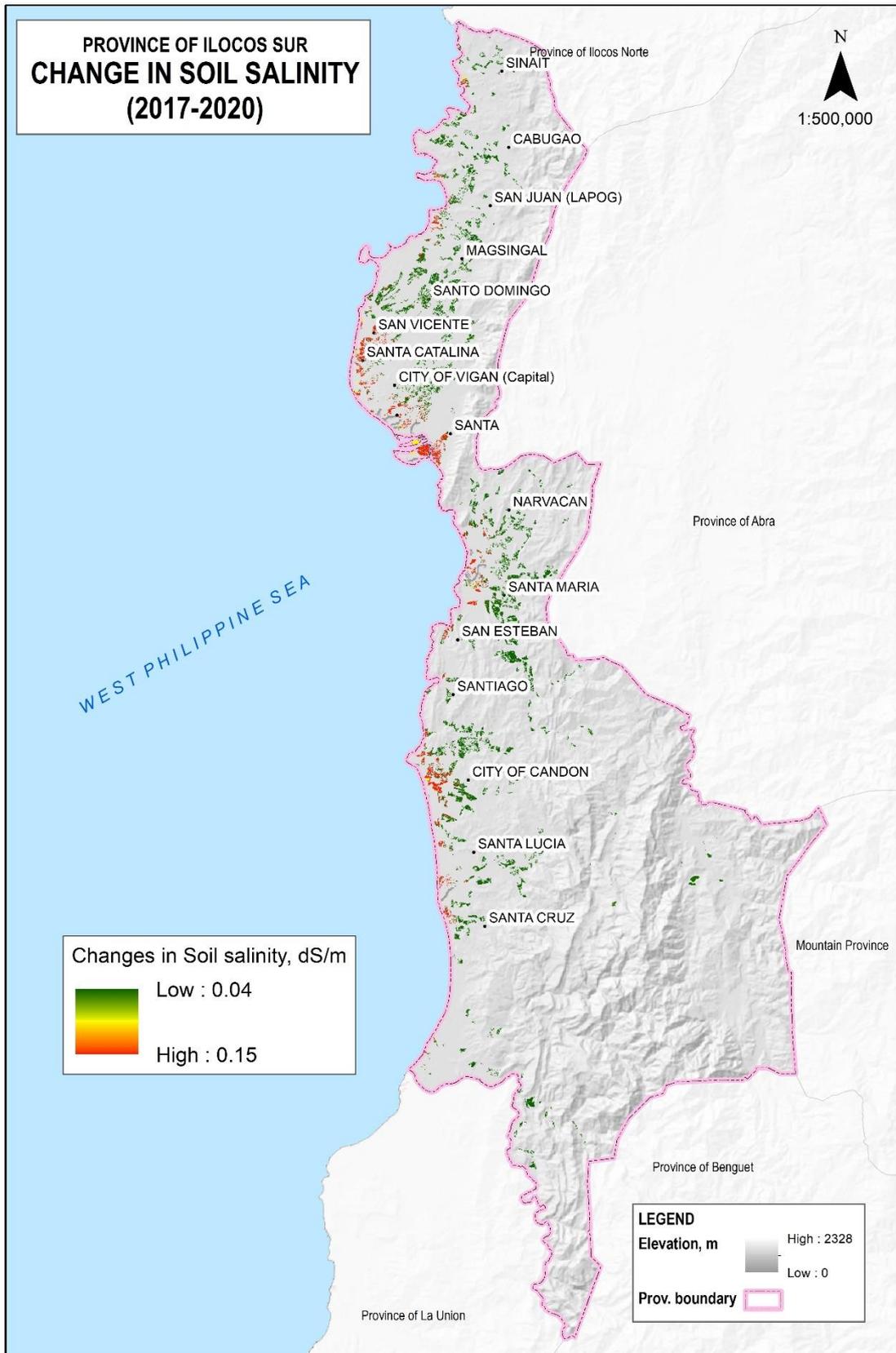


Figure 28. Soil salinity change

### 3.6.2 Farmers and farmer associations

This study provides information on areas affected by soil salinity and the possible farmers that may be affected by it. Once the farmer is informed about the status of their soil, it would be better for them to manage it and the “guessing game” stops. The presence of drainage, limitation of groundwater extraction, and good irrigation schedules could help alleviate the problem of soil salinity. Farmers can also participate with the government's projects in correcting the land and soil by scraping, land leveling, subsoiling, and improving planting techniques. In addition, they can improve the soil texture by applying organic matter, mulching, green manuring, and crop rotation.

### 3.7 Limitations of the Research and Future Work

Like any other scientific model, the methodology used in this study is not perfect. The uncertainties of this research were found on soil data, delineated rice areas, remote sensing data, and the interpretation of results. For the soil data and delineation of rice areas, it was mentioned in Sections 2.3.1 and 2.5.4 that the geographical coordinates of the soil samples do not hold the exact locations where the soil samples were collected. Thus, new random points were created inside the delineated rice areas that served as the representative soil samples in the area. However, the values of the electrical conductivity of these points are duplicated. Therefore, it is essential to take note of the geographical coordinates where the sub-soil samples are collected.

Furthermore, the delineated rice areas are manually made by the staff of DA-BSWM using Google Earth. This data may not have been very accurate due to differences in spatial resolution and imagery date. The delineated rice areas also include the rice areas with the standing crop, as mentioned in Section 2.5.4. The bare soil areas within the delineated rice areas were estimated using the averaged NDVI image, which may not also include a bare soil area in the ground. For future work, it is recommended to have more soil samples, and the exact geographical coordinates of soil samples were well noted.

Another limitation of the study is the availability of satellite data. The satellite data of the three sensors were averaged per date to compensate for the period of soil sampling. These were averaged because different image tiles with varying image dates can cover the whole study area in Sentinel-1. Keeping the images not averaged would be a promising avenue for analyzing the spatiotemporal variability of soil salinity in a limited time window. There is also a limitation on the relationship between soil data and remote-sensing data. The image features extracted in the satellite data are related to soil salinity at the soil surface, whereas the soil samples were taken 0 to 30 centimeters below the soil surface. The remote sensing data tells the salinity in the surface; the field data represents the soil salinity at the subsoil. However, they are assumed to be related because the soil samples were taken when the surface is not yet disturbed, and parts of the soil surface were still analyzed in the laboratory. Utilization of higher resolution remote sensing data is highly recommended for better prediction of soil salinity. Furthermore, only the soil surface salinity has been assessed in this study. Some studies show the importance of analyzing the root-zone soil salinity, especially when the soil salinity per season (wet season) was added per year.

We considered other methods for identifying relationships between soil salinity and remote sensing-based information, such as General Additive Models (GAMs). GAMs provide a simple solution between variables to see which fits with the linear regression. GAMs have been used in many studies in assessing soil salinity. However, there are limitations on the complex relationships between variables. Since we had 59 explanatory variables for predicting field-measured soil salinity, we opted for a machine learning method that could handle interactions between so many variables.

One machine learning method was used in the study, although many models could be used. Other machine learning algorithms such as Neural Network, Support Vector Regression, Gaussian Process, and many more can also be explored to see which ML methods the multi-sensor approaches perform best. The use of Random Forests was due to its characteristics such as fast training speed, handling of high dimensional data, and there is low bias in each decision tree based on the previous related studies. In addition, RF has a built-in feature selection. According to Genuer et al., 2010 and Grömping, 2009, the variable importance based from RF is efficient for problems that has high number of input variables and low number of samples.

However, the interpretability of the model is the main drawback of this machine learning algorithm. Most of the uncertainties of this study are concentrated on the interpretation of predicted multiyear soil salinity. Since the model for the year 2019 was used for predicting soil salinity for other years, there are no ground-truth data available that could support it. However, they can use spatial and temporal variability to guide future soil sampling or cross-checking. Testing the model in other areas and other crops are also recommended to see how well the overall modeling approach can generalize. In addition, spatial variability of soil salinity was also found near the river systems. Therefore, information on the distance to river systems could also be another explanatory variable for future work.

Lastly, this study is not providing information on which human activities contribute to soil salinity but the variability of soil salinity from environmental conditions. Therefore, it is essential to know where and how the soil salinity is becoming a problem in the study area for the more effective spreading of information and making policies by interviewing farmers and getting the actual situation on the ground.

## 4. CONCLUSION AND RECOMMENDATIONS

Soil salinization is one of the most common soil degradation types and is considered one of the world's most widespread soil problems. In the Philippines, the information about soil salinity has not been systematically updated since the 1980s. Soil salinity has been detected in the rice areas in the Province of Ilocos Sur using the combination of field data, remote sensing data, and ancillary data to train a machine learning regression algorithm. The affected areas, if left unmanaged, may increase and could lead to serious environmental problems, especially for rice production. Therefore, it is necessary to identify and determine where soil salinity occurs to rehabilitate the areas affected in its early stage. Two multi-sensor approaches were used in the study: Optimized Multi-Sensor Predictors and Multi-Sensor Predictors. In relation to the research questions expressed for this study, the following are concluded and recommended:

### ***4.1 What is the spatial distribution of soil salinity predicted from the variables generated from Sentinel-1, Sentinel-2, and Landsat-8?***

The predictor variables derived from Sentinel-1, Sentinel-2, and Landsat-8 were used to configure the two models: Optimized Multi-Sensor Predictors and Multi-Sensor Predictors. On a provincial scale, both of the approaches shows that the high values of soil salinity are more evident and clumped in the coastal rice areas of the municipalities of Banayoyo, Bantay, Caoayan, City of Candon, City of Vigan, Magsingal, Narvacan, San Esteban, San Juan, San Vicente, Santa, Santa Catalina, Santa Cruz, Santa Lucia, Santa Maria, and Santiago. Meanwhile, soil salinity becomes lower as the rice areas are located further from the coast side. Geographically, high soil salinity is also located in low-lying areas and mainly in soil textures with clay and sandy loam. It is also important to note that remote-sensed ancillary data plays a significant role in detecting and retrieving soil salinity. Without the ancillary data, the results of the metrics will be lower, and the model will not perform well.

### ***4.2 What is the accuracy of using the multi-sensor approach in retrieving soil salinity?***

There are two multi-sensor approaches used in the study: Optimized Multi-sensor Predictors and Multi-sensor Predictors. These approaches showed excellent performance in predicting and retrieving soil salinity based on the RMSE,  $R^2$  and Pearson correlation coefficient values. The Multi-sensor Predictors model provided better accuracy than the Optimized approach because of higher correlation values with the same error value with RMSE of 0.15, the  $R^2$  value of 0.82, and the Pearson correlation coefficient of 0.91. This approach also utilized fewer predictor variables than the Optimized by not including the land surface temperature from Landsat-8 because the LST was not within the significance threshold. It was attempted to include LST in configuring the model, although the results of the statistical metrics show that the RMSE increased and the correlations decreased. Therefore, LST has been influenced by other variables from other sensors upon the first training of the data. It can also be concluded that integrating the three sensors is more efficient and more accurate in detecting and retrieving soil salinity than using a single sensor. The ancillary data also plays a vital role in the configuration of model of the two approaches. Even though they are outside of the significance threshold, they still affect the metrics of the model and maintaining all of them resulted in higher accuracy of predicting soil salinity.

#### ***4.3 To what extent and pattern do soil salinity change over multiple years based on the multi-sensor approach?***

Using the Multi-Sensor Predictors model, soil salinity changes over multiple years increasingly. For example, soil salinity changed from 2017 to 2020, showing an increase of soil salinity from 0.04 to 0.15 dS/m for four years. From the coastal rice areas, the soil salinity extends inwardly. Although the resulting predictions were not validated with field-measured soil salinity of different years, it is still valuable to understand the salt accumulation with climate. In addition, there is no information yet if the management practices and anthropological activities have affected these changes and increase soil salinity.

In conclusion, the study's approaches present an efficient, cost-effective, and practical expert system to detect, predict, and update the soil salinity in the Philippines. The values of soil salinity predicted in this study are considered low or none to slight according to the USDA's classification of soil salinity. This finding means that soil salinity is not a real problem to deal with. However, it has been detected that soil salinity was increasing per year. Thus, it is assumed that soil salinity is even higher for the year 2021. Utilizing recent data or collecting new soil samples would be helpful for further validation of the model.

## REFERENCES

- Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., & Hegewisch, K. C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958-2015. *Scientific Data*, 5(1), 1–12. <https://doi.org/10.1038/sdata.2017.191>
- Abbas, A., Khan, S., Hussain, N., Hanjra, M., & Akbar, S. (2013). Characterizing soil salinity in irrigated agriculture using a remote sensing approach. *Physics and Chemistry of the Earth*, 55–57, 43–52.
- ABS-CBN News. (2018, September 18). *6-meter storm surges may hit Cagayan, Isabela, Ilocos Sur: PAGASA* | ABS-CBN News. <https://news.abs-cbn.com/news/09/13/18/6-meter-storm-surges-to-hit-cagayan-isabela-ilocos-sur-pagasa>
- Aldakheel, Y. Y. (2011). Assessing NDVI Spatial Pattern as Related to Irrigation and Soil Salinity Management in Al-Hassa Oasis, Saudi Arabia. *Journal of the Indian Society of Remote Sensing*, 39(2), 171–180. <https://doi.org/10.1007/s12524-010-0057-z>
- Ali, R., Saber, M., Nizinski, J., Montoroi, J.-P., & Zaghoul, A. (2016). Land surface analysis of salt affected soils using DEM and GIS. *European Journal of Scientific Research*, 138(2), 197–202. [https://www.researchgate.net/publication/301290862\\_Land\\_surface\\_analysis\\_of\\_salt\\_affected\\_soils\\_using\\_DEM\\_and\\_GIS](https://www.researchgate.net/publication/301290862_Land_surface_analysis_of_salt_affected_soils_using_DEM_and_GIS)
- Allbed, A., & Kumar, L. (2013). Soil Salinity Mapping and Monitoring in Arid and Semi-Arid Regions Using Remote Sensing Technology: A Review. *Advances in Remote Sensing*, 02(04), 373–385. <https://doi.org/10.4236/ars.2013.24040>
- Allbed, A., Kumar, L., & Aldakheel, Y. Y. (2014). Assessing soil salinity using soil salinity and vegetation indices derived from IKONOS high-spatial resolution imagery: Applications in a date palm dominated region. *Geoderma*, 230–231, 1–8. <https://doi.org/10.1016/j.geoderma.2014.03.025>
- Arceo, A. (2019, November 18). *Ramon nears, northern Ilocos Norte added to areas under Signal No. 2*. Rappler. <https://www.rappler.com/nation/weather/tropical-storm-ramon-lpa-pagasa-forecast-november-18-2019-11am>
- Asio, V., Jahn, R., Perez, F., Navarrete, I., & Abit, S. (2009). A review of soil degradation in the Philippines. *Annals of Tropical Research*, 69–94. [https://www.academia.edu/22898396/A\\_review\\_of\\_soil\\_degradation\\_in\\_the\\_Philippines](https://www.academia.edu/22898396/A_review_of_soil_degradation_in_the_Philippines)
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- California Envirothon. (2017). *SOIL ACIDITY, ALKALINITY, AND SALINITY*. [https://caenvirothon.com/wp-content/uploads/2018/02/Envirothon-07\\_Soil-Acidity-Alkalinity-and-Salinity-ver.-2017.pdf](https://caenvirothon.com/wp-content/uploads/2018/02/Envirothon-07_Soil-Acidity-Alkalinity-and-Salinity-ver.-2017.pdf)
- Carating, R. (2015). *Land Degradation Assessment (LADA) in the in the Philippines: The BSWM-FAO Technical Cooperation Final Report*. [https://www.academia.edu/23584038/Land\\_Degradation\\_Assessment\\_LADA\\_in\\_the\\_in\\_the\\_Philippines\\_The\\_BSWM\\_FAO\\_Technical\\_Cooperation\\_Final\\_Report](https://www.academia.edu/23584038/Land_Degradation_Assessment_LADA_in_the_in_the_Philippines_The_BSWM_FAO_Technical_Cooperation_Final_Report)
- CEOS.org. (2018). *A Layman's Interpretation Guide to L-band and C-band Synthetic Aperture Radar data*.
- Chakure, A. (2019, May 29). *Random Forest Regression. In this blog we'll try to understand...* | by Afroz Chakure | *The Startup* | Medium. Start It Up. <https://medium.com/swlh/random-forest-and-its-implementation-71824ced454f>
- DA-BSWM. (2019, March 4). *BSWM conducts Stakeholders Consultation for National Mapping, Characterization, and Development of Spatial Database for the Coastal Areas Affected by Salinity*. BSWM Online. <http://bswm.da.gov.ph/news/0084/bswm-conducts-stakeholders-consultation-for-national-mapping-characterization-and-development-of-spatial-database-for-the-coastal-areas-affected-by-salinity>
- DA-BSWM. (2021). *National Mapping, Characterization and Development of Spatial Database for the Coastal Areas Affected by Salinity*. <http://bswm.da.gov.ph/program/salinity/>
- DA-PhilRice. (2011). *Rice Technology Bulletin Series*. <https://www.pinoyrice.com/wp->

- content/uploads/management-of-salt-affected-soils-for-rice-production-2.pdf
- Dobermann, A., & Fairhurst, T. (2000). *Nutrient Disorders & Nutrient Management Rice Rice ecosystems Nutrient management Nutrient deficiencies Mineral toxicities Tools and information* (1st ed.). PPI. [http://books.irri.org/9810427425\\_content.pdf](http://books.irri.org/9810427425_content.pdf)
- Douaoui, A. E. K., Nicolas, H., & Walter, C. (2006). Detecting salinity hazards within a semiarid context by means of combining soil and remote-sensing data. *Geoderma*, 134(1–2), 217–230. <https://doi.org/10.1016/j.geoderma.2005.10.009>
- Dutkiewicz, A., Lewis, M., & Ostendorf, B. (2009). Evaluation and comparison of hyperspectral imagery for mapping surface symptoms of dryland salinity. *International Journal of Remote Sensing*, 30(3), 693–719. <https://doi.org/10.1080/01431160802392612>
- El-Haddad, A., & Garcia, L. A. (2006). *Detecting Soil Salinity Levels in Agricultural Lands Using Satellite Imagery*.
- Eldeiry, A. A., & Garcia, L. A. (2008). Detecting Soil Salinity in Alfalfa Fields using Spatial Modeling and Remote Sensing. *Soil Science Society of America Journal*, 72(1), 201–211. <https://doi.org/10.2136/sssaj2007.0013>
- ESA. (n.d.). *Sentinel-2 - Missions - Sentinel Online*. Retrieved October 7, 2020, from <https://sentinel.esa.int/web/sentinel/missions/sentinel-2>
- FAO. (1976). 3. *SALINE SOILS AND THEIR MANAGEMENT*. Soils Bulletin. <http://www.fao.org/3/x5871e/x5871e04.htm>
- FAO. (1986). *CHAPTER 7 - SALTY SOILS*. <http://www.fao.org/3/r4082e/r4082e08.htm>
- FAO. (2019). *Call for Proposals | Soil Salinity Mitigation and Adaptation Projects in the Eurasian region | Global Soil Partnership | Food and Agriculture Organization of the United Nations*. FAO of the United Nations. <http://www.fao.org/global-soil-partnership/resources/highlights/detail/en/c/1208623/>
- FAO and ITPS. (2015). *Status of the World's Soil Resources (SWSR) – Main Report*. <http://www.fao.org/3/a-i5199e.pdf>
- Fenix, M. (2020, September 10). The vanishing traditions of salt making | Inquirer Lifestyle. *Philippine Inquirer*. <https://lifestyle.inquirer.net/370488/the-vanishing-traditions-of-salt-making/>
- Fernández-Buces, N., Siebe, C., Cram, S., & Palacio, J. L. (2006). Mapping soil salinity using a combined spectral response index for bare soil and vegetation: A case study in the former lake Texcoco, Mexico. *Journal of Arid Environments*, 65(4), 644–667. <https://doi.org/10.1016/j.jaridenv.2005.08.005>
- Fischer, G., Nachtergaele, F., Prieler, S., van Velthuisen, H. T. ., Verelst, L., & Wiberg, D. (2018). *Soil Qualities for Crop Production*. FAO. <http://www.fao.org/soils-portal/data-hub/soil-maps-and-databases/harmonized-world-soil-database-v12/en/>
- Flores, H. (2019, March 6). *Drought due to El Niño to hit more provinces | Philstar.com*. Philippine Star. <https://www.philstar.com/headlines/2019/03/06/1899107/drought-due-el-nio-hit-more-provinces>
- Gao, B. (1996). NDWI--a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58, 257–266.
- Genuer, R., Poggi, J. M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters*, 31(14), 2225–2236. <https://doi.org/10.1016/j.patrec.2010.03.014>
- Glick, B. R., Cheng, Z., Czarny, J., & Duan, J. (2007). Promotion of plant growth by ACC deaminase-producing soil bacteria. In *New Perspectives and Approaches in Plant Growth-Promoting Rhizobacteria Research* (pp. 329–339). Springer Netherlands. [https://doi.org/10.1007/978-1-4020-6776-1\\_8](https://doi.org/10.1007/978-1-4020-6776-1_8)
- Goosens, R., Ghabour, T. K., Ongena, T., & Gad, A. (1994). Waterlogging and soil salinity in the newly reclaimed areas of the western Nile Delta of Egypt. *Environmental Change in Drylands: Biogeographical and Geomorphological Perspectives*, 365–377. <https://agris.fao.org/agris-search/search.do?recordID=GB9509373>
- Grömping, U. (2009). Variable importance assessment in regression: Linear regression versus random forest. *American Statistician*, 63(4), 308–319. <https://doi.org/10.1198/tast.2009.08199>
- Habibi, V., Ahmadi, H., Jafari, M., & Moeini, A. (2021). Quantitative assessment of soil salinity using remote sensing data based on the artificial neural network, case study: Sharif Abad Plain, Central Iran. *Modeling Earth Systems and Environment*, 7(2), 1373–1383. <https://doi.org/10.1007/s40808-020-01015-1>
- Haefele, S. M., Nelson, A., & Hijmans, R. J. (2014). Soil quality and constraints in global rice production. *Geoderma*, 235–236, 250–259. <https://doi.org/10.1016/j.geoderma.2014.07.019>
- Hanson, B. (2011). *Agriculture and Natural Resources Making a Difference for California Drip Irrigation Salinity Management for Row Crops*. <http://anrcatalog.ucdavis.edu>

- Healthy soils are essential to achieve Zero Hunger, peace and prosperity - World | ReliefWeb.* (2018, August 13). Relief Web. <https://reliefweb.int/report/world/healthy-soils-are-essential-achieve-zero-hunger-peace-and-prosperity>
- Hick, P. T., & Russell, W. G. R. (1990). Some spectral considerations for remote sensing of soil salinity. *Australian Journal of Soil Research*, 28(3), 417–431. <https://doi.org/10.1071/SR9900417>
- Hiphen. (2021). *What is a Vegetation Index?* <https://www.hiphen-plant.com/vegetation-index/3582/>
- Hoa, P., Giang, N., Binh, N., Hai, L., Pham, T.-D., Hasanlou, M., & Tien Bui, D. (2019). Soil Salinity Mapping Using SAR Sentinel-1 Data and Advanced Machine Learning Algorithms: A Case Study at Ben Tre Province of the Mekong River Delta (Vietnam). *Remote Sensing*, 11(2), 128. <https://doi.org/10.3390/rs11020128>
- Hu, J., Peng, J., Zhou, Y., Xu, D., Zhao, R., Jiang, Q., Fu, T., Wang, F., & Shi, Z. (2019). Quantitative Estimation of Soil Salinity Using UAV-Borne Hyperspectral and Satellite Multispectral Images. *Remote Sensing*, 11(7), 736. <https://doi.org/10.3390/rs11070736>
- Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25(3), 295–309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)
- Humboldt State University. (2014). *Vegetation indices*. [http://gsp.humboldt.edu/OLM/Courses/GSP\\_216\\_Online/lesson5-1/NDVI.html](http://gsp.humboldt.edu/OLM/Courses/GSP_216_Online/lesson5-1/NDVI.html)
- Ilocosur.gov.ph. (n.d.). *Climate*. Retrieved June 6, 2021, from <https://www.ilocosur.gov.ph/index.php/gen-info/geography/climate>
- Ivushkin, K., Bartholomeus, H., Bregt, A. K., Pulatov, A., Kempen, B., & de Sousa, L. (2019). Global mapping of soil salinity change. *Remote Sensing of Environment*, 231, 111260. <https://doi.org/10.1016/j.rse.2019.111260>
- Khan, M., Rastoskuev, V., Shalina, E., & Sato, Y. (2001). Mapping Salt-Affected Soils Using Remote Sensing Indicators—A Simple Approach with the Use of GIS IDRISI. *Proceedings of the 22th Asian Conference on Remote Sensing, Singapore*, 5–9, 5.
- Khan, M., & Sato, Y. (2001). Monitoring hydro-salinity status and its impact in irrigated semi-arid areas using IRS-1B LISS-II data. *Asian J. Geoinf*, 1(3), 63–73.
- Khan, N., Rastoskuev, V., Sato, Y., & Shiozawa, S. (2005). Assessment of hydrosaline land degradation by using a simple approach of remote sensing indicators. *Agricultural Water Management*, 77(1–3), 96–109. <https://doi.org/10.1016/j.agwat.2004.09.038>
- Kupidura, P. (2019). The comparison of different methods of texture analysis for their efficacy for land use classification in satellite imagery. *Remote Sensing*, 11(10), 1233. <https://doi.org/10.3390/rs11101233>
- Liu, H., & Huete, A. (1995). A feedback based modification of the NDV I to minimize canopy background and atmospheric noise. *IEEE Transactions on Geoscience and Remote Sensing*, 33, 457–465.
- Lugassi, R., Goldshleger, N., & Chudnovsky, A. (2017). Studying Vegetation Salinity: From the Field View to a Satellite-Based Perspective. *Remote Sensing*, 9(2), 122. <https://doi.org/10.3390/rs9020122>
- Mariano, J., Romero, I., & Tingzon, J. (1954). *Soil Survey of Ilocos Sur Province, Philippines*.
- Metternicht, G. I., & Zinck, J. A. (2003). Remote sensing of soil salinity: Potentials and constraints. In *Remote Sensing of Environment* (Vol. 85, Issue 1, pp. 1–20). Elsevier Inc. [https://doi.org/10.1016/S0034-4257\(02\)00188-8](https://doi.org/10.1016/S0034-4257(02)00188-8)
- Mokarram, M., Hojjati, M., Roshan, G., & Negahban, S. (2015). Modeling the behavior of Vegetation Indices in the salt dome of Korsia in North-East of Darab, Fars, Iran. *Modeling Earth Systems and Environment*, 1(3), 1–9. <https://doi.org/10.1007/s40808-015-0029-y>
- Mougenot, B., & Pouget, M. (1993). Remote sensing of salt affected soils. *Remote Sensing Reviews*. [https://www.researchgate.net/publication/282168150\\_Remote\\_sensing\\_of\\_salt\\_affected\\_soils](https://www.researchgate.net/publication/282168150_Remote_sensing_of_salt_affected_soils)
- Mwiti, D. (2021, May 25). *Random Forest Regression: When Does It Fail and Why?* - *neptune.ai*. <https://neptune.ai/blog/random-forest-regression-when-does-it-fail-and-why>
- National Statistics Office. (2012). *2010 CENSUS POPULATION BY PROVINCE CITY/MUNICIPALITY BARANGAY OF POPULATION AND HOUSING REGION National Statistics Office ILOCOS REGION*.
- Paliwal, A., Laborte, A., Nelson, A., & Singh, R. K. (2019). Salinity stress detection in rice crops using time series MODIS VI data. *International Journal of Remote Sensing*, 40(21), 8186–8202. <https://doi.org/10.1080/01431161.2018.1513667>
- Panagos, P., Van Liedekerke, M., Jones, A., & Montanarella, L. (2012). European Soil Data Centre: Response to European policy support and public data requirements. *Land Use Policy*, 29(2), 329–

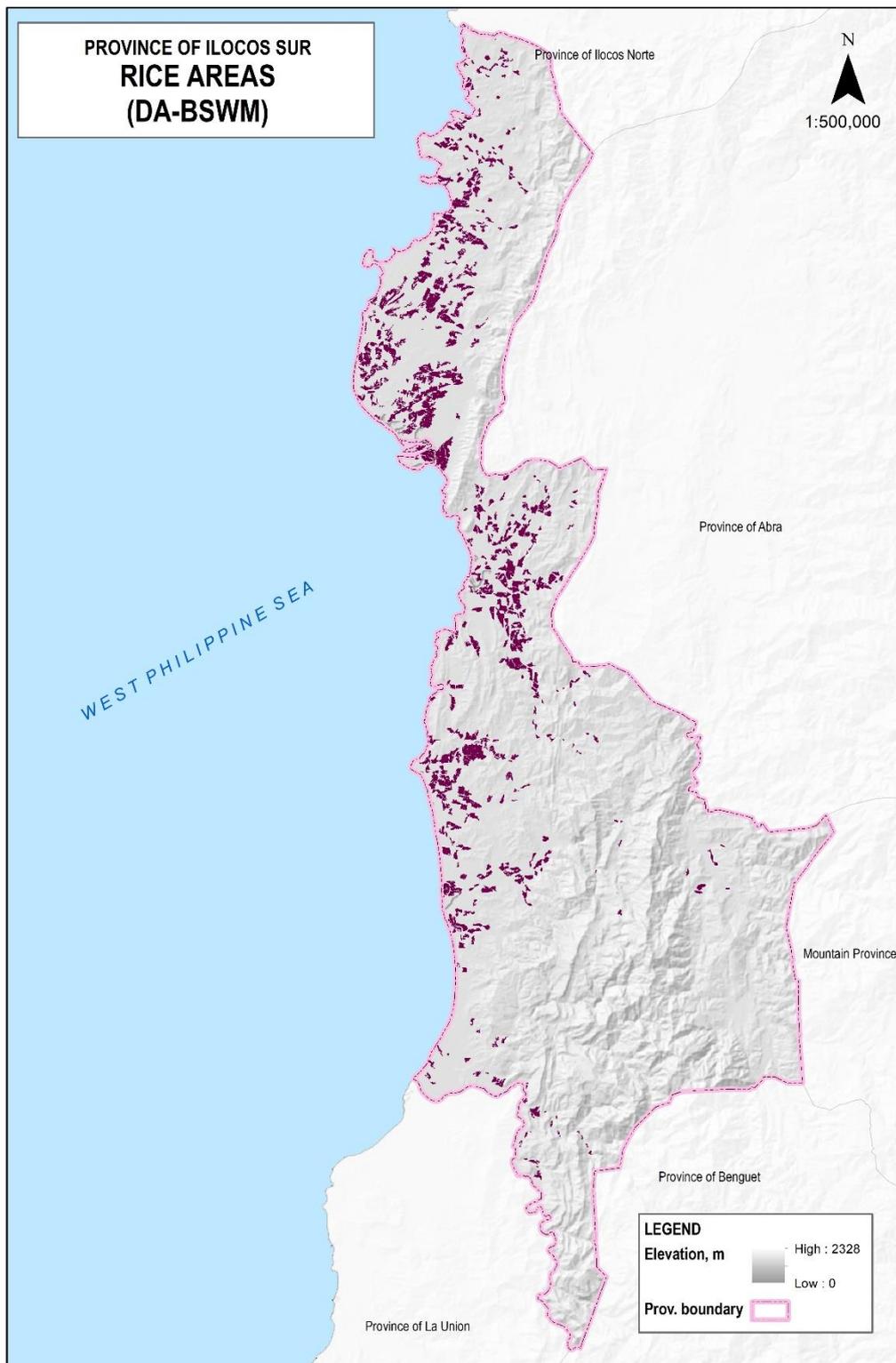
338. <https://doi.org/10.1016/j.landusepol.2011.07.003>
- Paul, J., Verleyesen, M., & Dupont, P. (2013). Identification of Statistically Significant Features from Random Forests. *Undefined*.
- Pearson, R. L., Miller, L. D., Pearson, R. L., & Miller, L. D. (1972). Remote Mapping of Standing Crop Biomass for Estimation of the Productivity of the Shortgrass Prairie. *Eighth International Symposium on Remote Sensing of Environment, University of Michigan, Ann Arbor, Mich*, 1977. <https://www.scopus.com/record/display.uri?eid=2-s2.0-84858446148&origin=inward&txGid=afe248519f5ea36c2530f21ec76efbd0>
- Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., & Rossiter, D. (2021). SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *SOIL*, 7(1), 217–240. <https://doi.org/10.5194/soil-7-217-2021>
- Rappler. (2018, October 30). *Typhoon Rosita makes landfall in Isabela*. <https://www.rappler.com/nation/weather/typhoon-rosita-pagasa-forecast-october-30-2018-5am>
- Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1973). Monitoring Vegetation Systems in the Great Plains with ERTS. *3rd ERTS Symposium*, 10(14), 309–317.
- Scudiero, E., Skaggs, T. H., & Corwin, D. L. (2014). Regional scale soil salinity evaluation using Landsat 7, Western San Joaquin Valley, California, USA. *Geoderma Regional*, 2–3(C), 82–90. <https://doi.org/10.1016/j.geodrs.2014.10.004>
- Shahid, S. A., Zaman, M., & Heng, L. (2018). Introduction to Soil Salinity, Sodicity and Diagnostics Techniques. In *Guideline for Salinity Assessment, Mitigation and Adaptation Using Nuclear and Related Techniques* (pp. 1–42). Springer International Publishing. [https://doi.org/10.1007/978-3-319-96190-3\\_1](https://doi.org/10.1007/978-3-319-96190-3_1)
- Shrivastava, P., & Kumar, R. (2015). Soil salinity: A serious environmental issue and plant growth promoting bacteria as one of the tools for its alleviation. In *Saudi Journal of Biological Sciences* (Vol. 22, Issue 2, pp. 123–131). Elsevier. <https://doi.org/10.1016/j.sjbs.2014.12.001>
- Singh, R. P., Setia, R., Verma, V. K., Arora, S., Kumar, P., & Pateriya, B. (2017). Satellite remote sensing of salt-affected soils: Potential and limitations. *Journal of Soil and Water Conservation*, 16(2), 97. <https://doi.org/10.5958/2455-7145.2017.00015.7>
- Szabolcs, I. (1998). *Methods for Assessment of Soil Degradation* (R. Lal, W. H. Blum, C. Valentine, & B. A. Stewart (Eds.)). Library of Congress Cataloging-In-Publication Data.
- Taghadosi, M. M., Hasanlou, M., & Eftekhari, K. (2019a). *Retrieval of soil salinity from Sentinel-2 multispectral imagery*. <https://doi.org/10.1080/22797254.2019.1571870>
- Taghadosi, M. M., Hasanlou, M., & Eftekhari, K. (2019b). Soil salinity mapping using dual-polarized SAR Sentinel-1 imagery. *International Journal of Remote Sensing*, 40(1), 237–252. <https://doi.org/10.1080/01431161.2018.1512767>
- Toldo, M. (2019, March 15). *Filipinos show varying levels of concern about drought - Philippines | ReliefWeb*. News and Press Release. <https://reliefweb.int/report/philippines/filipinos-show-varying-levels-concern-about-drought>
- USDA Natural Resources Conservation Service. (1998). *Soil Quality Information Sheet Soil Quality Resource Concerns: Salinization*. <http://soils.usda.gov>
- USGS. (n.d.). *Landsat Soil Adjusted Vegetation Index*. Retrieved June 15, 2021, from <https://www.usgs.gov/core-science-systems/nli/landsat/landsat-soil-adjusted-vegetation-index>
- Valley, J., Scudiero, E., Skaggs, T. H., & Corwin, D. L. (2014). Regional scale soil salinity evaluation using Landsat 7, western San Joaquin Valley, California, USA. *Geoderma Regional*, 2(3). <https://doi.org/10.1016/j.geodrs.2014.10.004>
- Verma, K. S., Saxena, R. K., Barthwal, A. K., & Deshmukh, S. N. (1994). Remote sensing technique for mapping salt affected soils. *International Journal of Remote Sensing*, 15(9), 1901–1914. <https://doi.org/10.1080/01431169408954215>
- Wu, W., Zucca, C., Muhaimed, A. S., Al-Shafie, W. M., Fadhil Al-Quraishi, A. M., Nangia, V., Zhu, M., & Liu, G. (2018). Soil salinity prediction and mapping by machine learning regression in Central Mesopotamia, Iraq. *Land Degradation and Development*, 29(11), 4005–4014. <https://doi.org/10.1002/ldr.3148>
- Yahiaoui, I., Douaoui, A., Zhang, Q., & Ziane, A. (2015). Soil salinity prediction in the Lower Cheliff plain (Algeria) based on remote sensing and topographic feature analysis. *Journal of Arid Land*, 7(6), 794–805. <https://doi.org/10.1007/s40333-015-0053-9>
- Zamann, M., Shahid, S. A., & Heng, L. (2018). *Guideline for Salinity Assessment, Mitigation and Adaptation Using Nuclear and Related Techniques*.
- Zhang, T. T., Qi, J. G., Gao, Y., Ouyang, Z. T., Zeng, S. L., & Zhao, B. (2015). Detecting soil salinity

with MODIS time series VI data. *Ecological Indicators*, 52, 480–489.  
<https://doi.org/10.1016/j.ecolind.2015.01.004>

Zhang, T. T., Zeng, S. L., Gao, Y., Ouyang, Z. T., Li, B., Fang, C. M., & Zhao, B. (2011). Using hyperspectral vegetation indices as a proxy to monitor soil salinity. *Ecological Indicators*, 11(6), 1552–1562. <https://doi.org/10.1016/j.ecolind.2011.03.025>

# APPENDICES

## Appendix A. Map of Rice Areas



## Appendix B. Soils of the Area

Ilocos Sur has three distinct topographic units: the coastal plain, intermediate uplands, and mountains. According to the reconnaissance soil survey of the province (BSWM, 1954), the coastal plain part of the study area comprises four main soil orders in terms of the USDA soil taxonomic system. These are Fluvisols, Vertisols, Arenosol and Regosol (BSWM, 2019). In particular, the coastal plain soils are represented by the Bantog, San Manuel, and Umingan series and two miscellaneous land types: beach sand and river wash. All these soils manifest slight profile development: they are usually insufficient to moderately well-drained and medium to heavy textures, ranging from sandy loam to light clay loam (BSWM, 1954). For this research, only the soils of the coastal plain are discussed.

### Fluvisols

Fluvisols are young soils that have the “fluvic soil properties.” For all practical purposes, this means that the soils receive fresh sediment during regular floods (unless empoldered) and still show stratification and irregular organic matter profile (ISRIC). The soil series of San Manuel and Umingan belong to this soil order.

#### A. San Manuel sandy clay loam

This series comprises the most extensive and of the best soils of the plain. Generally, these soils are well-drained as they usually occur along rivers. They are developed from alluvial material washed down from the higher areas' underlain by igneous and sedimentary rocks. This type occurs in widely scattered areas in the province. Still, the largest contiguous area occurs in the plain from Vigan northward up to midway between Lapog and Cabugao. It covers an area of 23,275.65 hectares or 9.53 percent of the total soils of the province. This soil occupies nearly level to level areas with the elevation ranging from 15 to 30 meters above sea level. Drainage, both external and internal, is fair to good despite the nearly level topography. A few low areas, especially those devoted to lowland rice, have slow external drainage (BSWM, 1954).

#### B. Umingan sandy loam

This type occurs in the small alluvial plain below the town of Cervantes in the southeastern part of the province and along the highway from Santa northward up to Banaoang near the Abra river gap. It covers a total area of 1,356.00 hectares or 0.55 percent of the total soils mapped on the province. Due to the looseness of the surface and subsurface soil, this type is slightly erodible in the sloping areas. The surface soil is much thinner than the normal depth in some places, and the gravelly layer gets nearer the surface (BSWM, 1954).

### Vertisols

Vertisols are churning heavy clay soils with a high proportion of swelling clays. These are found in depressions, level to undulating areas, mainly in tropical areas. They become very hard in the dry season and are sticky during the wet season (ISRIC). The soil series of Bantog fall into this type of soil order.

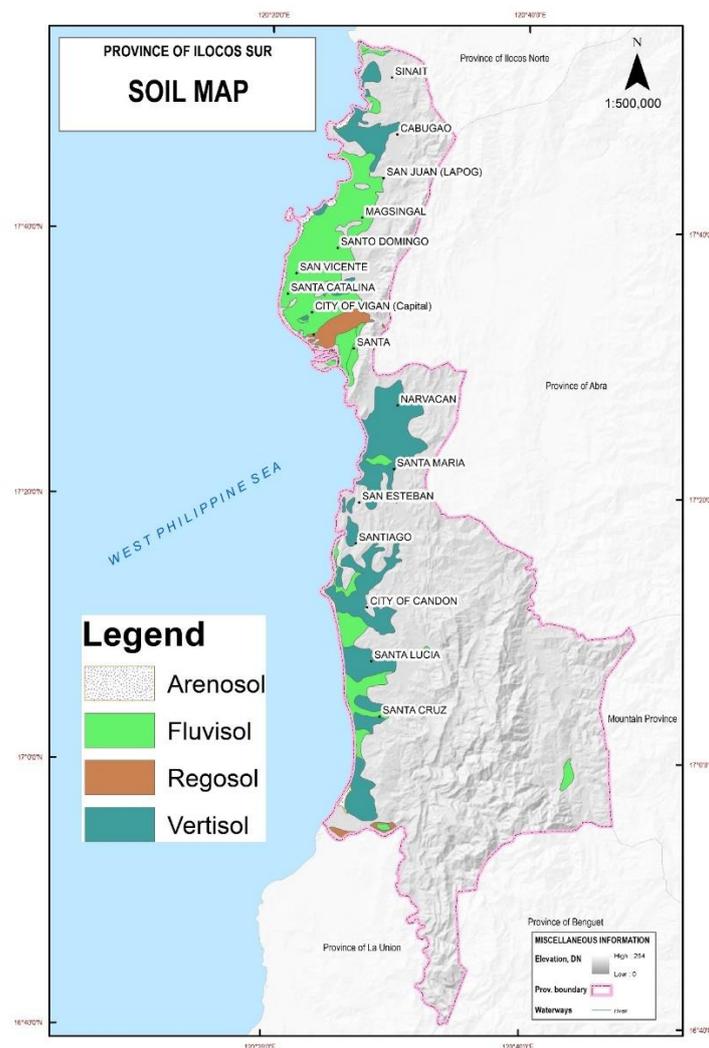
The soil series of Bantog constitutes the second most extensive soil unit. During the dry season, this soil cracks forming massive blocks with irregular cleavage. It is widely scattered in several parts of the plain, but the large areas of the type are in Narvacan, Cabugao, Candon, Sta. Maria, Sta. Lucia, Banayoyo and Tagudin. It covers 24 480.04 hectares or 10.02 percent of the mapped soils in the province. Generally, this soil occupies lower elevations than the San Manuel silt loam. Of all the soils of Ilocos Sur, this soil series has the highest pH, going into the alkaline side. This makes the maximum expected yield of rice impossible if left unmanaged (BSWM, 1954).

### Arenosols

This soil type consists mainly of sand mixed with some hummus or clay and is commonly found in arid and tropical regions (ISRIC). In the province, this is the beach sand soils. This occurs in the different locations along the coast from north to south of the province, with a mapped area of 1,484.31 hectares or 0.61 percent. There is little importance of agriculture production in this type of soil (BSWM, 1954).

### Regosols

Regosols are the type of soil that could not be accommodated in any other Reference Soil Group. This occurs in eroding lands of arid and semi-arid areas and in mountain regions (ISRIC). The local name for this is the river wash. The land under this type consists of stony, gravelly, and sandy material that is generally bare and useless for plants. This has a mapped area of 2,566.24 hectares or 1.05 percent (BSWM, 1954).



Soil map of Ilocos Sur Province (lowlands only) based on World Reference Based for Soil Resources

## Appendix C. Preprocessing of Sentinel-1

The whole preprocessing of Sentinel-1 data was done in the NRS Dunne server using the following steps in SNAP v.8:

B.1 TOPS Split Used to select the interferometric swath (IW) in the image. IW2 was selected for Sentinel-1 B, and IW3 was selected for Sentinel-1 A products. VV and VH polarizations were also selected in the processing parameters.

B.2 Apply Orbit File This improved the geolocation accuracy of the retrieved image.

B.3 Calibration to Beta0 Normalized the backscatter signal to beta0 band and derived reliable radar backscattering coefficients.

B.4 TOPSAR Deburst Removed the seamlines in the image.

B.5 Radiometric Terrain Flattening This process performed the radiometric correction. “SRTM 3Sec” was selected as the digital elevation model, bilinear interpolation as the DEM resampling method, and the other parameters were set to default. This method produced the Gamma0 band.

B.6 Speckle filtering This reduced the speckle amount resulting from the resampling method of radiometric terrain flattening. Refined Lee was used as the filter. This filter averaged the image while the edges are preserved.

B.7 Range doppler Terrain Correction This geocoded the image by correcting SAR geometric distortion using DEM. This method produces a projected image. “SRTM 3Sec” was also the DEM used in the image, and the other parameters were set at default.

B.8 Linear to dB: After getting the final gamma0 bands, they were converted from linear to dB and saved.

B.9 Exporting of Gamma0 bands The final gamma bands in dB units were exported in GeoTIFF/BigTIFF format. This resulted in one raster file having two bands—Gamma0 in VH for the first band and GammaVV in VV for the second band.

B.10 Generating Gray Level Co-occurrence Matrix (GLCM) Features For processing parameters, only the Gamma0 in dB bands were used with a 5x5 window size, quantification into 32 bins, and the other parameters were set as default. Furthermore, nine features were selected. This results in a product with 18 bands (9 features in VV and 9 for VH polarization) with 13.89m spatial resolution.

The selected features were contrast, dissimilarity, homogeneity, energy, max, entropy, GLCM mean, GLCM variance, and GLCM correlation. The Angular Second Moment (ASM), which was on the list for generating the GLCM features on SNAP, was not included due to its similar definition to energy.

B.11 Exporting the GLCM features. This process was different from the earlier export of Gamma0 bands. When the GLCM features were created on the product, it will automatically remove the Gamma0 bands and replace them with the generated GLCM features. The GCLM features were also exported into GeoTIFF/BigTIFF file format resulting in one raster file having 18 bands: 9 bands for GLCM in VH polarization and nine bands for GLCM in VV polarization. Table 6 shows the bands of the output raster file.

## B.12. Processing in ArcMap

B.12.1 Setting No Data Values Each band of the raster file was extracted to process them individually. A raster calculator in ArcMap was used to set the No Data value (-9999).

B.12.2 Mosaicking The images were mosaicked using the cell statistics function in ArcMap to make the image pairs as one image.

B.12.3 Reprojection and resampling The mosaicked images were reprojected to WGS84 UTM Zone 51N. The image was resampled into 10 x 10-meter spatial resolution using the nearest neighbor resampling method.

B.12.4 Clipping to Study Area The administrative boundary of the Province of Ilocos Sur was used as a mask to extract the study area. The function is used as an extract by mask in ArcMap. This function is essential because the random forest classification reads and transforms the images into a data frame, so each image must have the same extent and cell size.

## Appendix D. Preprocessing of Sentinel-2

Preprocessing of Sentinel-2 MSI Level 1 data was done in SNAP using Sen2Cor. Sen2Cor is a processor for Sentinel-2 Level 2A product generation and formatting; it performs the atmospheric, terrain and cirrus correction of Top-Of- Atmosphere Level 1C input data. In addition, Sen2Cor creates Bottom-Of-Atmosphere, optionally terrain- and cirrus corrected reflectance images; additional, Aerosol Optical Thickness-, Water Vapor-, Scene Classification Maps and Quality Indicators for cloud and snow probabilities (ESA, n.d).

C.1 Sen2Cor processing Sen2Cor version 2.5.5 plugin was first installed in SNAP, and the environment settings for this plugin were set in the local directory. Then, the downloaded images were opened in SNAP. In the processing parameters, “all” was chosen in the resolution, and the remaining parameters are set to default. The result of this is the Level 2-A of the product.

C.2 Resampling Using the resampling tool in SNAP, the Level 2A products were resampled to 10m resolution.

C.3 Exporting into GeoTIFF and BigTIFF All the bands (B1 to B12) were exported into GeoTIFF/BigTIFF file format. Which results in a raster file with 12 bands.

C.4 Generating indices related to soil salinity Seventeen indices were explored in relating soil salinity to optical remote sensing. In ArcMap, bands 2, 3, 4, and 8 of Sentinel-2 Level 2A products were used to generate the indices using the raster calculator tool. These bands correspond to the Blue (B), Green (G), Red (R), and Near-Infrared (NIR) regions of the electromagnetic spectrum, as shown in Table 9. Raster calculator tool in ArcMap was used to generate the indices based on the following formulas shown in Table 10.

## Appendix E. Preprocessing of Landsat-8

The formula used to convert the digital number in Band 11 (Thermal Infrared 1) into radiance is

### D.1 Converting a digital number to radiance

The formula used to convert digital number into radiance is

$$L\lambda = M_L Q_{cal} + A_L$$

Where  $L\lambda$  is the top of atmosphere (TOA) spectral radiance

$M_L$  is the band-specific multiplicative rescaling factor from the metadata (radiance-multi-band)

$Q_{cal}$  is the quantized and calibrated standard product pixel values (DN)

$A_L$  is the band-specific additive factor from metadata (radiance-add-band)

Therefore, for  $L\lambda$  of the study area, the applied formula for Landsat-8 is

$$L\lambda = (0.0003342 * \text{Band10} + 0.1)$$

The same formula was applied in Band 11

### D.2 Converting radiance to brightness temperature

This step uses the thermal constant (K values) given in the metadata file. The formula used is

$$T = \frac{K_2}{\ln\left(\frac{K_1}{L\lambda} + 1\right)} - 273.15$$

Where  $T$  is the At-satellite brightness temperature in Kelvin

$L\lambda$  TOA spectral radiance result from step 2.a

$K_1$  and  $K_2$  are the band-specific conversions from the metadata

Therefore, for  $T$  of the study area, the applied formula for Landsat-8 is

$$T_{B10} = \left( \frac{1321.0789}{\ln\left(\frac{774.8853}{L\lambda_{B10}} + 1\right)} - 273.15 \right)$$

$$T_{B11} = \left( \frac{1201.1442}{\ln\left(\frac{480.8883}{L\lambda_{B11}} + 1\right)} - 273.15 \right)$$

### D.3 Calculating Mean Statistics

This step calculated the average temperature of both  $T_{B10}$  and  $T_{B11}$  using the Cell Statistics Tool and selected the mean statistics option.

### D.4 Calculating the NDVI

The NDVI was calculated using  $((\text{Band5} - \text{Band4}) / ((\text{Band5} + \text{Band4})))$

The maximum and minimum values of NDVI for both images are 1 and -1, respectively.

#### D.5 Calculating proportion of vegetation

This was calculated using the NDVI values by the formula

$$pv = \left( \frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \right)^2$$

#### D.6 Deriving Land Surface Emissivity (LSE)

The formula for LSE is

$$LSE = \text{float}(0.004 * pv + 0.986)$$

#### D.7 Calculating Land Surface Temperature

Using equation 8, the LST was calculated using step B.2 as a substitute for BT, that is, brightness temperature in Celsius and the output of B.6 to replace e or emissivity. Using a raster calculator, the LST was calculated using the formula

$$LST = \left( \frac{T_{B10}}{\left( 1 + B_{10} * \frac{T_{B10}}{14380} \right) * \ln LSE} \right)$$

#### D.8 Processing in ArcMap

After generating the LST raster data, the further processing of the image includes setting no data values, mosaicking, reprojection, resampling, and clipping to the study area. Refer to sub-sections C.1 to C.4 of section 3.4.1 for the methods used. Furthermore, the mean of two generated LSTs was calculated using the cell statistics tool.

## Appendix F. Protocol on Soil Sampling

Protocol on Soil Sampling

Soils Survey Division

Bureau of Soils and Water Management

Philippines

- I. Pre-field Survey Activities
  - Coordination with stakeholders (DA-RFOs, LGUs, and partner agencies)
  - Preparation of base maps (topographic maps of 1:50,000 scale)
  - Procurement and preparation of office and field supplies and materials
- II. Field Survey Activities
  1. Preparation of materials, tools, and equipment
    - Soil auger
    - Shovel
    - Bolo
    - Soil sampling bags
    - Pail
    - Sack (for mixing and quartering)
    - Notebook and pens
    - Camera
    - Handheld GPS (or mobile tablets with GPS applications)
    - Interview Forms
    - Field maps
  2. Courtesy call with LGUs or barangay officials – upon meeting the stakeholders, determine the following: phenological stages of standing rice and area without fertilizer application
  3. Rice Boundary verification/validation – verify/check the location in the GPS and field map, observe the surroundings and check the differences in land use, slope, topography, and soil drainage in the area and further segregate the boundary delineation in the working sheet
  4. Collecting soil samples
    - a. Homogenous Area – should there be homogeneity in land use and topography, collect one composite soil sample in 10 random sampling areas covering 50 hectares
    - b. Heterogenous Area – if there is heterogeneity in land use and topography, the minimum area may be less than 50 hectares. Collect one composite soil sample in 3-5 sampling points and trace the separation delineation
      - Using the soil auger - To break through the soil, push the soil auger downwards, rotating clockwise until the auger becomes loose, indicating that it is already full. Slowly pull the auger upward. Repeat the procedure until the desired depth (30 cm) is reached
      - Using the shovel - Remove the litter on top of the soil. Dig a “V-cut” hole at an approximately plow depth of 30 cm. From the newly exposed soil surface, get a slice of a soil sample, approximately 2 cm thick and 10 cm wide.
      - Gently remove the soil sample and place it in a pail. Repeat the procedure until the appropriate number of sampling points is reached.
      - Pour the collected soil samples taken from different sampling points into a sack, break the clods, and mix thoroughly.

- Conduct the quartering method and repeat this procedure for each sampling area until 1 kg of the composite soil sample is produced.
- Place the 1kg of the composite soil sample in the sampling bag, tightly seal the bag with a rubber band and label it with sampling number, date, location, land use (irrigated or not), and farmer's name (if available)

Note: the following should be avoided when taking soil samples:

1. Withstanding crop except when they are all ready for harvest
2. With weeds, shrubs, and animal manure
3. Near or beside a garbage pile, drainage system, irrigation canal, and road network

5. Management of soil samples

- Upon return from the field, prepare a master list of all samples collected within the day and submit/endorse the samples to the soil testing supervisor
- Air-dry the soil samples and grind them. Sieve the ground soil samples using a 2mm mesh
- Divide the soil samples into two, one for submission for complete (routine analysis) and one for soil testing using Soil Test Kit and place them on their respective sampling bags
- Submit/endorse the soil samples to the Regional Soils Laboratory or soil testing supervisor

## Appendix G. Python Script for Random Forests regression

"""

Title: Random Forests Regression for Predicting Soil Salinity

Created on Tue Jan 19 07:44:34 2021

@author: SarahJoey

contact: s.j.salgado@student.utwente.nl

Welcome

This script is intended for the assessment of spatial variability of soil salinity in coastal agricultural areas in the Province of Ilocos Sur, Philippines using satellite data derived from Sentinel-1, Sentinel-2 and Landsat-8 and field data taken from the DA- BSWM.

"""

#Importing plugins

import os

import sys

from datetime import datetime

import matplotlib.pyplot as plt

import numpy as np

import pandas as pd

import random

from pylab import hist2d

from sklearn.model\_selection import train\_test\_split

from sklearn.metrics import mean\_squared\_error

from RF\_Functions import file\_path\_check, read\_tif, Run\_RF, writeTiff, cal\_ubrmse

from joblib import dump

from scipy.stats import pearsonr

# time (start running)

time\_0 = datetime.now()

start\_time = time\_0.strftime("%Y-%m-%d %H:%M:%S")

print('Program start running at: %s. ' % start\_time)

"""

Step 1. Specifying the working path.

"""

# Set the working path

print('-- Step 1. Specifying the working path: Started! --')

work\_path = r'D:\Data\Personal\s2245531\RANDOMFOREST\_21Apr'

user\_name = 'Sarah Joey Salgado'

os.chdir(work\_path)

satellite\_folder = 'sensor'

output\_folder = 'result\_train\_MSDA\_2'

if not os.path.exists(output\_folder):

    os.mkdir(output\_folder)

print('-- Step 1. Setting the work path: Completed! --\n')

"""

Step 2. Reading the excel data and transforming into pandas dataframe

```

'''
print('-- Step 2. Read the excel data: Started! --')
# read excel data
file_insitu = 'in_situ\\extracted_26May.xls'
df_insitu = pd.read_excel(file_insitu, sheet_name='Sheet3') # Transformed into pandas
print('--Step 2. Read the excel data: Completed --\n')

'''

Step 3. Train and test the excel data
'''

print('-- Step 3. Train and test the Random Forest Regression Model: Started! --')
df_insitu = df_insitu.dropna()
# set the labels
labels = np.array(df_insitu['EC'])
features = df_insitu.drop('EC', axis=1)

# remove the labels from the features
features = features.drop('ID', axis=1)

# dropping non-significant variables
feature_list = list(features.columns)
features = np.array(features)

# Randomly split the features and labels .
train_features, test_features, train_labels, test_labels = train_test_split(features, labels, test_size=0.25,
random_state=42)
# Run the random forest regression model
rf_model = Run_RF(train_features, test_features, train_labels, test_labels, feature_list, output_folder)
rf, predictions, importances, std, indices, rmse, r2, r = rf_model.train_model()
# Calculate the Unbiased Root Mean Square Error of the test model.
ubrmse, a_mean, b_mean = cal_ubrmse(test_labels, predictions)
print("Test of the model:")
print('rmse: %.2f. \n ubrmse: %.2f % (rmse, ubrmse)

# Save the trained RF model.
file_model = os.path.join(output_folder, 'RF_model_new_multi2.joblib')
dump(rf, file_model)
print('-- Step 3. Train and test the model: Completed!\n')

'''-- Step 4. Draw the figure and save the result'''
# Save the result
print('-- Step 4. Showing Result and draw figures: started! --')
file_txt = os.path.join(output_folder, 'result.txt')
f1 = open(file_txt, 'w')
f1.write("**** This is %s's running result of the RF model ****" % user_name)
f1.write('\nStart running at: %s\n' % start_time)
# sort the feature_list
feature_list_new = feature_list.copy()
for idx in np.arange(len(feature_list)):
    feature_list_new[idx] = feature_list[list(indices)[idx]]

```

```

# print the feature importances:
f1.write('--- Train the model, importance of features ---')
for count_k in np.arange(len(indices)):
    print('Feature: %s, Importance: %.2f % (feature_list_new[indices[count_k]],
importances[indices[count_k]])
    f1.write('\nFeature: %s, Importance: %.2f % (feature_list_new[indices[count_k]],
importances[indices[count_k]]))

f1.write('\n ---Test the model ---')
print('\nMetrics:'), f1.write('\nTest the Model:')
print("RMSE: %.2f [saturation degree]" % rmse), f1.write("\nRMSE: %.2f [saturation degree]" %
rmse)
print("r2: %.2f" % r2), f1.write("\nr2: %.2f" % r2)
print("Pearson correlation coefficient: %.2f" % r), f1.write("\nPearson correlation coefficient: %.2f" %
r)

# Show the Images of Importance of Features
plt.figure()
plt.title("Importance of features")
plt.bar(range(train_features.shape[1]), importances[indices],
        color='b', yerr=std[indices], align='center')
# set the x labels
plt.xticks(range(train_features.shape[1]), feature_list_new)
plt.xlim(-1, train_features.shape[1])

# Add the xlabel and ylabel
plt.xlabel('Name of features')
plt.ylabel('Relative importance [%]')
out_name = os.path.join(output_folder, 'Feature_Importances.png')
plt.savefig(out_name)
# plt.pause(5) # Show 5 seconds
plt.close()

#making a scatterplot
x = test_labels
y = predictions
colors = (0,0,0)
area=np.pi*3
plt.scatter(x,y,s=area,c=colors,alpha=0.5)
plt.title('Multi-sensor Data Association')
plt.xlabel('Field-measured EC')
plt.ylabel('Predicted soil salinity')
plt.savefig(os.path.join(output_folder, 'Scatterplot of Soil Salinity.png'))
plt.close()
print('-- Step 4. Draw the figure and save the result: Completed!\n')

'''
--Step 5. Read and load satellite data--
'''

```

```

# Set the file path
print('-- Step 5.1. Reading and loading satellite data: Started! --\n')

# Set the file path
file_enervh = 'sensor\\enervh.tif'
file_rvi = 'sensor\\rvi.tif'
file_savi = 'sensor\\savivi.tif'

# Set the file path for Soil Properties Data (Ancillary)
file_bk = 'prop\\bk.tif'
file_silt = 'prop\\silt.tif'
file_sand = 'prop\\sand.tif'
file_clay = 'prop\\clay.tif'
file_ph = 'prop\\ph1.tif'
file_mois = 'prop\\2019_mois.tif'

# Set the file path for geographical data
file_dem = 'prop\\dem.tif'
file_long = 'prop\\longitude_null1.tif'
file_lat = 'prop\\latitude_null1.tif'

# Set the file path for climate data
file_ppt = 'prop\\2019_ppt.tif'
file_tave = 'prop\\2019_tave.tif'

# Check the file path
file_path_check([file_enervh, file_rvi, file_savi,
                 file_bk, file_silt, file_sand, file_clay, file_ph, file_mois,
                 file_dem, file_long, file_lat,
                 file_ppt, file_tave])
print('-- Step 5.1. Setting the file path: Completed! --\n')

# Read the data
print('--Step 5.2. Read the data: Started!--')

# read the data [important features]
ds_enervh = read_tif(file_enervh)
ds_rvi = read_tif(file_rvi)
ds_savi = read_tif(file_savi)

# read the data [Bulk Density, DEM]
ds_bk, ds_dem = read_tif(file_bk), read_tif(file_dem)
ds_bk = (ds_bk[0] * 0.001, ds_bk[1], ds_bk[2], ds_bk[3], ds_bk[4], ds_bk[5])
# The scale of Bulk Density is 0.001 ; construct a new tuple

# read the data [Soil Texture: Sand, Silt, Clay]
ds_sand = read_tif(file_sand)
ds_silt = read_tif(file_silt)
ds_clay = read_tif(file_clay)

```

```

# read the data [pH and moisture]
ds_ph = read_tif(file_ph)
ds_mois = read_tif(file_mois)

# read the data [geographical data]
ds_long = read_tif(file_long)
ds_lat = read_tif(file_lat)

# read the data [climate]
ds_ppt = read_tif(file_ppt)
ds_tave = read_tif(file_tave)
print('--Step 5.2. Read the data: Completed!--')

'''
Step 6. Reformat the Array into pandas data frame
'''

# list of the data and name for convenience.
print('-- Step 6. Reformat the data: started! --')
series_data = [ds_enervh, ds_rvi, ds_savi]
series_data_list = ['enervh', 'rvi', 'savi']

property_data = [ds_bk, ds_dem, ds_sand, ds_silt, ds_clay, ds_ph, ds_mois, ds_long, ds_lat, ds_ppt,
ds_tave]
property_data_list = ['bk', 'dem', 'sand', 'silt', 'clay', 'ph', 'mois', 'long', 'lat', 'ppt', 'tave']

# Set the column list of the data frame
columns_list = ['enervh', 'rvi', 'savi',
                'bk', 'dem', 'sand', 'silt', 'clay', 'ph', 'mois', 'lat', 'long', 'ppt', 'tave']

# Initialize the data frame
df_all = pd.DataFrame(columns=columns_list)
df_all = df_all.dropna()

print('Allocating DataFrame of', len(columns_list), 'columns')
# Obtain the time-series data
col = 1
for idx, column in enumerate(series_data_list):
    item = series_data[idx]
    data = item[0]
    pixels_x = item[1]
    pixels_y = item[2]
    nr_bands = item[3]
    print('[', col, '/', len(columns_list), '] Filling column', column, 'with timeseries of', pixels_x, '*', pixels_y,
'pixels *', nr_bands, 'bands =', pixels_x * pixels_y * nr_bands, 'samples')
    df_all[column] = data.transpose().flatten()
    col = col + 1

# Obtain the 'stable' data
for idx, column in enumerate(property_data_list):
    item = property_data[idx]

```

```

data = item[0]
pixels_x = item[1]
pixels_y = item[2]
print(['col,/',len(columns_list),] Filling column', column, 'with fixed data of', pixels_x, '*', pixels_y,
'pixels expanded to', nr_bands, 'bands =', pixels_x * pixels_y * nr_bands, 'samples')
df_all[column] = data.transpose().flatten().repeat(nr_bands)
col = col + 1
print('-- Step 6. Reformat the data: Completed!--\n')

''' Predicting soil salinity'''
print('-- Step 6. soil Salinity Prediction: started! --')
lines = ds_rvi[0][0, :].shape[0] #most important feature
columns = ds_rvi[0][0, :].shape[1]

# format the features into a line

enervh_a_line = ds_enervh[0].reshape(1, lines * columns)
enervh_a_line[ds_enervh == -3.40282346639e+038] = 0
rvi_a_line = ds_rvi[0].reshape(1, lines * columns)
rvi_a_line[ds_rvi == -3.40282346639e+038] = 0
savi_a_line = ds_savi[0].reshape(1, lines * columns)
savi_a_line[ds_savi == -3.40282306074e+038] = 0

bk_a_line = ds_bk[0].reshape(1, lines * columns)
bk_a_line[ds_bk == 32767] = 0
clay_a_line = ds_clay[0].reshape(1, lines * columns)
clay_a_line[ds_clay == 32767] = 0
dem_a_line = ds_dem[0].reshape(1, lines * columns)
dem_a_line[ds_dem == -3.40282306074e+38] = 0
ph_a_line = ds_ph[0].reshape(1, lines * columns)
ph_a_line[ds_ph == 32767] = 0
sand_a_line = ds_sand[0].reshape(1, lines * columns)
sand_a_line[ds_sand == 32767] = 0
silt_a_line = ds_silt[0].reshape(1, lines * columns)
silt_a_line[ds_silt == 32767] = 0

lat_a_line = ds_lat[0].reshape(1, lines * columns)
lat_a_line[ds_lat == -3.40282306074e+038] = 0
long_a_line = ds_long[0].reshape(1, lines * columns)
long_a_line[ds_long == -3.40282306074e+038] = 0

mois_a_line = ds_mois[0].reshape(1, lines * columns)
mois_a_line[ds_mois == -3.40282306074e+038] = 0
ppt_a_line = ds_ppt[0].reshape(1, lines * columns)
ppt_a_line[ds_ppt == -3.40282306074e+038] = 0
tave_a_line = ds_tave[0].reshape(1, lines * columns)
tave_a_line[ds_tave == -3.40282306074e+038] = 0

# concatenate the lines
ms1_feature_arr = rvi_a_line.copy()

```

```

ms1_feature_arr = np.concatenate((ms1_feature_arr, enervh_a_line), axis=0)
ms1_feature_arr = np.concatenate((ms1_feature_arr, savi_a_line), axis=0)

ms1_feature_arr = np.concatenate((ms1_feature_arr, bk_a_line), axis=0)
ms1_feature_arr = np.concatenate((ms1_feature_arr, clay_a_line), axis=0)
ms1_feature_arr = np.concatenate((ms1_feature_arr, dem_a_line), axis=0)
ms1_feature_arr = np.concatenate((ms1_feature_arr, mois_a_line), axis=0)
ms1_feature_arr = np.concatenate((ms1_feature_arr, ph_a_line), axis=0)
ms1_feature_arr = np.concatenate((ms1_feature_arr, sand_a_line), axis=0)
ms1_feature_arr = np.concatenate((ms1_feature_arr, silt_a_line), axis=0)

ms1_feature_arr = np.concatenate((ms1_feature_arr, lat_a_line), axis=0)
ms1_feature_arr = np.concatenate((ms1_feature_arr, long_a_line), axis=0)

ms1_feature_arr = np.concatenate((ms1_feature_arr, ppt_a_line), axis=0)
ms1_feature_arr = np.concatenate((ms1_feature_arr, tave_a_line), axis=0)

# transpose the array
ms1_feature_arr = ms1_feature_arr.transpose()
test_ms1_ss = rf.predict(ms1_feature_arr) #prediction of salinity
test_ms1_ss = np.nan_to_num(test_ms1_ss)
test_ms1_ss_rs = test_ms1_ss.reshape(lines, columns)

# get the prediction result and create soil salinity map
proj = ds_rvi[5]
proj = proj.__call__()
file_soilsalinity = os.path.join(output_folder, 'multisensor soil_salinity.tif')
writeTiff(test_ms1_ss_rs, columns, lines, 1, ds_enervh[4], proj, file_soilsalinity)
print('-- Step 6. Soil Salinity Prediction: completed! --')

# time (end running)
time_1 = datetime.now()
end_time = time_1.strftime("%Y-%m-%d %H:%M:%S")
print('program end running at: %s.' % end_time), f1.write('\nEnd running at: %s.' % end_time)
time_consuming = (time_1 - time_0).seconds
print('It takes %d seconds to run this scripts' % int(time_consuming))
f1.write('\nIt takes %d seconds to run this scripts' % int(time_consuming))
f1.write("\n**** This is %s's running result of the RF model ****" % user_name)
f1.close()

```

### Appendix H. Predicted Soil Salinity in Different Municipalities Using OMSP Model

Municipality	Multi-sensor Data Fusion				Total Area, ha
	Low	Moderately Low	Moderately High	High	
Alilem	118.15	-	-	-	118.15
Banayoyo	132.84	-	-	-	132.84
Bantay	222.04	0.23	-	-	222.27
Burgos	488.80	-	-	-	488.80
Cabugao	364.71	23.69	-	-	388.40
Caoayan	132.64	7.32	0.50	-	140.46
City of Candon	846.70	15.88	0.30	-	862.88
City of Vigan	235.07	10.72	-	-	245.79
Galimuyod	108.98	-	-	-	108.98
Gregorio del Pilar	6.70	-	-	-	6.70
Lidlidda	85.31	-	-	-	85.31
Magsingal	531.04	-	-	-	531.04
Nagbukel	94.11	-	-	-	94.11
Narvacan	641.71	0.79	-	-	642.50
Quirino	75.27	-	-	-	75.27
Salcedo	91.32	-	-	-	91.32
San Emilio	47.51	-	-	-	47.51
San Esteban	102.09	0.10	-	-	102.19
San Ildefonso	64.72	-	-	-	64.72
San Juan	316.74	4.70	-	-	321.44
San Vicente	155.11	4.07	-	-	159.18
Santa	300.23	39.18	9.61	-	349.02
Santa Catalina	138.60	7.27	-	-	145.87
Santa Cruz	368.00	1.44	-	-	369.44
Santa Lucia	251.97	1.80	-	-	253.77
Santa Maria	661.41	19.37	-	-	680.78
Santiago	170.36	1.49	-	-	171.85
Santo Domingo	586.16	3.34	-	-	589.50
Sigay	14.77	-	-	-	14.77
Sinait	151.82	14.71	-	-	166.53
Sugpon	47.91	-	-	-	47.91
Suyo	26.93	-	-	-	26.93
Tagudin	83.84	-	-	-	83.84
<b>Total area, ha</b>	<b>7,663.56</b>	<b>156.10</b>	<b>10.41</b>	<b>0.00</b>	<b>7830.07</b>

**Appendix I. Predicted Soil Salinity in Different Municipalities Using MSP Model**

Municipality	Multisensor Data Association				Total Area, ha
	Low	Moderately Low	Moderately High	High	
Alilem	118.15				118.15
Banayoyo	121.6		11.24		132.84
Bantay	202.52		19.7		222.22
Burgos	489.9				489.9
Cabugao	389.1	0.02	0.51		389.63
Caoayan	43.8	6.8	89.59		140.19
City of Candon	593.6	13.21	256.04		862.85
City of Vigan	179.63	8.6	56.89		245.12
Galimuyod	108.83				108.83
Gregorio del Pilar	6.7				6.7
Lidlidda	85.2				85.2
Magsingal	484.83		46.12		530.95
Nagbukel	94.44				94.44
Narvacan	550.91	0.9	92.16		643.97
Quirino	75.45				75.45
Salcedo	91.02				91.02
San Emilio	47.57				47.57
San Esteban	74.68		27.61		102.29
San Ildefonso	60.21		4.38		64.59
San Juan	272.36	3.71	45.45		321.52
San Vicente	69.87	3.96	86.01		159.84
Santa	90.63	34.19	220.85	4.01	349.68
Santa Catalina	26.47	6.24	112.5		145.21
Santa Cruz	298.15	2.5	69.02		369.67
Santa Lucia	215	1.8	35.28		252.08
Santa Maria	591.31	8.77	80.74		680.82
Santiago	157.9	0.73	13.88		172.51
Santo Domingo	572.47	0.94	12.8		586.21
Sigay	14.92				14.92
Sinait	143.34	8.16	14.73		166.23
Sugpon	47.79				47.79
Suyo	28.62				28.62
Tagudin	77.94	5.7			83.64
<b>Total area, ha</b>	<b>6,424.91</b>	<b>106.23</b>	<b>1,295.50</b>	<b>4.01</b>	<b>7,830.65</b>