Faculty of Environmental Sciences Department of Geosciences Institute of Cartography

Master Thesis

”Analyzing and Visualizing the Relationships between Vegetation Traits and Soil Moisture in Wetland Environment from Remote Sensing Data

submitted by **Mengyu Liang**
born on 12.06.1994 in Kunming, China

submitted for the academic degree of
Master of Science (M.Sc.)

Submission on 29/07/2019

Supervisors Dr. rer. nat. Marion Pause
TU-Dresden

Dr.-Ing. Matthias Schramm
TU-Wien

Dr. rer. nat. Nikolas Prechtel
TU-Dresden

Statement of Authorship

Herewith I declare that I am the sole author of the thesis named

"Analyzing and Visualizing the Relationships between Vegetation Traits and Soil Moisture in Wetland Environment from Remote Sensing Data"

which has been submitted to the thesis assessment board today.

I have fully referenced the ideas and work of others, whether published or unpublished. Literal or analogous citations are clearly marked as such.

Dresden, 29/07/2019 Signature Mengyu Liang

Acknowledgement

I would like to thank my supervisors, Dr. Marion Pause, Dr. Matthias Schramm and Dr. Nikolas Prechtel for their insights, time, and guidance. I also would like to thank the MSc Cartography Examination Committee for their additional efforts to reschedule my thesis evaluation and defense dates. This thesis would not be able to finish in time for the start of my Ph.D. in mid-August if not for their help and coordination.

I also wish to thank all the instructors that I have had the pleasure to work with at the three institutes, TUM, TUW, and TUD. Their knowledge and insights into geosciences are the exact reasons why I wanted to study in Europe – to learn how European scientists approach science.

To my colleagues, I also wish to thank them for their genuine encouragement and for the inspirations they provide through their unique points of view.

Abstract

Soil moisture is one of the most important environmental variables for understanding different types of terrestrial ecosystems—agriculture, forest, grassland, and wetland. Although it is one of the smallest components in the hydrological cycle in volume, soil moisture, plays a critical role in many hydrological, biological and biogeochemical processes. It measures the quantity of water contained in soil, thus, it strongly links to vegetation conditions in many ecosystems. Understanding soil moisture could provide farmers, scientists, and policymakers a better chance to make wiser land management decisions and prevent disasters, such as flooding. However, measuring soil moisture has been a challenging topic. Measuring soil moisture with remote sensing technology excelled field measurements in obtaining more continuous and frequent monitoring. Satellite missions and products of soil moisture monitoring include the Soil Moisture Active Passive (SMAP), Soil Moisture Ocean Salinity (SMOS) and Advanced Scatterometer-Soil Water Index (ASCAT-SWI). On the other hand, coarse spatial resolution and complex interactions of microwave radiation with surface roughness and vegetation structure present limitations within these products to monitor soil moisture variations on the landscapes with a high resolution.

The thesis seeks to explore an alternative approach to monitor soil moisture. Instead of treating vegetation canopy as obstacles that interfere with microwave remote sensing signals, vegetation traits observed in passive remote sensing have a great potential in indicating the soil conditions underneath. Vegetation Indices (VIs) algorithms are developed by taking advantage of the electromagnetic wave reflectance information from canopies using passive sensors. They are simple yet powerful ways to quantitatively and qualitatively monitor different aspects of vegetation conditions—greenness, water stress, coverage, vigour, and growth dynamics. These vegetation traits can, in turn, be used to reflect and monitor soil water condition.

With this objective in mind, SWI product derived from ASCAT are chosen to be paired with four VIs (NDVI, NDWI, LAI, and FAPAR) to identifying relationships between soil moisture and vegetation traits. The VIs are computed from Sentinel-2 imagery collected for the year of 2016 to 2018 over two cycles of dry and wet seasons in Okavango Delta based on the regional flow dynamic measured in situ. Seasonality helps to uncover the temporal dimension of the relationship. Spatial heterogeneity is addressed through the spatial variance and spatial structure of the VI values over the landscape. Later, this research discusses what could be contributing to these variations and patterns by examining the land cover and identifying the dominant vegetation types in the study area. Lastly, a web-based interactive mapping product is developed to visualize the complex spatial-temporal relationships discussed in the remote sensing analysis. This web map platform exploits the most popular JavaScript-based graphic design libraries, Leaflet and D3, as well as other web-development techniques (HTML/CSS). The goal of this visualization is to include dense information related to the project while reducing the interaction complexities, so that the targeted users, students in remote sensing classrooms and decision-makers could benefit from the visual thinking provided by interaction with the data.

Keywords: Vegetation traits, soil moisture, remote sensing, spatial-temporal patterns, cartographic visualization, interactive maps

Table of Contents

Figures

Tables

Formulas

Abbreviations

Acronym	Full Name
AIC	Akaike Information Criterion
ASCAT	Advanced Scatterometer
AMSR-E	Advanced Microwave Scanning Radiometer - Earth Observing System
BOA	Bottom of Atmosphere
CGLS	Copernicus Global Land Service
CSS	Cascading Style Sheets
CV	Coefficient of Variation
DBOF	Deciduous Broadleaf Open Forest
ECV	Essential Climate Variable
ESA	European Space Agency
FAPAR	Fraction of Absorbed Photosynthetically Active Radiation
GCOS	Global Climate Observing System
GMLJP2	Geographic Markup Language JPEG2000
GPT	Graph Processing Tool
HTML	Hypertext Markup Language
ISMN	International Soil Moisture Network
LAI	Leaf Area Index
LCCS	Land Cover Classification System
MetOp	Meteorological Operational Satellite
MODIS	Moderate Resolution Imaging Spectroradiometer
MSE	Mean Squared Error
MSEP	Mean Squared Error of Prediction
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NetCDF	Network Common Data Form
NIR	Near Infrared
PROBA-V	PROBA-Vegetation
SD	Standard Deviation
SNAP	Sentinel Application Platform
SMAP	Soil Moisture Active Passive
SMOS	Soil Moisture Ocean Salinity
SSM	Surface Soil Moisture
STEP	Science Toolbox Exploitation Platform
SWI	soil water index
SWIR	Short Wave Infrared
TOA	Top of Atmosphere

TOC	Top of Canopy
WMS	Web Map Service
WWW	World Wide Web
VI	Vegetation Index
UI	User Interface
UNFAO	Food and Agriculture Organization of the United Nations

1 Introduction

In this chapter, the background and motivation of the research will be presented. The research questions and objectives, as well as an overview of the thesis document, will be specified.

1.1 Background

Even with its small volume, soil moisture is a critical component of the hydrological cycle. It is defined as the quantity of water contained in the space between soil particles. Soil moisture is closely linked to many hydrological, biological and biogeochemical processes in local and regional scales, such as runoffs, flooding, plant evapotranspiration, soil erosion, water quality and so on. Therefore, it is of importance for governmental agencies, scientists, farmers to monitor its change and relations with climate and weather.

Traditional field-based measurement of soil moisture generally has high accuracy but fails to provide continuous spatial coverage and regular sampling rate. Volumetric soil moisture content is calculated through weighting water and soil separately and calculating the ratio of the volume of water to the volume of soil. This drying and weighing method (gravimetric method) was destructive and time-consuming. Later on, portable field measurement devices/sensors were developed to measure the horizontal and vertical profiles of soil water in situ (Johnson, 1962; Romano, 2014). Nevertheless, the problems with the spatial coverage and the revisiting rate remain a big challenge for regional-scale soil monitoring.

Great alleviation of the above issues is accompanied by the development of remote sensing-based soil moisture monitoring products. Traditional remote sensing methods observe the signals reflected in the microwave region. Passive microwave radiation remote sensing is sensitive to surface soil moisture, and active microwave detect energy emitted from the deeper layer of soil (Calvet et al., 2011). The most well-known soil moisture products developed from micro-wave radiation instruments includes the Advanced Scatterometer (ASCAT) onboard Meteorological Operational Satellite (MetOp) developed by the European Space Agency (ESA) in 2006, the Soil Moisture Ocean Salinity (SMOS) products by ESA in 2009, and the Soil Moisture Active Passive (SMAP) products by National Aeronautics and Space Administration (NASA) in 2015. Limitations with the microwave radiometer for soil moisture monitoring is related to the coarse spatial resolution and the complex interaction of the signal with the surface vertical structure, such as vegetation structure, surface roughness, and atmosphere. Therefore, an alternative approach to monitoring soil condition needs to be proposed to both take advantage of the continuous spatial coverage and high temporal resolution of remote sensing techniques and to increase the spatial resolution while reducing some of the signal distortions from interaction with ground canopy.

The reasons for specifically targeting at the wetland environment in this analysis of soil moisture/vegetation traits relations are — first, the microwave radiation remote sensing products for wetland soil monitoring have been very noisy in the past due to the complex land cover mixing between water and vegetation; second, peatlands present important ecological significance and the degradation of peatlands from land use and climate change could lead to emissions of greenhouse gases back to the atmosphere, thus, more understanding of peatlands' ecological behaviors are in need. Another inspiration for examining this relationship in the Okavango Delta is due to its wetter-than-normal and wetter-than-surrounding soil moisture signal as shown in the soil moisture anomaly analysis using the Essential Climate Variable (ECV) Soil Moisture product done by the Institute for Photogrammetry and Remote Sensing at the Vienna University of Technology (Dorigo et al., 2015). Last but not least, Okavango Delta, located in southern central Africa, has a great ecological significance by providing food, water and shelter for animals and humans with its dense vegetation canopy. The diverse vegetation types and complex land-water interaction presents challenges, but it also presents opportunities for scientists to exploit the vegetation information for soil moisture information retrieval.

1.2 Motivation and Problem Statement

High-resolution soil moisture information at local to regional scale is critical for understanding soil moisture's role in hydrological and agricultural processes at a decision-making level. Some examples of these processes are flooding, drought, irrigation, infiltration, runoff, and these processes occur mostly at local to regional scale. Soil moisture also plays a part in the energy cycle - soil water acts as source for evapotranspiration, and approximately 60% of precipitation over land is returned to the atmosphere through evapotranspiration (Oki & Kanae, 2006), and this process uses more than 50% of radiation over land (Dirmeyer et al., 2006). Other studies have also indicated that through coupling with evapotranspiration, soil moisture has an impact on the climatic system through influencing temperature and precipitation (Seneviratne et al., 2010), yet a non-negligible factor in soil moisture-climate interaction is the vegetation and its spatial-temporal variations. To further dissect the roles of soil moisture in any of these processes and make truly relevant decisions in land use and management, soil moisture observation at a high spatial resolution and temporal resolution is of high demand.

Current remote sensing soil moisture products retrieve information in the microwave region of the electromagnetic spectrum, and they commonly have high temporal repetition rate but very coarse to coarse spatial resolution. The ASCAT-SWI product is derived from Surface Soil Moisture (SSM) using a simple water balance model (Wagner et al., 1999); the data is available for 2007 to present at 25km spatial resolution for the world. Higher-resolution SWI (1km) is recently available for Europe using Sentinel-1/C-band SAR and MetOp/ASCAT for 2015 to present (Bauer-Marschallinger et al., 2018). Additionally, the well-known SMAP product is released since 2007 and is available at 3-5 days interval, and the SMOS data provides global soil moisture measurement at 35-50 km resolution with a revisit time of 1-3 days since the first data release in 2010. There are also many on-going efforts related to the spatial downscaling of soil moisture product, the SWI product using Sentinel-1 data is a good example, however, the spatial resolution and coverage remain challenging.

On the other hand, optical remote sensing tends to have a higher spatial resolution—Landsat has a spatial resolution of 30m and Sentinel-2 has a spatial resolution of 10m in selected bands (RGB and Near Infrared channels). Vegetation traits derived from optical remote sensing, thus, could provide a great opportunity to indirectly measure various environmental conditions at a finer resolution, including soil humidity conditions. Using vegetation as sensors for soil moisture also reduces the noise from the complex interaction between microwave radiation and vertical structure of the canopy. Instead of treating vegetation as obstacles for soil moisture information retrieval, VIs calculated from spectral information of remote sensing shed light on the soil condition underneath.

Specifically in wetland, land cover is particularly complex. Seasonal wetlands, permanent wetlands and different types of vegetation intermingled with each other form an extremely challenging landscape for microwave spectrometer detection. The vertical structures over these types of landscapes can be extremely heterogeneous over a short distance; the mixing of open water, vegetation, and peats within one footprint of coarse resolution microwave sensor present difficulties for understanding the true spatial variations in the wetland environment. For instance, with the 25 km resolution ASCAT product, soil moisture cannot be estimated if a large portion of the footprint is composed of open water or dense vegetation as specified in the product handbook (Bartalis et al., 2008). Therefore, obtaining vegetation information from fine spatial resolution multispectral data and using this information to estimate soil moisture condition provide an opportunity to closely examine variations over the wetland ecosystem.

Another aspect this topic will be how to visualize the complex natural processes with multiple dimensions on a 2-dimensional cartographic interface so that the targeted audience can benefit from the research results more easily. The targeted users for such a visualization are students in a remote sensing classroom seeking to understand the data-fusion approach to develop a case study and decision-makers who are interested in building in-situ sensor networks to monitor different environmental variables. Hence, this visualization product needs to provide dense information/data of both spatial and temporal aspects in an easy-to-use interface. With the web-based interface and the fusion of many JavaScript graphic design libraries, multidimensional information can be added through User Interface (UI) elements more easily, such as a slider bar and overlays, thus, allowing users to interact with the data and graphics more dynamically and rapidly (Roth, 2017). This cartographic visualization, therefore, adds value to the remote sensing analysis of the research by allowing the complex spatial-temporal aspects of the data to be communicated with the targeted audience more interactively.

1.3 Research Questions and Objectives

Based on the motivations and problem statements, this research aims to understand the relationships between vegetation and soil moisture in the Okavango Delta through fusing passive and active remote sensing data. More specifically, this project has the following four objectives:

First, to understand the seasonal variations of the relationship between soil moisture and vegetation measured in remote sensing data.

The temporal dimension of the problem is approached through data retrieved on 10 dates over a course two year, 4 of which are dates from the dry season, 6 of which are representing the wet season. The relationship between soil and vegetation are analyzed for these two seasons.

Second, to understand how the relationship between soil moisture and vegetation varies with different levels of spatial heterogeneity and within different land cover.

The spatial aspect of the problem is approached through sampling sites across the Delta and examining the land cover condition in each site. Analysis is conducted for sites with different dominant land cover separately for comparison. Characteristic statistics addressing the spatial variation and vegetation structure are retrieved for each site.

Third, to evaluate to what extent the vegetation traits observed through remote sensing data could be used to estimate soil moisture in the Okavango Delta environment.

Correlation between individual vegetation characters and SWI are used to examine correlation strength. Regression analysis is performed to summarize the relations between vegetation traits and soil moisture.

Fourth, to present appropriate web-based cartographic techniques for visualizing these multi-dimensional patterns and results to a wider audience.

The cartographic visualization utilizes the web browsers and provides dynamic and interactive displays of remote sensing data and geospatial data used in the remote sensing analysis by taking advantage of the popular JavaScript libraries.

1.4 Overview of Content

This thesis document contains six chapters. The first chapter introduces the issues related to soil moisture information retrieval and the opportunities for using vegetation as sensors for soil moisture monitoring. The second chapter discusses the achievements, challenges and recent advancements in the field of soil moisture monitoring with remote sensing technology and multidimensional visualization in modern cartography. The third chapter explains the data and methods adopted to sample, analyze and visualize data to approach the spatial and temporal aspects of the research questions. The fourth chapter presents the results of the analysis results of different aspects of the relations. The fifth chapter is dedicated to discussing the potential reasons for the observed results, to address limitation in the analysis and problems encountered. The sixth chapter summarizes the results and emphasizes some additional aspects of the research that could be analyzed more in-depth in the future using additional datasets. Lastly, the work cited will be presented.

2 Related Work

2.1 Soil Moisture Monitoring

2.1.1 Importance of Soil Moisture Monitoring

Soil moisture or the available soil water content is defined as the amount of water contained in the root zone that can be utilized by plants (Mahmood, 1996). It has long been recognized as one of the most critical natural variables for understanding environmental processes in local and regional scales. It is closely related to hydrological, atmospheric, geomorphic, hydrologic, and biological processes, and interacts with many components in the ecosystem, such as vegetation, climate, and energy (Legates et al., 2011). Practically speaking, soil moisture provides essential information for agriculture management, flood/runoff monitoring and weather/climate forecast in local to regional scale. Therefore, soil water content is an integrated topic in Earth System studies.

2.1.2 Efforts in Soil Moisture Monitoring

Due to its high significance in natural environments, soil moisture has solicited much attention in its measurement and monitoring. Soil moisture has been measured in the field or in situ with destructive methods such as volumetric soil moisture measurement by weighting-drying-weighting soil samples, or with portable field equipment and sensors (Johnson, 1962; Romano, 2014). Soil moisture has also been monitored indirectly by remote sensors at a larger spatial scale (Ochsner et al., 2013). Higher temporal resolution and regular observation can be achieved when using remote sensing techniques. Surface soil moisture observations are retrieved from sensors sensing at microwave and optical/thermal infrared wavelength. Some examples of soil moisture products from microwave observations include the well-known SMOS, SMAP, ASCAT, and AMSR-E (Advanced Microwave Scanning Radiometer - Earth Observing System) products as discussed below.

One of the earliest satellite-borne remote sensing soil moisture products is derived from ASCAT on board of MetOp. The SSM product from ASCAT is available since 2007 to present day and has a spatial resolution of 25km. The SWI product is calculated from the ASCAT SSM using a two-layer water balance model accounts for moisture conditions as water infiltrates the soil profile over a recursive time period of observation; this model does not account for soil texture (Wagner et al., 1999). The ASCAT-SWI is now available at a higher spatial resolution (1km) for Europe when calculated fusing Sentinel-1/C-band SAR as the latest advancement in soil moisture product derivation (Bauer-Marschallinger et al., 2018). Another well-known soil moisture products, SMAP, is a NASA mission that combines an L-band radar and an L-band radiometer for con-current soil moisture measurement. L-band SAR provides backscatter measurements at resolutions from 1km to 3km. The SMAP data is released since 2007 and is available at 3-5 days interval. SMOS product is developed by ESA provides global soil moisture measurement at 35-50 km resolution with a revisit time of 1-3 days since first data release in 2010. Recent advancement in soil moisture measurement is accompanied by the release of the Sentinel-1 mission. The new SWI product, currently only available for Europe publicly, is derived by a fusion of SSM observations from Sentinel-1 and ASCAT based on the same two-layer water model discussed above (Bauer-Marschallinger et al., 2018), it is able to increase the spatial resolution to 1km through the fusion with higher resolution data from Sentinel-1.

Moreover, atmospheric and hydrologic modeling (Fulakeza et al., 2002; Steiner et al., 2005), and soil moisture networks at national and international levels (Jackson et al., 2012; Sanchez et al., 2012) also contribute to the efforts of increasing the amount and improving the quality of soil moisture data. Especially, soil moisture monitoring sensor networks provide valuable information for calibrating and validating remote sensing observations and models (Dorigo et al., 2011; Jackson et al., 2012; Sanchez et al., 2012). For example, the International Soil Moisture Network (ISMN) is an international effort to host and maintain global in-situ soil moisture datasets. Therefore, there has been no shortage of efforts in the past two decades to obtain and improve soil moisture observations worldwide; many different techniques in field measurements, remote sensing observation, and modeling have been adopted along the way.

2.1.3 Challenges of Current Soil Moisture Monitoring

Even though with the number of advancements related to soil moisture information retrieval methods, many challenges remain. The challenges related to the soil moisture retrieval processes are to improve the accuracy of the soil moisture data and generate long-term soil moisture products with high spatial and temporal resolution. Specifically, as the previous section has discussed, most the publicly available soil moisture datasets have a coarse spatial resolution in the 10-30km range, and the new 1km Sentinel-1 fused SWI dataset is currently only available for Europe (Bauer-Marschallinger et al., 2018). Another aspect of the challenge is related to the fact that most soil moisture observations are obtained from signals reflected in the microwave region (Calvet et al., 2011). Microwave radiation is sensitive to surface roughness; the microwave signal has a complex interaction with vegetation and other vertical structures on the ground. Generally, the increase of surface roughness corresponds to the increase in radar cross-sections for all combination of polarization (HH, VV, HV); cross-section is a measure of a targeted object's ability to reflect radar signals in the direction of the radar receiver, or how detectable the targeted object is by radar (Elachi & Van, 2006). Hence, as the surface roughness increases or the vegetation canopy gets higher, the difference between the signal polarization combination decreases and cause noise to be larger for sensing the soil moisture condition underneath. More importantly, a large percentage of the landscape of high ecological significance is covered by vegetation to different extents, hence using microwave signals to derive soil moisture data will always have some degree of uncertainty over vegetated land (Bartalis et al., 2008). Additionally, microwave signals can penetrate vegetation and soil on the ground surface to a certain extent, and the penetration capacity decreases as the depth increases, thus, to derive information about water content in the soil column will need to involve some infiltration models (Bartalis et al., 2008; Wagner et al., 1999). To conclude, challenges remain alongside the vast efforts and progress, and they are remainders that alternative soil moisture retrieval methods and data-fusion approaches should be closely examined to conquer the current issues.

2.2 Multidimensional Cartography Visualization

Modern cartography is not only an important tool that adds value to the geodata and geoscientific analysis but also a way to improve the non-domain audience's understanding of complex natural phenomena. As the internet and World Wide Web (WWW) mature and popularize, enabling a new medium for data dissemination, modern maps have advanced beyond the static or printed format as what maps are traditionally perceived (Plewe, 1997). Web maps are reinvented as tools not only for the presentation of data and results but also for dynamic data exploration, data processing and even analysis (Couclelis, 1998; Köb-ben & Kraak, 1999).

2.2.1 Characteristics of Multidimensional Data and Visualization

Multidimensional or multi-variate are both characteristics describing the data structure and the relations between the data items and the depicted physical world (Fuchs & Hauser, 2009). "Multidimensional" describes the independent relations among data attributes as well as their encoding of physical dimensions such as space and time (Fuchs & Hauser, 2009). Multidimensional data can be multi-channel, multimodal, multi-field, multi-valued depending on the physical quantities, acquisition methods, attributes and measurement time (Fuchs & Hauser, 2009). In the case of describing the spatial-temporal relations between soil moisture and vegetation, multi-channel is related to the physical quantities of water level, soil moisture and VIs involved in the analysis; multimodal is related to fusing in-situ records, microwave sensing, and multispectral data to acquired information; multi-field relates to the various indices and various statistics calculated to describe vegetation traits; lastly, multi-value is related to the repetitive measures of each variable described over time. Therefore, the remote sensing analysis conducted in this research fits into the case of a spatial-temporal, multi-dimensional scenario for visualization. Moreover, to enhance the user's perception and understanding of such complex dataset and relations, innovative visualization or mapping platform is in need to effectively convey the information.

2.2.2 Importance of Modern Cartography in Multidimensional Visualization

Modern cartography's development benefited greatly from the advancement of the internet. The internet allows the dissemination and exchange of a large amount of multi-formatted data and information from diverse sources quickly (Plewe, 1997). Web mapping emerged and gained more functionalities as the browser-interpretable JavaScript language popularizes (Flanagan, 2006). JavaScript, as a programming language, is high-level and object-oriented, meaning programs are developed around data or objects instead of logic or functions, and it allows methods and functionalities to be specified for the targeted objects. Hypertext Markup Language (HTML) /Cascade Style Sheet (CSS) are standard technologies for web page creation and layout design; when combined with JavaScript, they form the fundamentals for a modern web-based mapping platform.

Web maps can visualize complex multi-dimensional data because of the combination of the interaction operators. Firstly, a UI is what links human users to computer; in the context of web mapping, it is an element on the graphic interface to maximize the user's ability to manipulate maps and their underlying geographic information. UI design in the context of web mapping focuses on the interaction operator primitives. These interaction operators, such as panning, zooming, detail retrieval, and overlay, are fundamental for users to perform data exploration, visual thinking and to generate new insights via interacting with the graphics (Roth, 2017). More specifically, for display and rendering of geospatial information on the web at an acceptable response time, many web maps nowadays separate information into groups, they are basemaps, thematic layers, and interactive elements. Basemaps provide contextual geographical information for the map; they are not delivering the key information of the map, but the web map could be uninterpretable without them. Thematic layers are overlays that convey the key messages or present the thematic data for exploration. They could be toggled on/off by user to facilitate visual thinking. Another inclusive component is the interactive elements on the web maps; they are items on the maps such as popups and sliders. Interactive elements make maps alive, and in many situations, help organize the information displayed on the screen to prevent the information overload. To conclude, these theoretical concepts and technical element are the underlying bases for web-based multidimensional visualization development.

2.2.3 Remaining Challenges of Multidimensional Visualization

Even with such great advancements in mere two decades, multi-dimensional visualization still faces challenges. These challenges can be related to the new data sources, such as remote sensing observation and real-time data stream (Robinson et al., 2017), which result in large quantities of data. Facing these challenges from large volume, high velocity, high variety and high veracity, Robinson et al. (2017) pointed out the need for map-based interface for users to analyze geospatial big data and to express the spatial-temporal patterns they have in mind through such an interface (Robinson et al., 2017). They also raised the agenda that the modern cartography needs to connect with outside disciplines to produce useful and sustainable map-based visualization to tackle the multi-dimensional data challenge and benefit more audience. Another important aspect of the challenge comes from the purposes of such map-based visualization. MacEachren & Kraak (1997) have pointed out that a cartographic product should be more than a presentation tool but also an exploration tool that supports research and decision making. The traditional mindset that static maps display the location of spatial objects needs to be thoroughly replaced, and the role of cartographic products as user-oriented tools to enhance complex data exploration and to enable analytical tasks need to be demonstrated (MacEachren & Kraak, 1997). To conclude, maps-based visualization product for complex multi-dimensional data still needs innovative efforts to tackle the new and old challenges.

2.3 State-of-the-Art

While challenges remain in soil moisture product development and multidimensional visualization, there is no shortage of innovative efforts to tackle the issues by improving current technology and adopting new approaches.

2.3.1 Innovation in Soil Moisture Monitoring

As discussed above, the new Sentinel-1/ASCAT fusion product increased the spatial resolution of soil moisture monitoring to 1 km for Europe as of now (Bauer-Marschallinger et al., 2018); it is one of the most recent breakthroughs in soil moisture monitoring. Progress has been made by innovatively using vegetation traits observed from passive remote sensing data to the soil moisture condition underneath. Alexandridis et al. (2017) experimented with using optical and thermal data from the Moderate Resolution Imaging Spectroradiometer (MODIS) in combination with ancillary soil and meteorology data to estimate root-zone soil moisture based on energy fluxes. They produced soil moisture maps for their study sites in Europe at 250m resolution. This demonstrates the potential of using optical data in soil moisture estimation. Torres-Rua et al. (2016) tested using VIs (such as NDVI, NDWI, LAI) calculated from Landsat 7, surface energy balance product and meteorology data, as well as analytical methods such as statistical learning machine, stratified cross-validation and forward variable selection. Moreover, Klinke et al. (2018) used plant characteristics and temperature information calculated from the Sentinel's and Landsat's thermal and visible infrared wavelengths as indicators or proxies of mean soil moisture in a wetland environment. The progress demonstrates the great potential multispectral remote sensing data has in soil moisture information retrieval. Especially, the vegetation traits calculated using optical sensor data provide unique opportunities to examine the soil conditions under the canopy while overcoming the complex interaction microwave signals has with the vertical canopy structure.

2.3.2 Innovation in Multidimensional Visualization

Innovation in web-based multidimensional visualization is strongly linked to the development of new software and technological solutions. Roth et al. (2015) has discussed the importance of open libraries implemented in JavaScript and identified several important JavaScript-based libraries for web mapping production and teaching. The two most-used JavaScript-based graphics APIs are D3 (Data Driven Documents) developed and maintained by Mike Bostock, and Leaflet developed and maintained by Vladimir Agafonkin. D3 was initially released in 2011, and its most recent stable version, 5.9.7, was released in June of 2019. D3 has a prominent advantage in its flexibility to dynamically render and project graphics in general and also spatial data. It also supports explicitly multi-view visualization or coordinated display. This means that geodata could be projected dynamically based on the user-specified mathematical functions, and other information display in non-map formats, such as charts and graphs, could also be linked to aid data exploration and analysis. The other three libraries mainly display spatial data on slippy map style with conventional Web Mercator projection for fast tileset rendering. Comparatively, Leaflet was released initially in 2011, and its most recent stable release, version 1.4.0, was in December of 2018. Leaflet supports spatial dataset rendering on a tiled slippy map. The innate functionalities of the library include some critical interactions such as pan, zoom, and overlay. The small file size and the supported touch-based interactions make it easy to implement for mobile design. Multi-view or coordinated display is possible but only through implicit modification within the Leaflet library. Leaflet functionalities are also stretched by a series of plugins by its user communities due to it being open-source, for example, temporal sliders and non-tiled data from Web Map Service (WMS) are plugins developed by Leaflet users. With these two exemplary graphics libraries based in JavaScript and on the web browser, many flexible mapping platforms are made possible for users to interact with different aspects of the dataset. Other interactive, web-based mapping solutions also contribute to the efforts of improving the multidimensional data visualization; for example, Mapbox, Google Map API, OpenLayers, Shiny (R package) are also under active development and maintenance. They demonstrate the availability and advancements of technical solutions and software for web-mapping and multidimensional data visualizations.

3 Methodology

3.1 Overview

This research seeks to answer the spatial and temporal relations of soil-vegetation using a data-fusion approach. The in-situ datasets reveal the regional flow dynamics in the Okavango Delta, and the C-band microwave ASCAT soil moisture product is combined with the VIs computed from multispectral Sentinel-2A data to analyze the relations between soil moisture conditions and vegetation traits. Multiple characteristic statistics are calculated for each VI to exploit different aspects of the indices and better correlate vegetation with soil moisture.

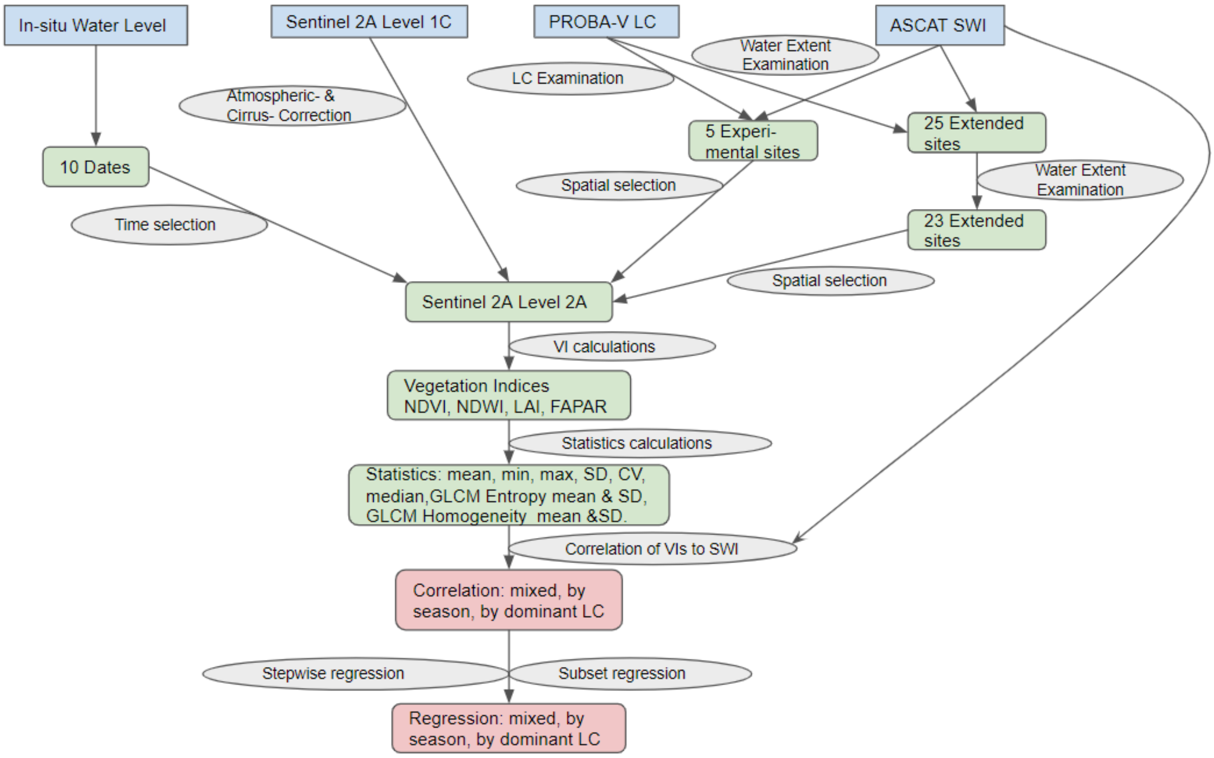


Figure 1. Research workflow.

3.2 Study Area

The Okavango Delta is situated in northern Botswana between -18.23 and -18.51 °S, 21.84 and 23.81 °E, occupying approximately 19,253 km^2 depending on the season. The climate in Botswana is semi-arid. The Okavango Delta region receives approximately 400~500mm precipitation annually on average, and the mean annual temperature ranges from 15~20°C. Traditionally, the wet season begins in December, peaks in January and February, and finishes by March. The water in the Okavango delta is mainly supplied through the Okavango River from the high plateau in central Angola instead of local rainfall. The altitude of the Angolan plateau ranges from 1,200 to 1,800m. The rainfall is more intense in the Angolan plateau with a cooler climate minimizing evaporation; the average rainfall ranges from 1000mm to 1300mm annually. Generally, the month with peak rainfall in Okavango delays the peak month in Angolan plateau by approximately a month, which indicates there is a delay in the discharge received in Okavango from when Angolan plateau receives its rainfall, in addition to the delay in other local hydrological processes such as infiltration.

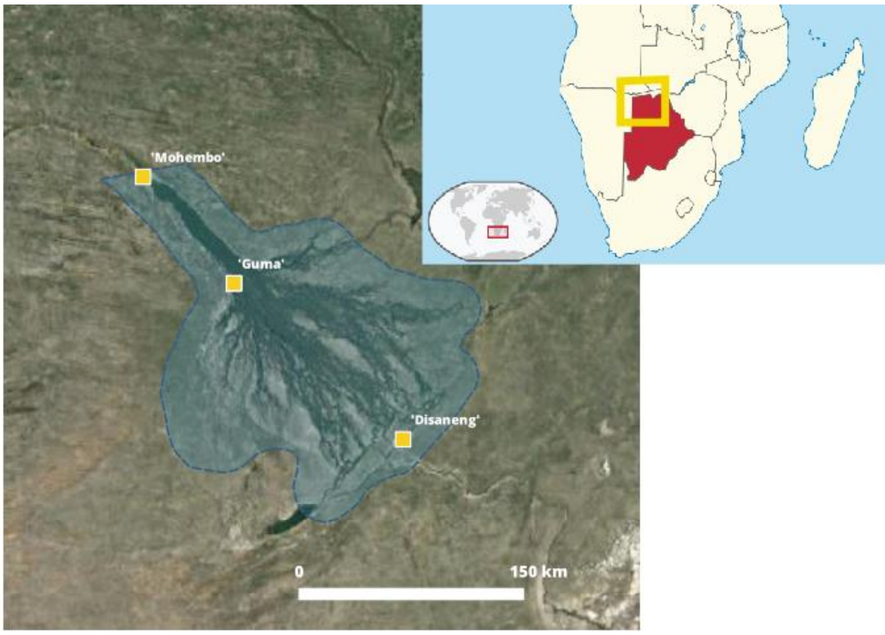


Figure 2. Extent of the Okavango Delta and the locations of the in-situ water stations.

3.3 Date and Data processing

Data	Spatial Resolution	Temporal Coverage	Source
Sentinel-2A Level-1C	10m, 20m, 60m	11/22/2016~06/05/2018	Earth Explorer
ASCAT-SWI	0.1 ° grid space	11/22/2016~06/05/2018	Copernicus Global Land Service (CGLS)
In-situ Water Level	Site-based	06/01/2016~06/01/2018	Okavango Delta Monitoring & Forecasting
PROBA-Vegetation (PROBA-V) Land Cover	100m	2015	CGLS

Table 1. Overview of primary and ancillary dataset.

3.3.1 Sentinel-2 Data

Sentinel-2A Level-1C data is archived by the USGS's EarthExploerer and downloadable as zip files, which contain image data in Geographic Markup Language JPEG2000 (GMLJP2) format and additional metadata. All images covering the Okavango Delta from 2016 to 2018 are initially examined, and 10 dates are chosen for the temporal analysis based on the cloud coverage and the reflection of seasonality in corresponding to the in-situ water level records. The dates chosen are 2016-11-22, 2016-12-02, 2017-04-01, 2017-04-11, 2017-04-21, 2017-11-07, 2018-01-06, 2018-04-26, 2018-05-16, 2018-06-05. Among these dates, 2016-11-22, 2016-12-02, 2017-11-07, 2018-01-06 (Figure 3) are representing the low water level or the dry season in the Delta, and the rest dates represents the wet season of the Delta. Spatially speaking, six 100 km x 100 km tiles with a UTM/WGS84 (Universal Transverse Mercator/World Geodetic System 1984) projection and datum are selected to cover the Delta.

Band number	S2A	
	Central wavelength (nm)	**Resolution (m)**
1 - Coastal Aerosol	427	27
2 - Blue	492.4	10
3 - Green	559.8	10
4 - Red	664.6	10
5 – Vegetation Red Edge	704.1	20
6 - Vegetation Red Edge	740.5	20
7 - Vegetation Red Edge	782.8	20
8 - Near Infrared (NIR)	832.8	10
8a - Vegetation Red Edge	864.7	20
9 – Water Vapor	945.1	60
10 – Short Wave Infrared (SWIR)-Cirrus	1373.5	60
11 - SWIR	1613.7	20
12 - SWIR	2202.4	20

Table 2. Sentinel-2A product bandwidth and resolution.

The Level 1C Top-of-Atmosphere (TOA) reflectance data then undergoes atmospheric- and cirrus correction using the Sen2Cor processor (version 2.8.0) distributed by Sentinel Toolbox Exploitation Platform (STEP). The batch processing is done with the command line tool of Sentinel Application Platform (SNAP), the SNAP Graph Processing Tool (GPT). The corrected results are reformatted to Level 2A Bottom-of-Atmosphere (BOA) reflectance data. The Sentinel-2 bands have different spatial resolutions (Table 2), therefore, they are resampled to Band 2 at 10m prior to any processing. The resampled and atmospherically-corrected data are later subset to spatial extents of the sample sites across the Delta.

3.3.2 ASCAT-SWI Data

The ASCAT-SWI data used in this research are retrieved from CGLS for the 10 dates corresponding to the Sentinel-2 dataset mentioned above. The SWI is calculated from the SSM from ASCAT on board MetOp based on a simple water infiltration model over a time period, T (Wagner et al., 1999). The SSM represents the moisture content in the thin top layer of soil, and the calculated SWI describes the percent moisture in the soil profile over the preceding time period. In other word, the SSM over the preceding time period is summed and exponentially weighted (Equation 1). T determines how fast the weights become smaller, and how strongly the SSM observations taken in the past influence the current SWI. The daily SWI data is available for different characteristic time length T (1, 5, 10, 15, 20, 40, 60 and 100 days). In this research, SWI at T=10 day is used. Another thing to note about the SWI model is, the SWI product is independent of soil texture and does not involve any vegetation information in its calculation, thus, makes it logical to use in the later correlation analysis with VIs.

$$SWI(t_n) = \frac{\sum_i^n SSM(t_i) e^{-\frac{t_n - t_i}{T}}}{\sum_i^n e^{-\frac{t_n - t_i}{T}}} \; for\; t_i \le t_n$$

Equation 1

The SWI is calculated at points located at the centers of 0.1°- wide grids. The SWI data is distributed by CGLS in Network Common Data Form (NetCDF) format; additional metadata and quality flags are also included when downloading from CGLS. The SWI at each location or grid point is matched by a cell number. When selecting the sample sites across the Delta, the pixel size of the SWI product is used as the spatial bound for each site and the cell number is used to reference each site (Figure 8).

3.3.3 Ancillary Data

The in-situ water level records measured at two stations, Mohembo (Latitude: -18.275733, Longitude: 21.787312) and Guma (Latitude: -18.96266, Longitude: 22.373213), are retrieved from the Okavango Delta Monitoring & Forecasting service. The water level is measured at a daily frequency with a small portion of missing dates in the records. The missing data are interpolated with Kriging method to remove holes for better plotting and establishing a clear overview (Figure 3). The water level records clearly show high and low seasons at these two measuring gauges of the Delta. As previously discussed, the Delta region receives a low amount of precipitation, the water levels measured at the Delta inlet stations can indicate the majority of water supply to the Delta. Therefore, dates are chosen at the extreme low and high regions of the water level; subtract from them the dates that the high cloud coverage is present in Sentinel-2 images, as a result, there are 10 dates left for further analysis. Specifically, the wet season for this research is defined as the time of the year that high water level occurs at Mohembo and Guma stations; dry season is when the water level is low at both gauges.

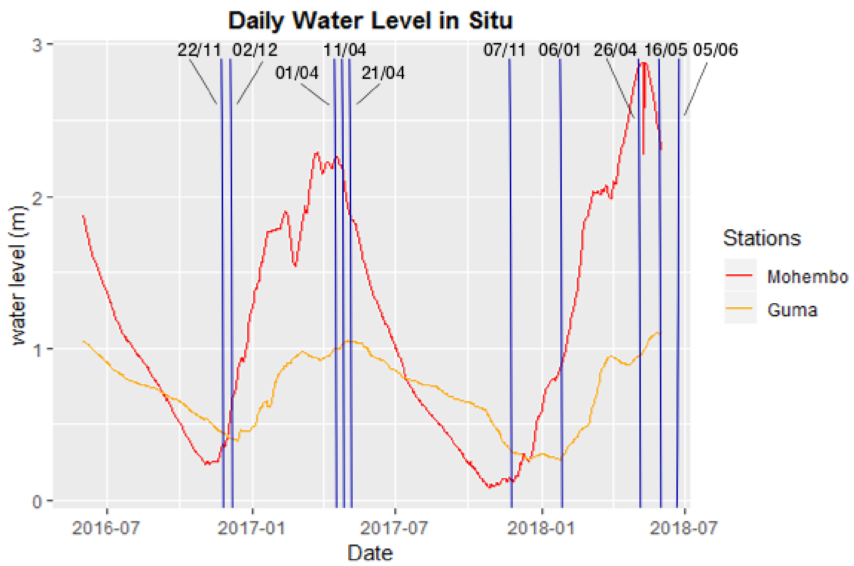


Figure 3. Water level records from 2016-06 to 2018-06 measured at Mohembo and Guma stations. Blue lines indicate the dates selected for data retrieval and further soil-vegetation correlation analysis.

Another important dataset included in this research for understanding the Delta vegetation and water extent is the Dynamic Land Cover map at 100 m resolution distribution by CGLS. This land cover product is derived from the PROBA-V time series for the year 2015 for Africa. The layer of the product mainly used in this project is the discrete land cover classification, which contains 20 classes based on the Food and Agriculture Organization of the United Nations (UNFAO) Land Cover Classification System (LCCS) standard (Figure 4), and the cover fraction layer for seasonal inland water for (Figure 5). Specifically, the cover fraction layer for seasonal inland water is used to understand the flooding extent in each SWI pixel and assess their fitness to use for analysis. The various vegetation classes are used to discuss if the patterns revealed in the later analysis are related to land cover.

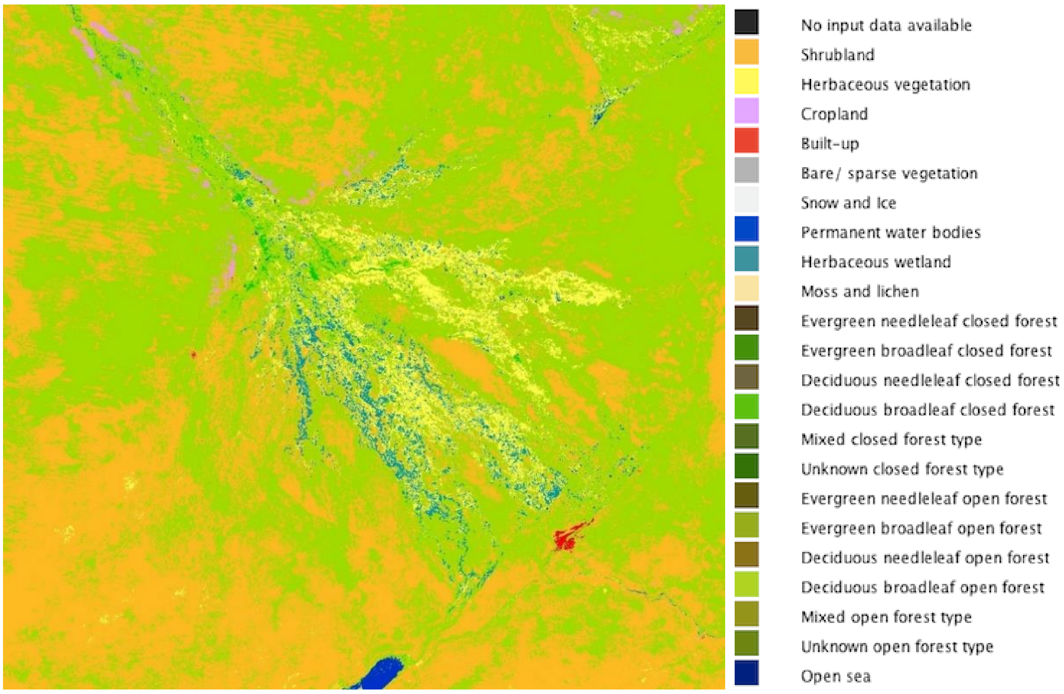


Figure 4. Dynamic Land Cover map at 100m resolution for Okavango Delta. Land cover derived from PROBA-V time series, and retrieved from CGLS.

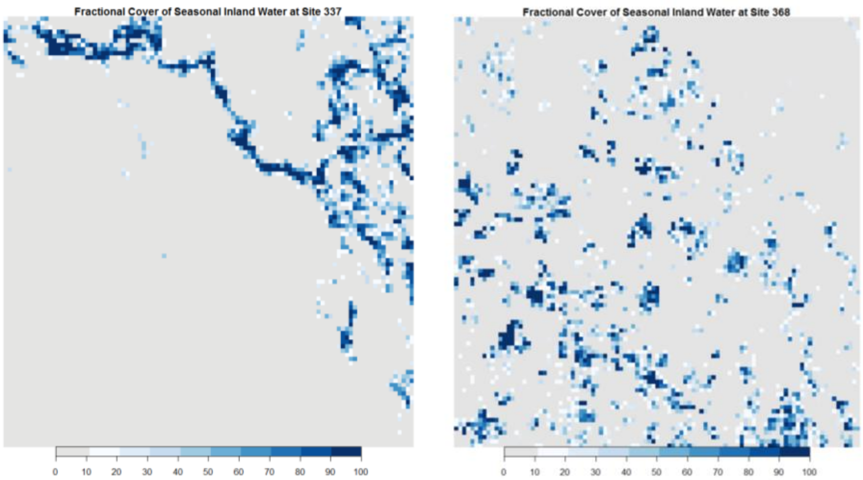


Figure 5. Cover Fractional layer of seasonal Inland water for site 337 (left) and 368 (right). Unit is percentage. These two sites are excluded from further analysis due to the relative high percentage of water that causes noise for SWI retrieval.

3.4 Sample Sites

In this project, a total of 30 sample sites are selected across the Delta. Each of the sample sites has the same spatial extent as the corresponding 0.1°- grid resolution SWI, and they will both be referenced with the cell number from the SWI grid. Firstly, five experimental sites were selected based on their geographical locations in the Delta and their land cover types to test out the methodologies and form a workflow. Then, 25 more extended sample sites are randomly selected across the Delta and are applied the same workflow as tested from the experimental sites. With 30 sites, the patterns observed and the analysis results will be more robust. However, the workflow developed for analyzing these sample sites is not suitable for every SWI pixel, because some SWI pixels are located at the boundaries of two Sentinel-2 tiles or where dense cloud cover or water occurs, therefore, analysis on a continuous spatial coverage is not suitable.

3.4.1 Experimental Sample Sites

Five sample sites are selected to test out the experimental workflow (Figure 6). The five pixels are selected to be distributed across the Delta, to contain different dominant vegetation types (Figure 7) and to avoid dense clouds and water. The site selection also avoid permanent water body to eliminate noise in SWI affecting analysis results. Site 135 is located adjacent to the in-situ water level station, Guma, in the inlet region of delta, and the dominant vegetation type here is Deciduous Broadleaf Open Forest (DBOF). Site 318 and site 359 are located at the central region of the Delta fan, but with DBOF and shrubs as dominant vegetation types respectively. In these two sites, seasonal flooding or temporary wetland still occurs, which may affect the ASCAT-SWI values. Site 373 and site 483 are located in the dry peripheral of the Delta where dominant vegetation types are DBOF and shrubs respectively.

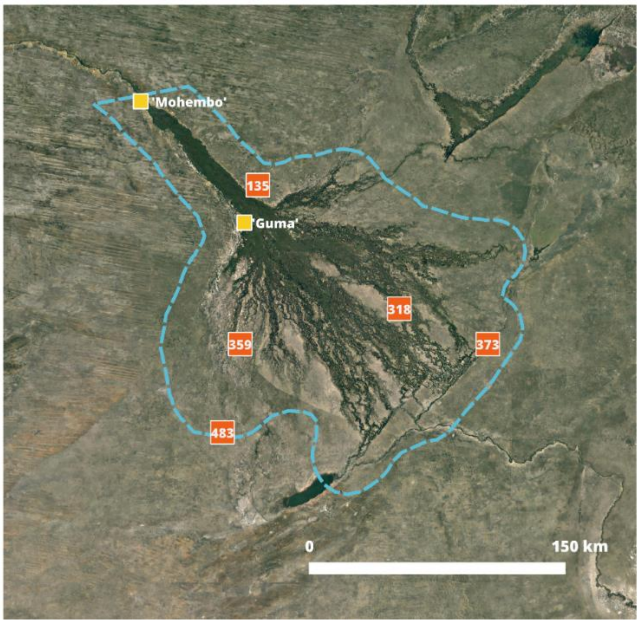


Figure 6. Locations and vegetation types of the five experimental sites. Grey rectangles indicating the Sentinel-2A tile locations.

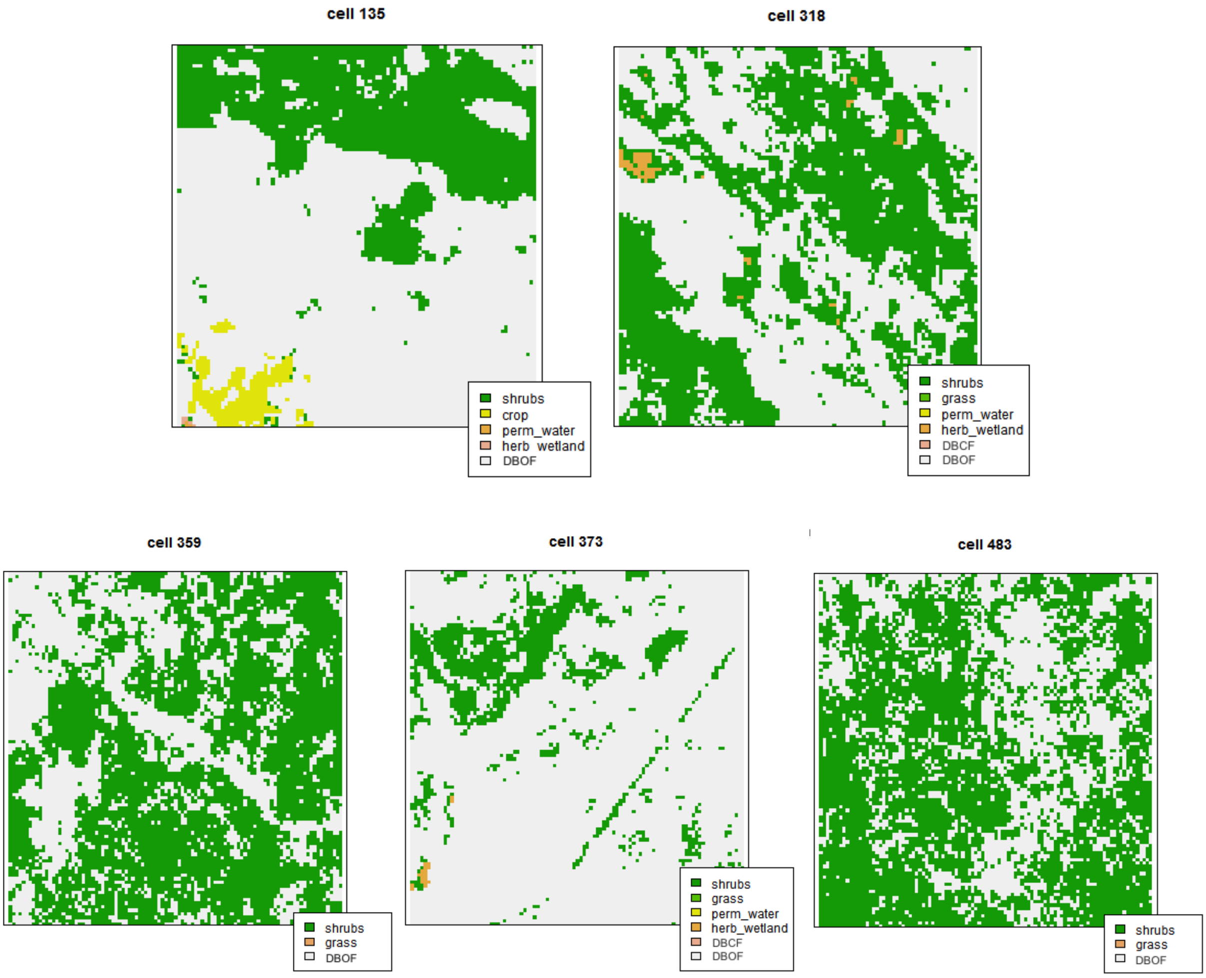


Figure 7. Land cover composite for the five experimental sites.

Upon examining the vegetation type and water extent within the experimental sites, four VIs, Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), Leaf Area Index (LAI), and Fraction of Absorbed Photosynthetically Active Radiation (FAPAR), are calculated for the atmospherically corrected Sentinel-2 scenes over the 10 selected dates. The grey value images of the VIs are then subset to the spatial extent of each site in RStudio using R package "ratser". Then, for each site, statistics with different summarizing capacities are calculated for each index to reflect on different characteristics of vegetation traits within each site location in RStudio. Afterward, each VI is correlated to SWI in scatterplots to examine correlation strength based on seasons and dominant vegetation types.

3.4.2 Extended Sample Sites

After testing the experiment plan with the five sites, a random site selection is conducted over all SWI pixels within the entire Delta extent in RStudio. The random selection first generated 60 sample sites across the Delta. Then the sites located in the Sentinel-2 tile boundaries, in the dense cloud covered regions or permanent swamp regions are manually removed, leaving 25 sites to form the extended sample sites (Figure 8). Upon the examination of the cover fraction of inland seasonal water (Figure 5), site 368 and 337 are removed because of their relatively high seasonal water coverage based on 2015 LCCS (6.57% and 4.92% respectively). On each of the extended sites, the experiment plan described above—land cover examination, VI retrieval, statistics retrieval, correlation analysis—is applied (Figure 1). With the extended dataset of VIs statistics, a more robust regression analysis can be conducted.

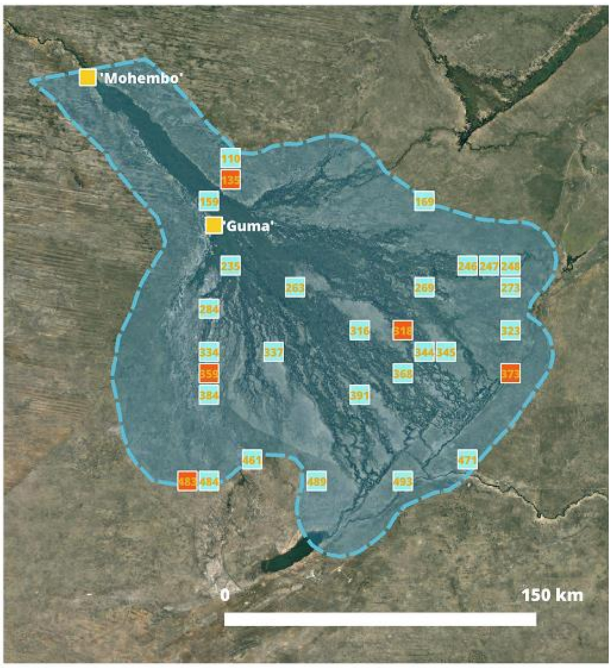


Figure 8 Locations of all 20 sites, including 5 experimental sites and 25 extended sites.

Site	Shrubs	DBOF	Grass	Herbaceous Wetland	DBCF	Cropland	Built-up	Permanent Water	Temporary water
110	40.92	59.08	0	0	0	0	0	0	0
135	29.06	67.55	0	0.05	0	3.31	0	0.03	0
159	4.93	64.25	15.21	6.07	9.34	0	0	0.20	0
169	33.03	66.97	0	0	0	0	0	0	0
235	9.52	57.84	25.49	5.26	1.88	0	0	0.01	0
246	8.45	50.90	40.35	0.21	0.01	0	0	0.09	0
247	8.61	85.93	5.27	0.05	0	0	0	0.14	0
248	12.07	87.64	0.20	0.07	0	0	0.02	0.01	0
263	5.50	68.96	18.82	5.83	0.88	0	0	0	0
269	17.61	64.51	17.29	0.38	0	0	0	0.20	0.01
273	8.49	91.51	0	0	0	0	0	0	0
284	42.51	44.98	6.66	5.86	0	0	0	0	0
316	35.08	56.32	5.69	2.91	0	0	0	0	0
318	17.11	52.68	29.34	0.68	0.12	0	0	0.08	0

323	29.74	70.26	0	0	0	0	0	0	0
334	50.53	47.98	1.16	0.33	0	0	0	0	0
337	26.44	58.15	5.89	9.52	0	0	0	0	0
344	28.08	57.35	13.71	0.51	0	0	0	0.35	0
345	58.34	41.57	0.09	0	0	0	0	0	0
359	59.09	39.27	1.64	0	0	0	0	0	0.0
368	3.82	26.61	51.75	17.81	0	0	0	0.01	0
373	15.52	84.00	0.28	0.16	0.01	0	0	0.03	0
384	57.76	42.23	0.01	0	0	0	0	0	0
391	32.36	57.93	5.17	4.54	0	0	0	0	0
461	37.48	62.52	0	0	0	0	0	0	0
471	42.64	56.91	0.21	0	0	0.24	0	0	0
483	60.78	37.87	1.35	0	0	0	0	0	0
484	79.59	19.89	0.52	0	0	0	0	0	0
489	82.28	17.72	0	0	0	0	0	0	0
493	70.63	28.10	0.96	0.01	0	0.21	0.039	0.06	0

Table 3. Discrete Land cover (%) for all sites; sites highlighted in orange are the experimental sites; blue highlights indicate the sites with high seasonal inland water based on the fractional land cover product and are removed from further analysis; red bolded cells indicate the dominant land cover type in each site.

3.5 VIs Retrieval

Four VIs, NDVI, NDWI, LAI, and FAPAR, are retrieved for the 6 tiles from the pre-processed Sentinel-2 imagery of the 10 selected dates. VIs are calculated using the algorithms built in SNAP with the band specifications for Sentinel-2 products. The batch calculation is also done using the SNAP GPT. The results are saved as NetCDF files. NDVI is one of the most used VIs that measures the photosynthetic activity of vegetation and describes the vitality of vegetation on Earth's Surface. It is included here to correlate with SWI and analyze if the vitality and greeness of vegetation are related to soil water content. The algorithms for calculating NDVI is as below (Equation 2) and Band 4 is used as RED and Band 8 is used as NIR.

$$NDVI = (NIR - RED)/(NIR + RED)$$

Equation 2

In addition to NDVI, NDWI is another important VI that measures the liquid water content in canopy that interacts with incoming solar radiation (Gao, 1996). NDWI is included to analyze if the water content in vegetation canopies are related to the water content in soil. NDWI generally increases as the vegetation fractions and the leaf layer increase, while NDWI is generally negative in areas with soil (Gao, 1996). Gao (1996) also suggested NDWI contain information independent to NDVI. NDWI is calculated using Band 8 as NIR band and Band 12 as SWIR band for the Sentinel-2 products based on the formula below (Equation 3).

$$NDWI = (NIR - SWIR) / (NIR + SWIR)$$

Equation 3

Following the NDVI and NDWI, LAI and FAPAR are calculated together using the Biophysical Processor in SNAP. LAI and FAPAR are both ECVs recognized by the Global Climate Observing System (GCOS). LAI is defined as half the developed area of photosynthetically active elements of the vegetation per unit horizontal ground area. LAI is used to determine the size of the interface for energy and mass exchange between canopy and atmosphere, thus, is related to vegetation's evapotranspiration and photosynthetic primary production capacity. Generally, vegetation LAI estimated from remote sensing observations take into account all green contributors (understory and top forest canopy) and is strongly scale dependent (Weiss & Baret, 2016). Weiss & Baret (2016) also explained that the remote sensing LAI is sensitive to "effective" LAI, which is the LAI that would produce the same remote sensing signal as that actually recorded under the assumption of random distribution of leaves. FAPAR measures the fraction of photosynthetically active radiation absorbed by the canopy, and it corresponds to the canopy's primary productivity of photosynthesis (Weiss & Baret, 2016). Both indices are included to understand the vegetation's evapotranspiration and photosynthetic primary production capacity in relation to soil water condition. Based on the ESA SNAP Toolbox algorithm descriptions, neutral networks are calibrated every time for each of the variables during calculation, which engages the Top of Canopy (TOC) reflectance data for canopy characteristic estimation and sets the angles for defining the observational configuration. Based on the SNAP documentation, the neutral network includes the following:

- one input layer made of the 11 normalized input data (cos(θ_s)°, cos(θ_v), cos($\theta_\emptyset$), and the TOC reflectances in the 8 SENTINEL2 wavebands).
- one hidden layers with 5 neurons with tangent sigmoid transfer functions.
- one output layer with a linear transfer function (Weiss & Baret, 2016).

3.6 Statistics Retrieval

After obtaining the VIs for the 6 tiles over 10 dates, characteristic statistics are calculated to describe the different aspects of the VIs. The NetCDF files containing VIs are first read into the local environment using package "raster" in RStudio, the variable layers corresponding to each VI is selected when rasters are read. Then the rasters are subset to the spatial extent of the individual sample site. A series of statistics are calculated on the subset images in RStudio, including mean, standard deviation (SD), coefficient of variation (CV), maximum, minimum, median, second order homogeneity (mean & SD) and entropy (mean & SD). These statistics are grouped to, the ones that represent the VI's central tendency at each site (mean); the ones that describe the VI's spatial heterogeneity (SD, CV), and the one describes the site's texture or vegetation structure (entropy, homogeneity).

Mean (μ) of the VIs is calculated using the following formula (Equation 4) to provide a direct representation of the central tendency of the entire dataset.

$$\mu = \frac{\sum x}{n}$$

Equation 4

SD (σ) is included to understand the variation of data values in the VIs about the mean, and it is calculated using the following formula (Equation 5):

$$\sigma = \sqrt{\frac{\sum |x-\bar{x}|^2}{n}}$$

Equation 5

CV is calculated as the ratio of SD to mean, it measures the relative variability in the dataset; it adjusts the variation for the mean, so it is useful for comparing across data values from different datasets with different mean; also, CV, in comparison with SD brings out the variability in data overshadowed by low SD (Equation 6).

$$CV = \frac{\sigma}{\mu}$$

Equation 6

Lastly, image texture of remotely sensed data is able to capture gradients in vegetation structure that may be overshadowed by the discrete land cover, thus, it is introduced to summarize another aspect of the VIs (Wood et al., 2012). Grey Level Co-occurrence Matrices (GLCM) is a statistical texture analysis method that describes the spatial distribution of the observed intensity pairs, their relative locations to each other, thus, it extracts second order spatial statistics such as Angular Second Moment (ASM), contrast, dissimilarity, correlation, energy, homogeneity, and entropy (Hall-Beyer, 2017). Due to the inertial correlations among some of these measurements, only two statistical measures, entropy and homogeneity are selected to each represent the orderliness group and contrast group of measures. Entropy (inversely correlated to ASM with r^2 = -0.87 under same window size) mainly measures the disorderness of image pixels, in other words, when GLCM matrix has the same values, the entropy is the highest (Equation 7). Homogeneity (inversely correlated to contrast with r^2 = -0.80 and dissimilarity with r^2 = -0.95 under same window size) measures the closeness of the distribution of elements in the GLCM to the

GLCM diagonal (Equation 8). The mean and SD for entropy and homogeneity are calculated to summarize the texture analysis for each VI. The calculation is conducted using the "glcm" package in R using N= 32 as the number of grey levels for all directions (0 degrees, 45 degrees, 90 degrees, and 135 degrees). A 3 × 3 window size for all image texture analyses. This window size has the advantage of capturing heterogeneity of pixel values over small distance (Lu & Batistella, 2005; Wood et al., 2012).

$$entropy = \sum_{i,j=0}^{N-1} -\ln(P_{ij})P_{ij}$$

Equation 7

$$homogenieity = \sum_{i,j=0}^{N-1} \frac{P_{ij}}{1+(i-j)^2}$$

Equation 8

i is the row number and j is the column number. P_{ij} is the probability value recorded for the cell i,j; N is the number of rows or columns.

3.7 Regression Analysis

Upon the retrieval of the statistics for each of the four VIs over 28 sample sites, scatterplots are used to examine the correlation between the VI and SWI by seasons and by dominant vegetation type. Scatterplots are created using the R package "ggplot2". Then multiple linear regressions are conducted to model the relationship between the selected VI statistics and SWI in RStudio. In the multiple linear regression, SWI is the response or dependent variable, and the independent variables used in the regression are selected from the VI statistics, leaving out min, max and median which are subject to the influence of outliers. Other statistics are also conditionally included based on their static significance in the correlation analysis. Only one statistic from each of the three statically group is included in the regression analysis, which include the mean, one of the spatial heterogeneity statistics, and one of the textural statistics, in order to reduce collinearity introduced into the model. The correlation strength will determine which one to include. The multiple regressions are run for observations at 28 sites as well as for the dry and wet season observations separately. Stepwise regression and subset regression by the R packages "MASS" and "olsrr" (Tools for Building OLS Regression Models) respectively are used to select a subset of independent variables to optimize the model. The stepwise regression analysis in "MASS" package, "stepAIC", adds and removes independent variables iteratively in both forward and backward selection direction to optimize the final model's AIC (Akaike Information Criterion); "ANOVA" test is used in the process to compare models in the sequence of steps taken and determine the reduction in AIC (Venables & Ripley, 2002). The new output model consists of a reduced set of independent variables, and with this reduction in the number of independent variables, the subset regression analysis can be computed in a reasonable amount of time. The subset regression analysis is useful for observing the inclusion or addition of new independent variables and how it affects the performance of models based on the defined objective criterion, such as R^2, Adjusted R^2, Mellow's Cp, Mean Squared Error (MSE), and AIC. Finally, the different VI statistics' contribution to SWI estimation can be accessed through these steps of regression analysis (Figure 9).

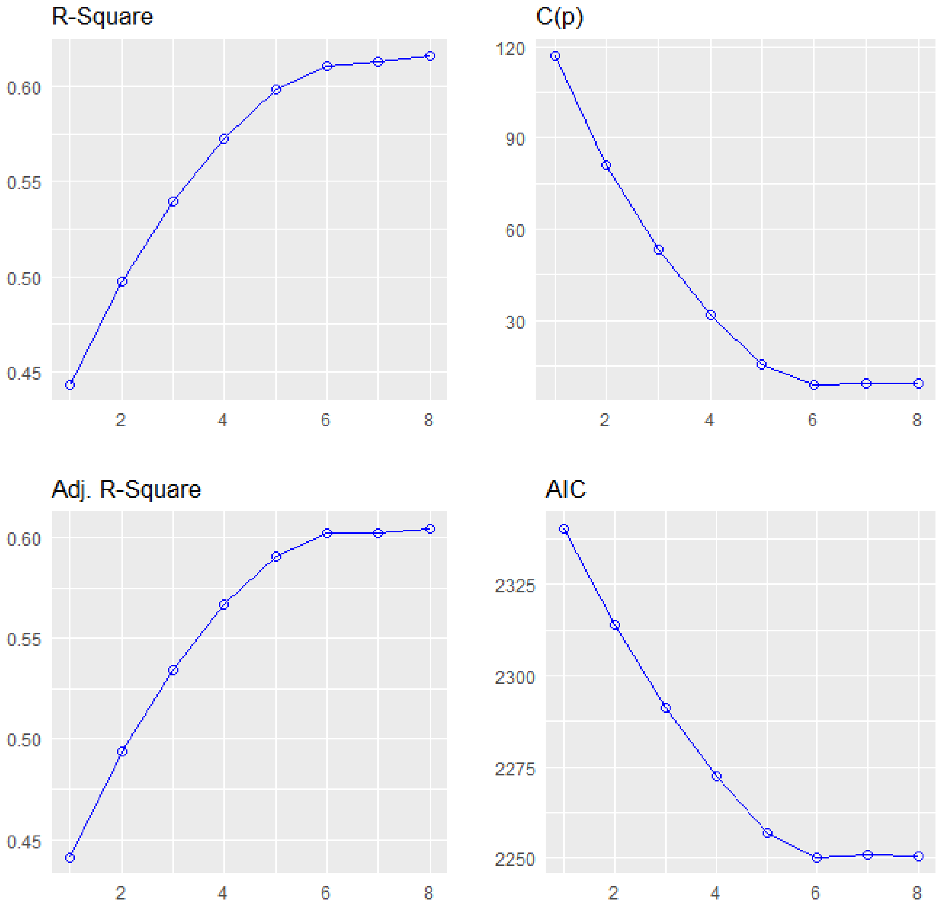


Figure 9. Plots generated by "olsrr" package in RStudio displaying the evaluation statistics for subset regression models.

3.8 Web-based Cartographic Visualization

The interactive cartographic visualization product is realized in a web-based browser environment with the support of HTML/CSS and various JavaScript libraries. HTML and CSS combination manages the structure and layout of the web platform. The most commonly used graphic JavaScript APIs and libraries are D3.js, Leaflet, and three.js. Due to the need for a simple slippy-style, 2-dimensional map interface for this visualization product, Leaflet library is selected to implement the main UI. D3.js is also used to produce a dynamic graph for information retrieval but not as the main UI. Responsive web design is included in the visualization design but only intended for large and medium screens, for example, regular desktop screens, laptop, and tablet screens (Figure 10).

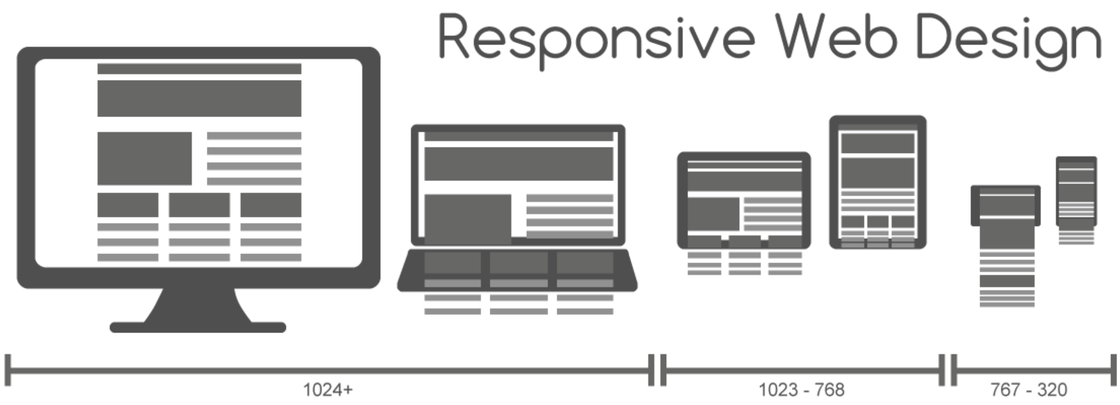


Figure 10. Responsive web design for the common screen sizes (Wikipedia Commons).

The data and overlays implemented in the visualization correspond to the methods and are mainly derived from the analysis. They are reformatted and assigned appropriate color tables (for raster images) for better visualization effect and easy comparison. The water level line graph is interpolated for the missing values in the original record using Kriging method in order for the interaction and information retrieval to be smoother (Figure 11). The VI panel is activated upon user request through clicking events on the highlighted experimental sites; the images showing different VIs are plotted in RStudio as level plots using R package "lattice" and styled with ColorBrewer's diverging color palettes (Figure 12). Corresponding legends are also produced for each VI image. Plotting the images as level plots will ensure consistent legend for each VI even when image values have different ranges. The images showing the spatially interpolated correlation strength between each VI and SWI for every sample site over 10 dates, and they are generated using ordinary Kriging method under the "Geostatistical Analyst" tool in ArcMap. The semivariogram model used is a Stable model (Equation 9) which takes account into parameters describing the variogram's shape and allows the model to change curvature while still maintaining the same nugget, range, and sill (automatically calculated from the data in ArcGIS). The Kriging results are exported as raster images to embed in Leaflet as map overlays (Figure 13).

$$\boldsymbol{\gamma}(\boldsymbol{h};\ \boldsymbol{\theta}) = \boldsymbol{\theta}_s\left[\mathbf{1} - \boldsymbol{exp}\left(-\mathbf{3}\left(\frac{||\boldsymbol{h}||}{\boldsymbol{\theta}_r}\right)^{\boldsymbol{\theta}_e}\right)\right] \boldsymbol{for\ all\ h};$$

Equation 9. $\theta_s \geq 0\ and\ 0 \leq \theta_e \leq 2$ (Esri, 2001).

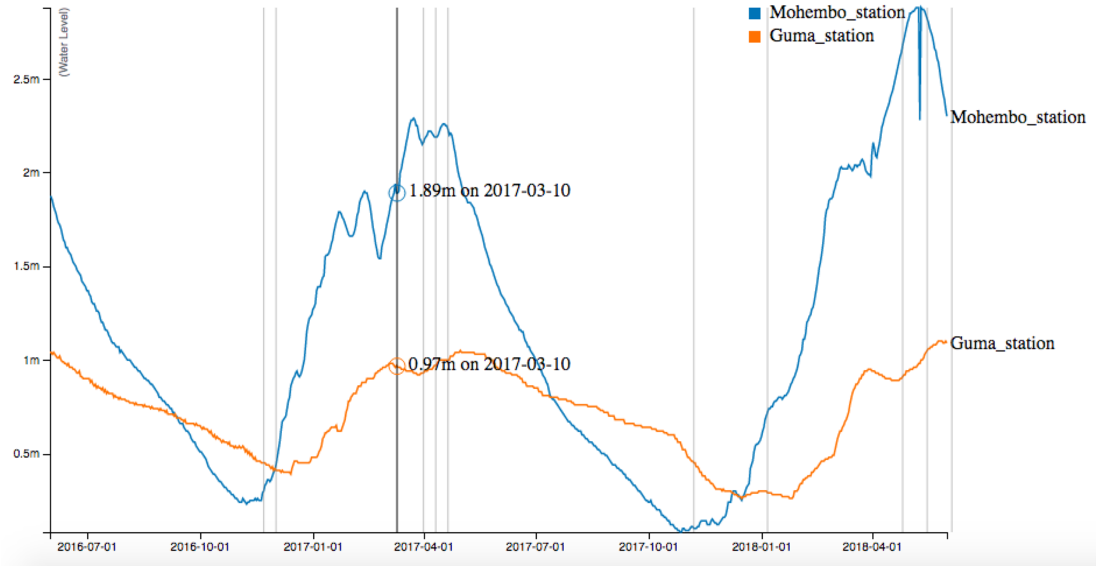


Figure 11. Water level chart UI; implemented in D3.js.

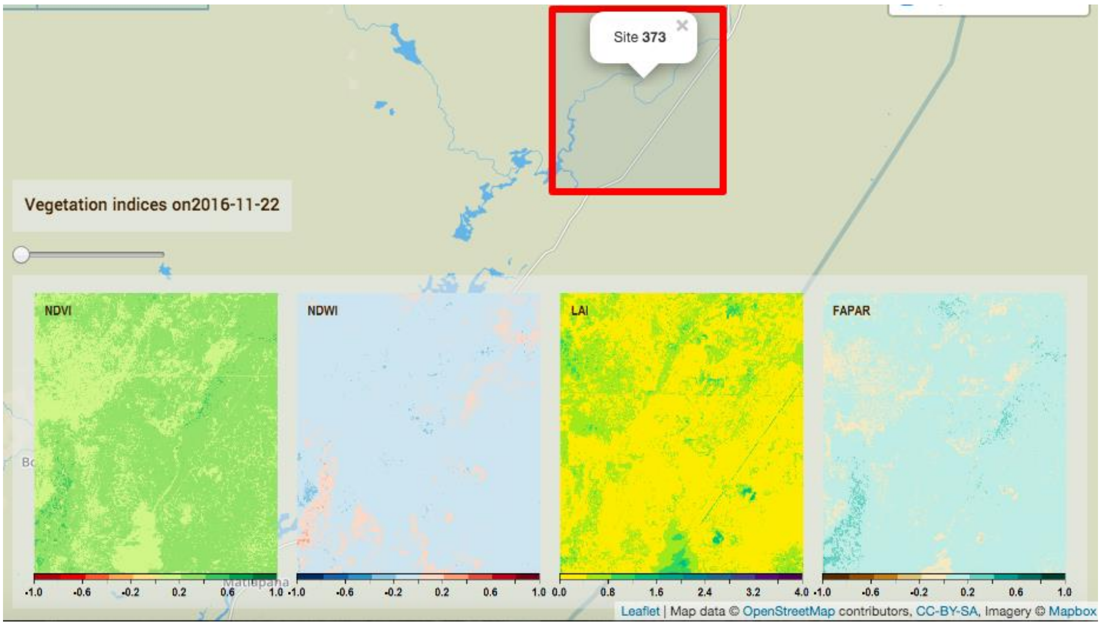


Figure 12. VI panel; square highlighted in red is the one of the experimental site.

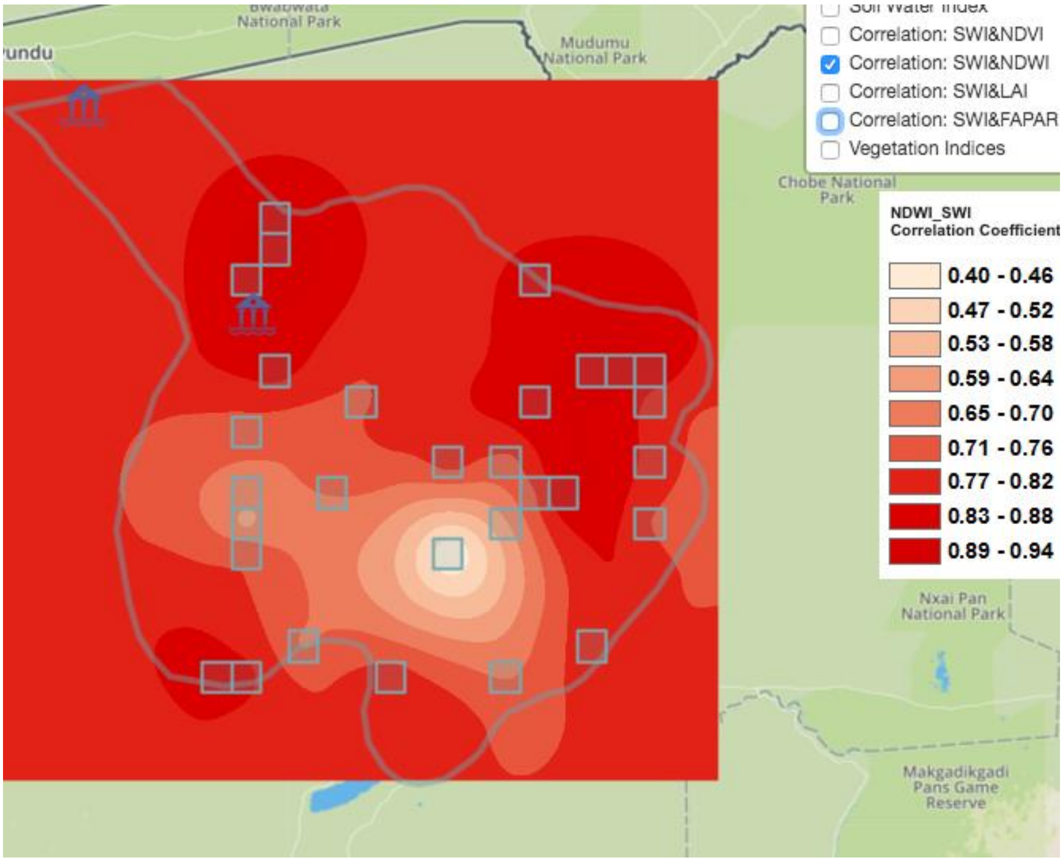


Figure 13. Kriging results of correlation between each VI and SWI; Kriging results are produced in ArcGIS with the Geostatistical Wizard.

4 Results

The results from implementing the methodology described in the previous chapter (Figure 1) will be presented here. The plots and tables generated from each step will be embedded to explain and summarize the relations and findings. The implemented web-based interactive map platform results will also be presented.

4.1 Land Cover and Water Extent Examination

Upon the examination of land cover composition of all 30 sample sites based on the 2015 100m PROBA-V Discrete Land Cover Classification and the cover fraction layer for seasonal water, the percentage of land cover in each site is calculated, dominant land cover is selected, and sites with relative high seasonal water extent are removed (Table 3; Figure 5).

4.1.1 Dominant Land Cover for Sample Sites

Dominant land cover types among the entire delta are the DBOF, shrubs, grassland, and herbaceous wetland. They each took up 54.51%, 37.43%, 5.19% and 1.92% of the total Delta area respectively. Based on the discrete land cover classes, permanent water is 0.24%, and temporary water is 0.1% of the Delta. Similar pattern can be discovered in all 30 sample sites. 70% of sample sites have dominant land cover being DBOF, 26.67% being shrubs, and 3.33% being grass.

4.1.2 Permanent and Seasonal Water Extent

Water extent in the Delta environment needs to be examined when using soil moisture data from ASCAT. Because the ASCAT-SWI product does not account for evapotranspiration. When groundwater or precipitation appears, the satellite observation could mistake the surface water being evaporated as the water filtrating into deeper soil layers, hence, causing the SWI method to have noise when the deep soil layers are affected by groundwater. In the 100m discrete land cover, very small percentages of permanent and seasonal water are observed, thus, the cover fraction layers for permanent and seasonal inland water are further examined. The cover fraction layer for permanent water only indicated sparse water extent (Figure 14), but the cover fraction layer for seasonal inland water showed a greater extent (. Figure 15). Upon the calculation the total area of fraction cover for seasonal water in each of the 30 sample sites (Table 4), site 368 and site 337 showed relative large extent, thus, are excluded (Figure 5).

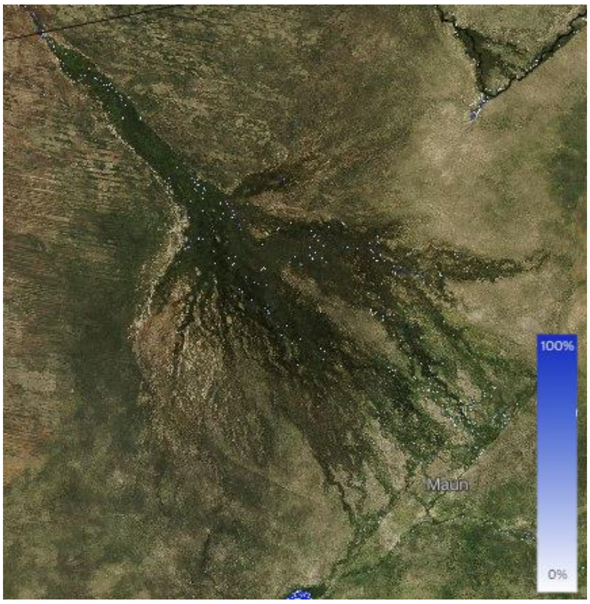


Figure 14. Fractional cover for permanent inland water.

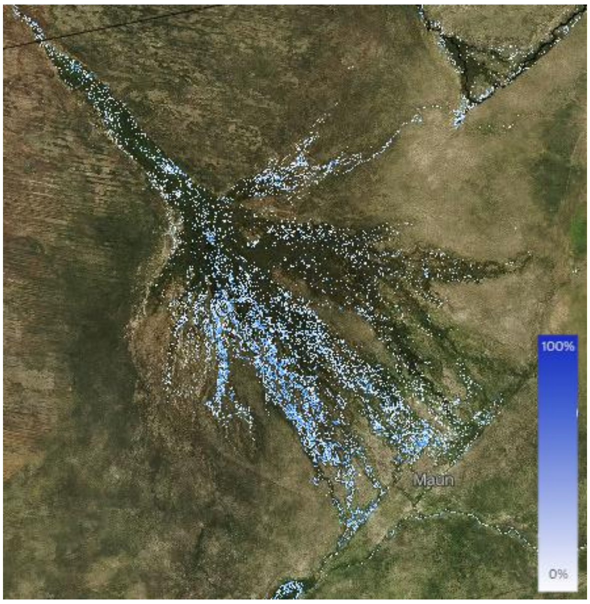


Figure 15. Fraction cover for layer seasonal inland water.

Site	Percent area of seasonal inland water
110	0
135	0.03
159	3.05
169	0
235	2.16
246	0.43
247	0.11
248	0.15
263	2.08
269	0.54
273	0
284	3.32
316	2.12
318	0.9
323	0
334	0.96
337	4.92
344	0.77
345	0.1
359	0.5
368	6.57
373	0.24
384	0
391	2.87
461	0
471	0
483	0
484	0
489	0
493	0.55

Table 4. Percent area of seasonal inland water in each site based on PROBA-V cover fraction layer.

4.2 Experimental Sites

4.2.1 VI over Time

With the land cover and water conditions examined for each site, the four vegetation indices, NDVI, NDWI, LAI, and FAPAR, are calculated for the ten selected dates (Figure 3) based on the water level to reflect seasonality. With the observations over 10 dates (Figure 16) it can be concluded that, one the VI and SWI signals are in phase with the water level dynamics observed from in-situ data; two, the mean VIs and SWI have similar trends and correlated relatively well at least in the selected time stamps. However, in Figure 3's water level records, Mohembo station has a higher peak in April 2018 than in April 2017, the SWI and mean VIs in Figure 16 shows higher values in April 2017 than April 2018. This misalignment can be related to the time delay in water extraction of vegetation from soils.

Again, due to the limitation of cloud-free Sentinel images to establish regularly spaced time series of VIs, this time delay behavior is hard to decide.

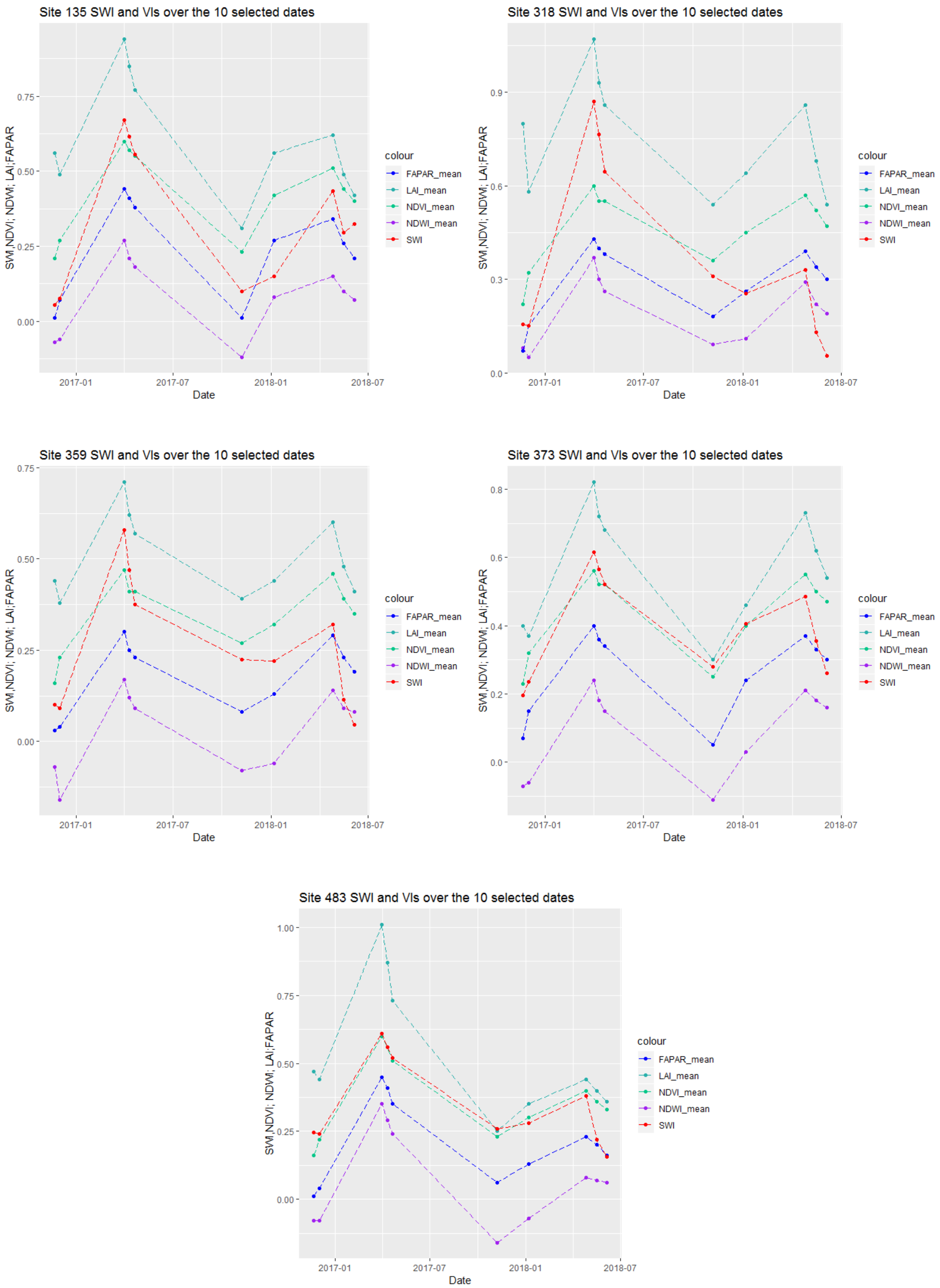


Figure 16. Unevenly spaced time series of SWI and VIs for 5 experimental sites.

4.2.2 VI and SWI Correlations

Corresponding behaviors can be observed between mean VI values and SWI from the unevenly spaced time series in Figure 16, and they are in phase with the regional water level dynamics. Correlation directions and strength between SWI and key VI statistics are shown and discussed in this section via scatterplots for the experimental sites from the time (season) aspect and the spatial (land cover) aspect.

4.2.2.1 Correlation by Seasons

Scatterplots below show the correlations of VI statistics to SWI; they are color-coded to show seasonal difference. Each point in the plot indicates one observation at one experimental site for one date, thus, 50 points are plotted per graph. Clear positive linear relationships can be observed in the mean plots of the four VIs (Figure 17). Stronger dispersions exist for observations of the dry season (red) than from the wet seasons (blue) as indicated in Figure 17; wet season observations show highly positive correlations in mean plots, especially the LAI mean. In the scatterplots, the observations for the two seasons do not completely form separate clusters, the lower range of the wet season observations intermingles with some dry season ones at mid to low VIs and SWI signals. This may be related to the contribution of other water sources, such as ground water and precipitation, in vegetation conditions, but they are not in the scope of the wet/dry season definition in this research.

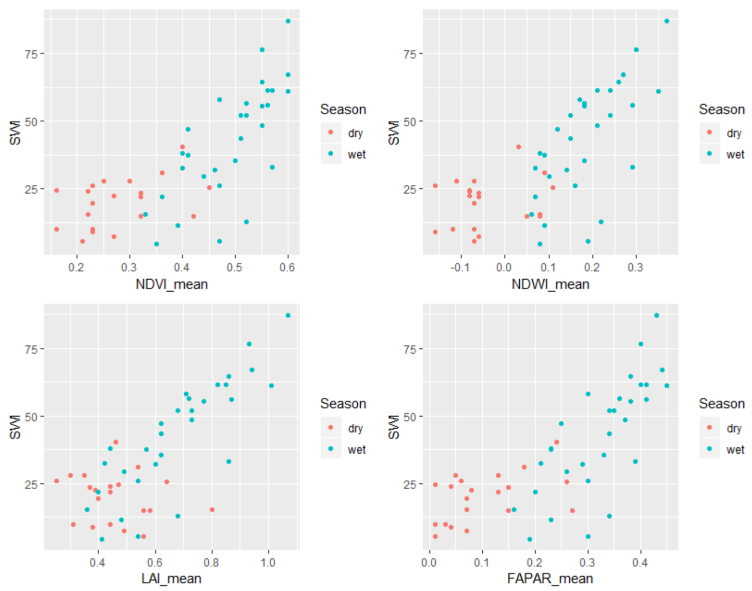


Figure 17. Scatterplot showing the correlation between mean values of the four VIs and SWI.

In Figure 18, observations for both dry and wet seasons are scattered widely for almost all the VIs, except a moderate positive linear correlation can be found in the LAI SD to SWI for the wet season.

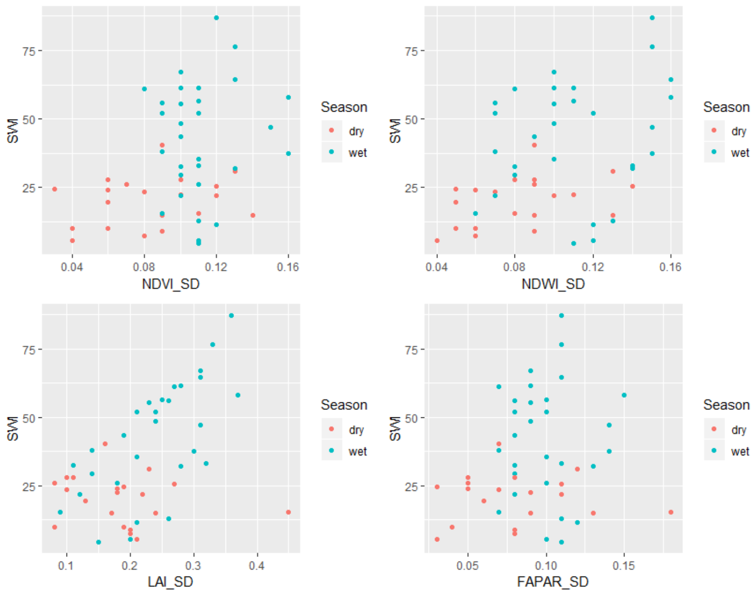


Figure 18. Scatterplots showing the correlation between SD value of the VIs and SWI.

CV values comparing to SD values brings out the variation in data at the lower end of SD. In Figure 19, the CV for NDWI are negative in the dry season, this is resulted from the negative means for NDWI at some sites, but they are not meaningful because CV measures relative variability. Generally speaking, wet season observations for NDVI, LAI, and FAPAR has a smaller range of CVs than in dry seasons. In NDVI and FAPAR, a negative correlation between CV and SWI can be observed.

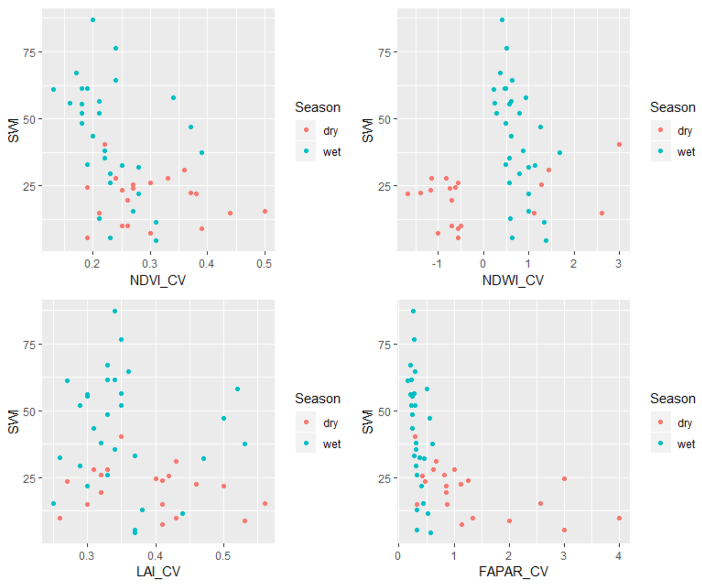


Figure 19. Scatterplots showing the correlation between CV value of the VIs and SWI.

Figure 20 and Figure 21 plot the mean and SD values for entropy (calculated from the GLCM) for the four VIs to indicate the level of disorderness in the VIs structure. Smooth image values usually result in high entropy. Clear correlations can be observed for LAIs, with mean of LAI entropy positively correlated to the SWI and LAI entropy SD negatively correlated to SWI.

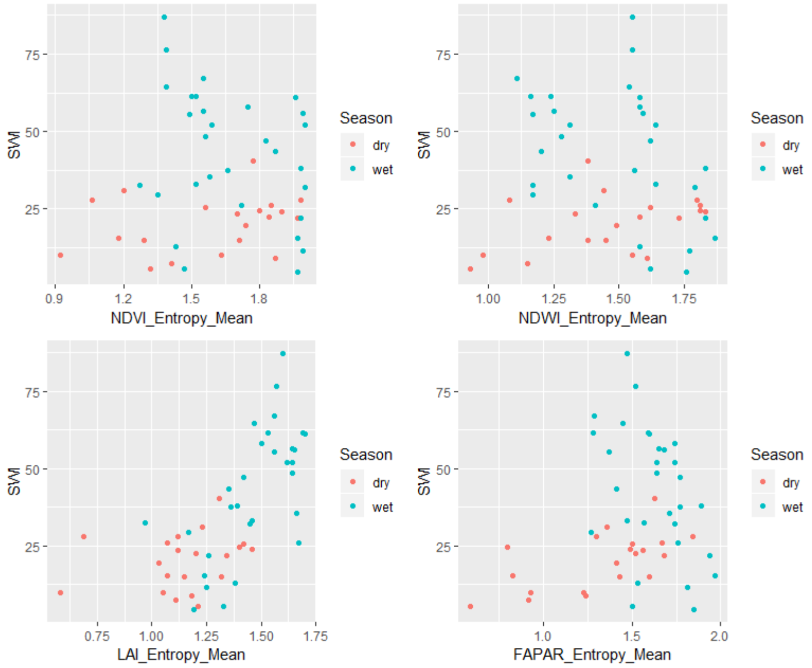


Figure 20. Scatterplots showing the correlation between mean Entropy of the VIs and SWI.

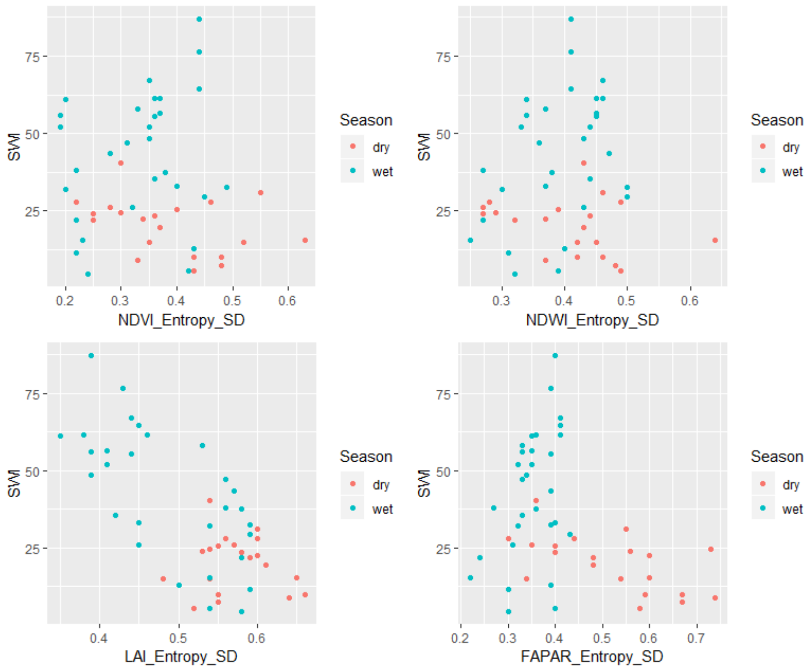


Figure 21. Scatterplots showing the correlation between Entropy SD of the VIs and SWI.

4.2.2.2 Correlation by Dominant Vegetation type

In the following figures, scatterplots correlating the VI statistics and SWI are color-coded by the dominant vegetation types. Base on Table 3, there are only two dominant land cover types, DBOF and shrubs, in the experimental sites and the extended sample sites which will be discussed more in detail in the next subchapter. In Figure 22 ~ Figure 25, the two vegetation types correlated similarly in the mean, SD and CV statistics to SWI. Shrubs are more scattered towards the higher ends of mean entropy and lower ends of Entropy SD in NDVI, NDWI and FAPAR, indicating low level of disorderness in these VIs. But in terms of the LAI, the entropy mean and SD observations are similar for both vegetation types. These can be explained as that the vitality in biomass, water content and photosynthetic energy absorption are relatively invariant in the shrub-dominant site, versus DBOF sites could have more heterogeneity in these three aspects. In terms of the LAI, both vegetation types varied to a similar extent, and reasons for this could be the heterogeneity in the vegetation's leaf density and that each site has a mixture of other vegetation type's signals in addition to the domain one.

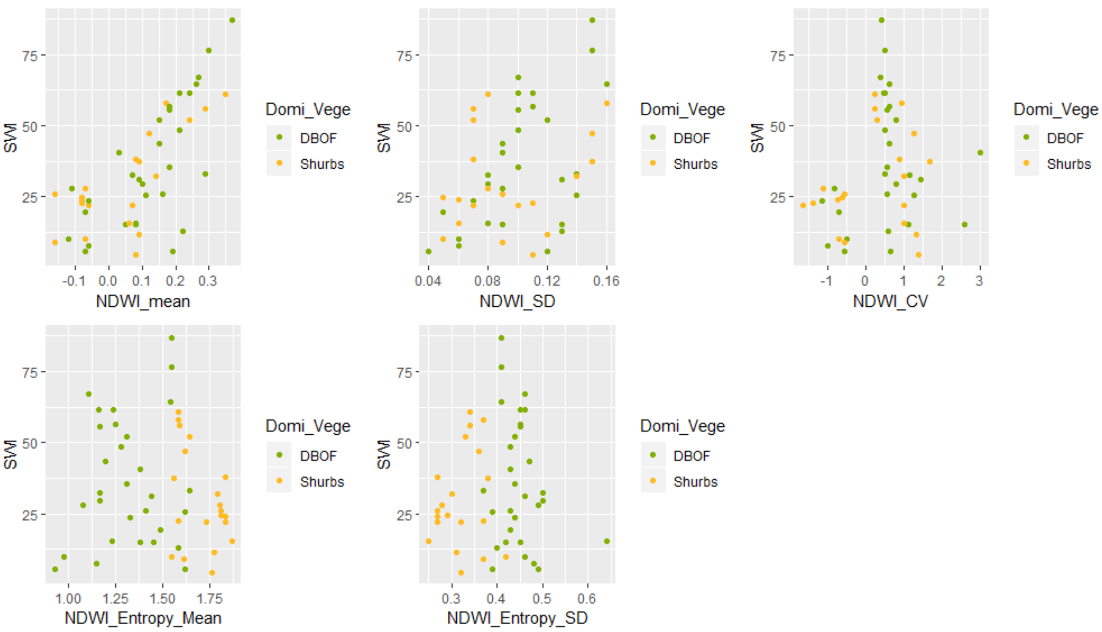


Figure 22. Scatterplot illustrates the relationship between NDWI statistics and SWI, coloured by dominant land cover type to show the variation based on land cover.

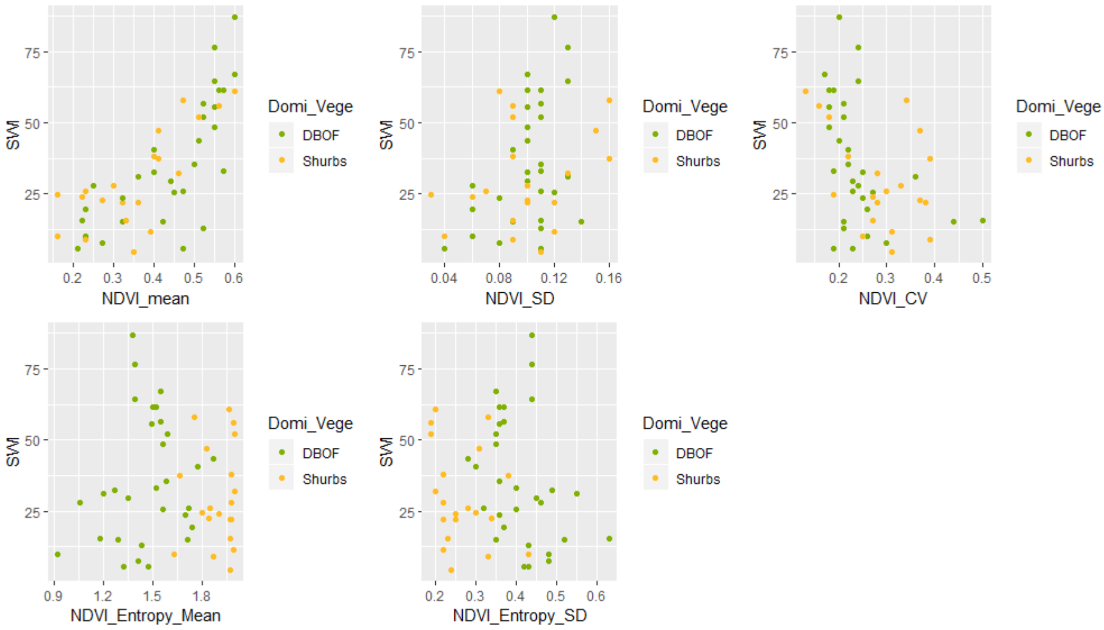


Figure 23. Scatterplot illustrates the relationship between NDVI statistics and SWI, coloured by dominant land cover type to show the variation based on land cover.

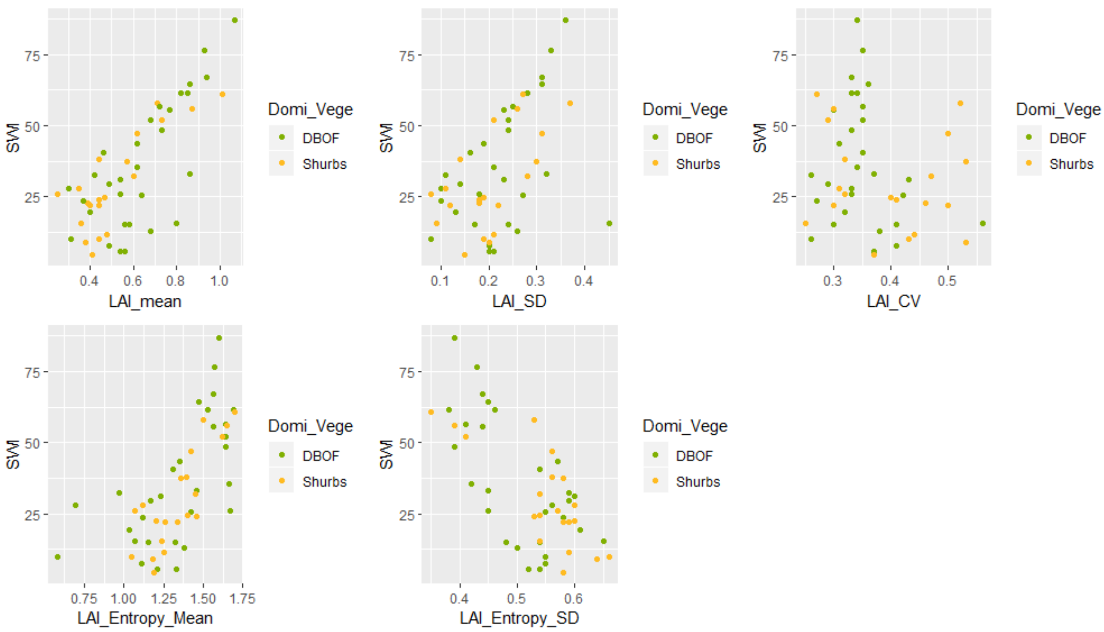


Figure 24. Scatterplot illustrates the relationship between LAI statistics and SWI, coloured by dominant land cover type to show the variation based on land cover.

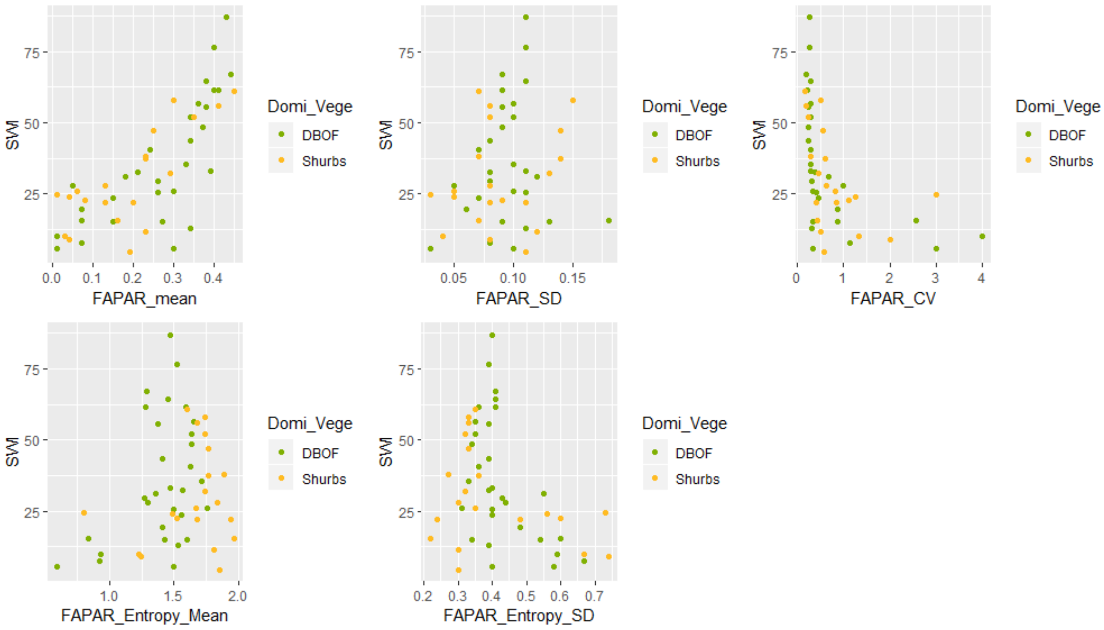


Figure 25. Scatterplot illustrates the relationship between FAPAR statistics and SWI, coloured by dominant land cover type to show the variation based on land cover.

4.3 Extended Sites

After examining the general correlation behaviors of VIs to SWI in the experimental sites, statistics retrieved for the four VIs for the total of 28 sites at the 10 time stamps will be presented here. With the complete set of results, correlations can be analyzed with higher statistical power.

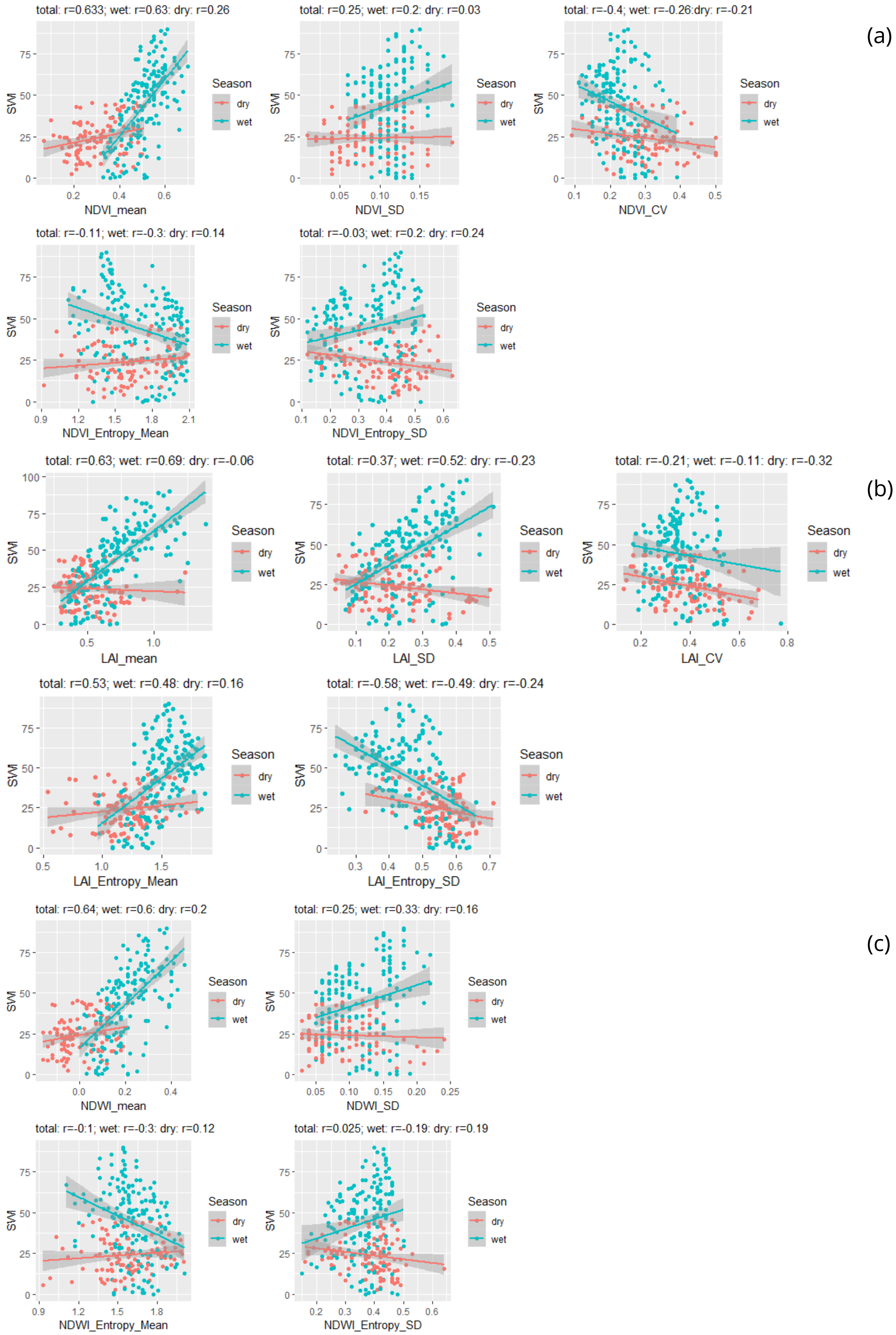
Comparing the correlation strength of the different statistics calculated for VIs to SWI, the mean values showed the strongest correlation (Table 5). The strongest positive linear correlation is from FAPAR, NDWI, NDVI and then LAI ($r > 0.6$). Texture information also contributes to explain the variance in SWI/VI relations. The mean entropy values of LAI show positive linear correlation with SWI at $r = 0.53$; the SDs of entropy for LAI are negatively correlated to SWI at $r = -0.58$. Note that the SD values describing variation of values about the means for FAPAR, NDWI, NDVI, and LAI all indicated weak positive linear correlation with SWI, with r values at 0.13, 0.25, 0.25, 0.37 respectively; the CVs depicting relative variability indicate negative linear correlation at a higher strength—$r = -0.4$ for NDVI SD and $r = -0.21$ for LAI SD. Again, CV comparing to SD, allows comparing variation in different datasets with different means, and able to bring out the variability in shadowed by low SDs. This potentially indicates, while the higher deviation of the individual values about the mean is correlated to higher the SWI, higher relative variability in NDVI and LAI corresponds to lower SWI. Because NDWI and FAPAR have means with 0 or negative values, the CV for NDWI and FAPAR are not valid. Therefore, NA values are entered for the correlation coefficients with SWI. The corresponding P-values indicating the confidence interval of the correlation can also be found in Table 5.

4.3.1 Correlation by Seasons

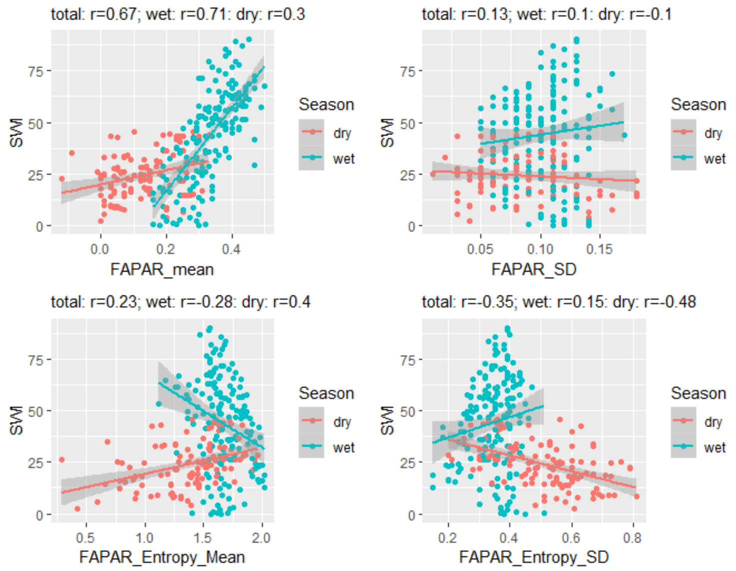
After evaluating the correlation of all VI observations to SWI, the values are grouped into wet season and dry season observations based on the time of the data. The critical statistics of the VIs are plotted in (a) ~ (d) of Figure 26. In terms of CVs, moderate to weak negative correlations are observed for both seasons. But consistent negative linear correlation indicate the negative relations in the relative variation (CV) in LAI and NDVI to SWI.

In terms of the means, in wet season observations, stronger positive correlations can be observed. The correlations for wet seasons are slightly higher than the ones for mixed observations, but generally they are not very distinct. However, it is obvious to see the dry season observations show little positive linear correlation to SWI; in the case of LAI, slight negative correlation can be observed but the correlation is not significant. This indicates that the mean values of four VIs are positively correlated to SWI in the wet season comparing to the dry season, meaning dry season vegetation conditions vary greatly over these sample sites, and other statistics other than mean might capture the relation better.

In the SD plots, moderate positive correlation can be observed for wet season observations (LAI SD at $r = 0.52$) and weak negative correlation are observed for the dry season ones. Similar patterns also appear in the entropy SD plots, but with moderately negative correlation for dry seasons, especially the FAPAR entropy SD ($r = -0.48$). This indicates there is a difference between the variations of vegetation condition in different seasons and how the variation correlates to soil humidity condition. For instance, in wet seasons, the higher the variation in vegetation’s leaf surface area or area to conduct photosynthetic activities, the higher the SWI; the higher variation in FAPAR entropy about the mean in the dry season, meaning the more variate the canopy’s absorption of photosynthetic energy, the lower the SWI.



(a)
total: r=0.633; wet: r=0.63: dry: r=0.26
NDVI_mean
total: r=0.25; wet: r=0.2: dry: r=0.03
NDVI_SD
total: r=-0.4; wet: r=-0.26:dry: r=-0.21
NDVI_CV
total: r=-0.11; wet: r=-0.3: dry: r=0.14
NDVI_Entropy_Mean
total: r=-0.03; wet: r=0.2: dry: r=0.24
NDVI_Entropy_SD
(b)
total: r=0.63; wet: r=0.69: dry: r=-0.06
LAI_mean
total: r=0.37; wet: r=0.52: dry: r=-0.23
LAI_SD
total: r=-0.21; wet: r=-0.11: dry: r=-0.32
LAI_CV
total: r=0.53; wet: r=0.48: dry: r=0.16
LAI_Entropy_Mean
total: r=-0.58; wet: r=-0.49: dry: r=-0.24
LAI_Entropy_SD
(c)
total: r=0.64; wet: r=0.6: dry: r=0.2
NDWI_mean
total: r=0.25; wet: r=0.33: dry: r=0.16
NDWI_SD
total: r=-0:1; wet: r=-0:3: dry: r=0.12
NDWI_Entropy_Mean
total: r=0.025; wet: r=-0.19: dry: r=0.19
NDWI_Entropy_SD
SMI
Season
dry
wet

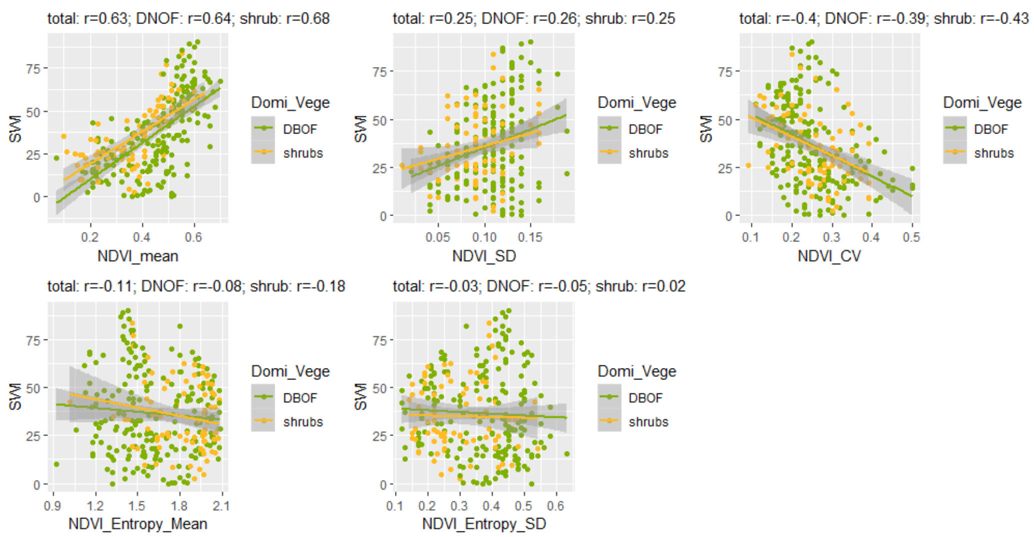


(d)

Figure 26. Scatterplots (a) (b) (c) (d) displays the key statistics of four Vis correlated to SWI grouped by seasons. Grey lines display confidence interval of 0.95.

4.3.2 Correlation by Dominant Vegetation Type

Another way to group the observations is by the dominant vegetation type in each site (Figure 27 (a) ~ (d)). This allows further examination of the differences in VI/SWI correlation by vegetation type. In terms of mean correlation plots, the correlation strength for DBOF and shrubs do not show drastic differences from the mixed observations. Except for LAI, shrubs show a slightly stronger positive correlation than DBOF to SWI. But again, the differences are not drastic, it can be concluded the mean of all VIs show a relatively strong positive correlation to SWI. Similar to the means, other statistics of VIs generally display little differences in correlation strength between the two dominant vegetation types. In terms of SDs, shrubs show a higher positive correlation to SWI than DBOF, indicating a stronger relation between the variations in values of shrub's VIs to SWI than DBOF's.



(a)

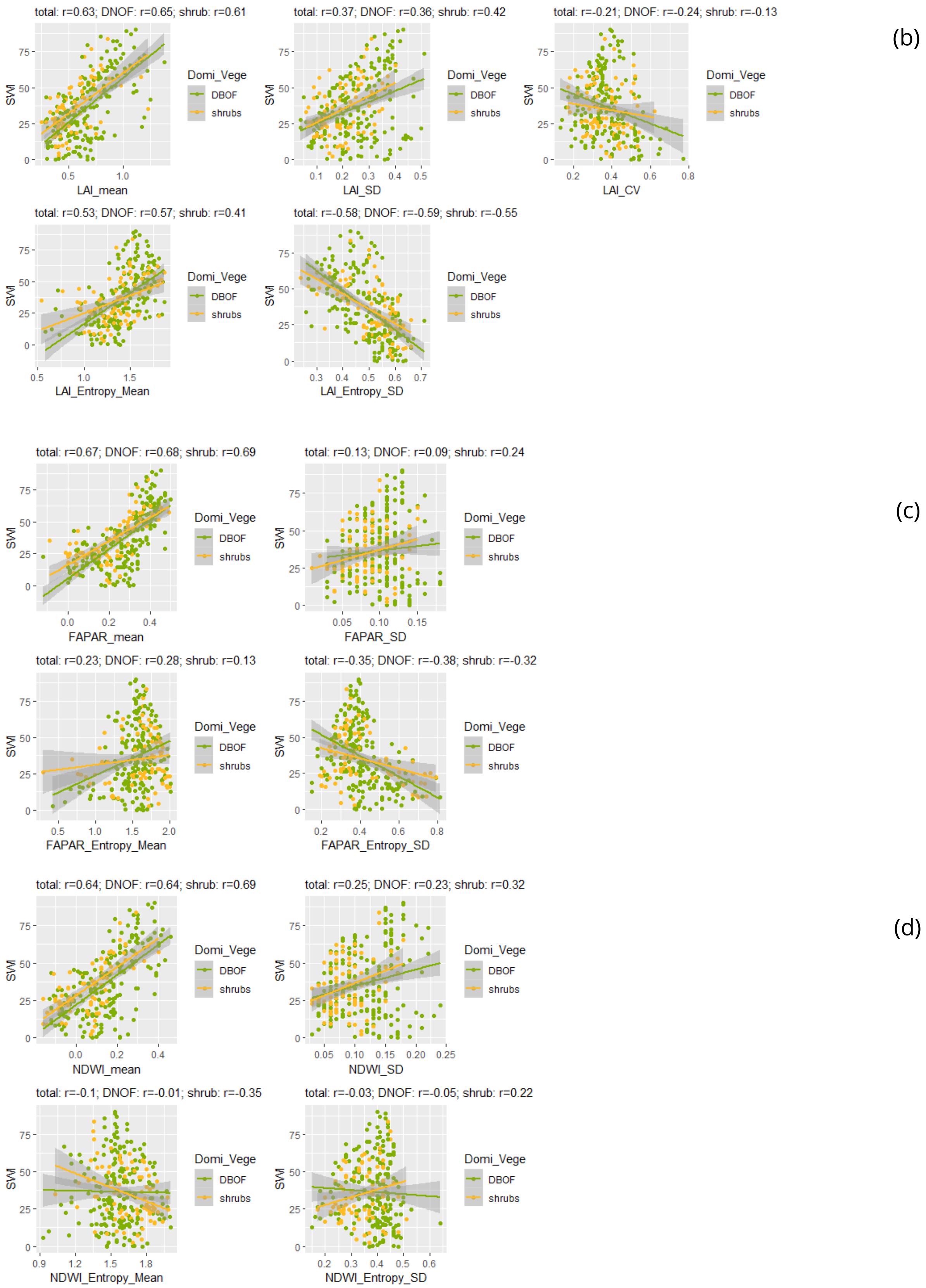


Figure 27. Scatterplots (a) (b) (c) (d) displays the key statistics of four VIs correlated to SWI grouped by dominant vegetation type. Grey lines display confidence interval of 0.95.

VI_Stats	r	P-value	Wet season		Dry season		DBOF		Shrubs	
			r	P-value	r	P-value	r	P-value	r	P-value
ndwi_mean	0.64	1.02E-33	0.60	1.11E-17	0.20	0.031546	0.64	1.42E-24	0.69	1.48E-12
ndwi_min	-0.31	1.92E-07	-0.30	8.94E-05	-0.17	0.07407	-0.30	1.75E-05	-0.33	0.002506
ndwi_max	0.45	2.37E-15	0.33	1.31E-05	0.16	0.082797	0.43	1.23E-10	0.54	2.55E-07
ndwi_sd	0.25	2.69E-05	0.25	0.001062	-0.05	0.58803	0.23	0.000983	0.32	0.00364
ndwi_cv	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
ndwi_ent_mean	-0.10	0.085639	-0.30	7.59E-05	0.12	0.222364	-0.01	0.836309	-0.35	0.001553
ndwi_ent_sd	0.03	0.678539	0.19	0.016004	-0.19	0.050311	-0.05	0.502817	0.22	0.049997
ndwi_hom_mean	0.15	0.011219	0.35	3.26E-06	-0.10	0.305072	0.07	0.325337	0.37	0.000659
ndwi_hom_sd	-0.11	0.078485	-0.13	0.105493	-0.19	0.045233	-0.12	0.096638	-0.11	0.346744
ndvi_mean	0.63	1.04E-32	0.63	5.27E-20	0.26	0.004892	0.64	1.73E-24	0.68	4.64E-12
ndvi_min	-0.45	5.18E-15	-0.46	3.07E-10	-0.23	0.016128	-0.45	4.01E-11	-0.50	2.93E-06
ndvi_max	0.43	2.9E-14	0.41	2.35E-08	0.20	0.035915	0.49	1.66E-13	0.32	0.003465
ndvi_sd	0.25	1.67E-05	0.20	0.010965	0.03	0.771316	0.26	0.00022	0.25	0.028115
ndvi_cv	-0.40	6.86E-12	-0.26	0.000648	-0.21	0.026494	-0.39	1.7E-08	-0.43	8.24E-05
ndvi_ent_mean	-0.11	0.076684	-0.30	5.94E-05	0.14	0.146985	-0.08	0.252808	-0.18	0.117745
ndvi_ent_sd	-0.03	0.614426	0.20	0.008361	-0.24	0.009973	-0.05	0.500267	-0.02	0.849776
ndvi_hom_mean	0.14	0.016998	0.33	1.44E-05	-0.13	0.167506	0.11	0.104969	0.22	0.04778
ndvi_hom_sd	-0.39	1.46E-11	-0.30	9.39E-05	-0.35	0.000148	-0.36	1.09E-07	-0.49	3.97E-06
fapar_mean	0.67	3.19E-37	0.71	1.37E-26	0.30	0.001237	0.68	3.62E-28	0.69	1.32E-12
fapar_min	0.15	0.014925	-0.11	0.172586	0.12	0.193451	0.18	0.012705	0.07	0.563563
fapar_max	0.40	4.52E-12	0.35	4.12E-06	0.05	0.630362	0.39	7.27E-09	0.43	6.66E-05
fapar_sd	0.17	0.034989	0.10	0.190869	-0.10	0.308832	0.09	0.223336	0.24	0.033731
fapar_cv	NA	NA	-0.49	2.36E-11	NA	NA	NA	NA	NA	NA
fapar_ent_mean	0.23	0.000141	-0.28	0.000306	0.40	1.38E-05	0.28	6.34E-05	0.13	0.245555
fapar_ent_sd	-0.35	2.18E-09	0.15	0.051419	-0.48	6.26E-08	-0.38	3.85E-08	-0.32	0.003295
fapar_hom_mean	-0.11	0.0568	0.32	2.36E-05	-0.35	0.000181	-0.18	0.012473	0.01	0.963085
fapar_hom_sd	-0.38	3.49E-11	-0.19	0.015037	-0.35	0.000176	-0.36	2.11E-07	-0.47	9.49E-06
lai_mean	0.63	1.84E-32	0.69	1.99E-25	-0.06	0.542757	0.65	2.3E-25	0.61	1.82E-09
lai_min	0.15	0.013705	0.28	0.000194	0.09	0.341703	0.17	0.013279	0.08	0.458904
lai_max	0.45	1.52E-15	0.46	2.99E-10	-0.08	0.394793	0.42	8.38E-10	0.60	3.23E-09
lai_sd	0.37	1.71E-10	0.52	6.29E-13	-0.23	0.014124	0.36	1.56E-07	0.42	0.000104
lai_cv	-0.21	0.000336	-0.11	0.147352	-0.32	0.000681	-0.24	0.000537	-0.13	0.267392
lai_ent_mean	0.53	1.43E-21	0.48	4.35E-11	0.16	0.085332	0.57	1.18E-18	0.41	0.000142
lai_ent_sd	-0.58	2.02E-26	-0.49	9.26E-12	-0.24	0.010575	-0.59	3.58E-20	-0.55	1.06E-07
lai_hom_mean	-0.50	2.43E-19	-0.42	1.57E-08	-0.17	0.075967	-0.55	5.2E-17	-0.38	0.000497
lai_hom_sd	-0.37	1.95E-10	-0.47	1.80E-10	0.01	0.91469	-0.33	1.84E-06	-0.49	3.50E-06

Table 5. Summarizing correlation coefficients between each VI statistic and SWI for all results together, dry/wet season separated, and DBOF/shrub separated.

4.4 Regression Results

With the correlation between the individual VI statistics and SWI analyzed, a further step to integrate the different aspects of the VIs and relate them to SWI is with multiple linear regression models. Therefore, the representative statistics from three characteristic statistics group are entered as independent variables or predictors, and the SWI observations at the corresponding time are the dependent variable or the outcome. Values retrieved for both the experimental sites and the extended sample sites are included to make the regression model more robust.

Before running the regression analysis, some variables are excluded. First, the CVs for NDWI and FAPAR are excluded because they contain undefined values. Second, min, max, and median values are not included because they are derived from single values in the sample site, thus, they are subject to bias and outliers. Third, based on the P-values in Table 5, variables with no statistically significant correlation to SWI which are the ones with P-value ≥ 0.05 are excluded. As discussed in the methodology, the statistics calculated for each VI are assigned to three groups: central discrepancy descriptor, spatial heterogeneity descriptors and texture descriptors. One statistic of each group is selected based on its correlation strength listed above (Table 5) when conducting the multiple regression analysis.

4.4.1 Step-wise Regression

With the deduction procedures discussed above, 15 variables are left. The test regression model with the 15 variables yield the following results (Table 6). The test model is only used as a starting point or as the input model for the further variable selection procedures.

Model ID	**R^2**	**Adjusted R^2**	**Residual standard error**	**P-value**	**# of Independent Variables**	**Model**
Test	0.62	0.60	13.25	< 2.2e-16	15	SWI= - 21.64 - 20.08×ndwi_mean + 145.61×ndwi_sd + 9.39×ndwi_hom_mean - 49.05×ndvi_mean + 96.78×ndvi_cv + 19.75×ndvi_hom_mean - 234.02×ndvi_hom_sd + 56.01× lai_mean - 19.6× lai_sd + 12.4×lai_ent_mean + 11.52×lai_ent_sd + 120.29×fapar_mean - 397.18×fapar_sd + 8.18×fapar_ent_mean - 22.47× fapar_hom_sd

Table 6. Test model statistics.

With this test model, stepwise variable selection by optimizing the models' overall AIC is performed to generate a new model with a subset of independent variables (Table 7). The diagnostic plots (Figure 28) indicate that the model's residuals are generally evenly spread around the horizontal line without a distinct pattern in the "residual vs Fitted" plot;

the QQ plot shows pretty good alignment of the residuals to the line with a few points at the top slightly offset; in the “Scale-Location” plot, the residuals are more randomly spread above and below and along the line; lastly, the “Residual vs. Leverage” plot also does not mark any influential cases that failed to be included or reduced.

Model ID	R^2	Adjusted R^2	Residual standard error	P-value	# of Inde-pendent variables	Model
Step-AIC	0.61	0.60	13.20	< 2.2e-16	8	SWI=1.29 + 107.94×ndwi_sd + 127.37×ndvi_cv + 15.45×ndvi_hom_mean - 263.49×ndvi_hom_sd + 38.51×lai_mean + 8.02× lai_ent_mean + 94.77× fapar_mean - 43.712× fapar_sd

Table 7. StepAIC model statistics.

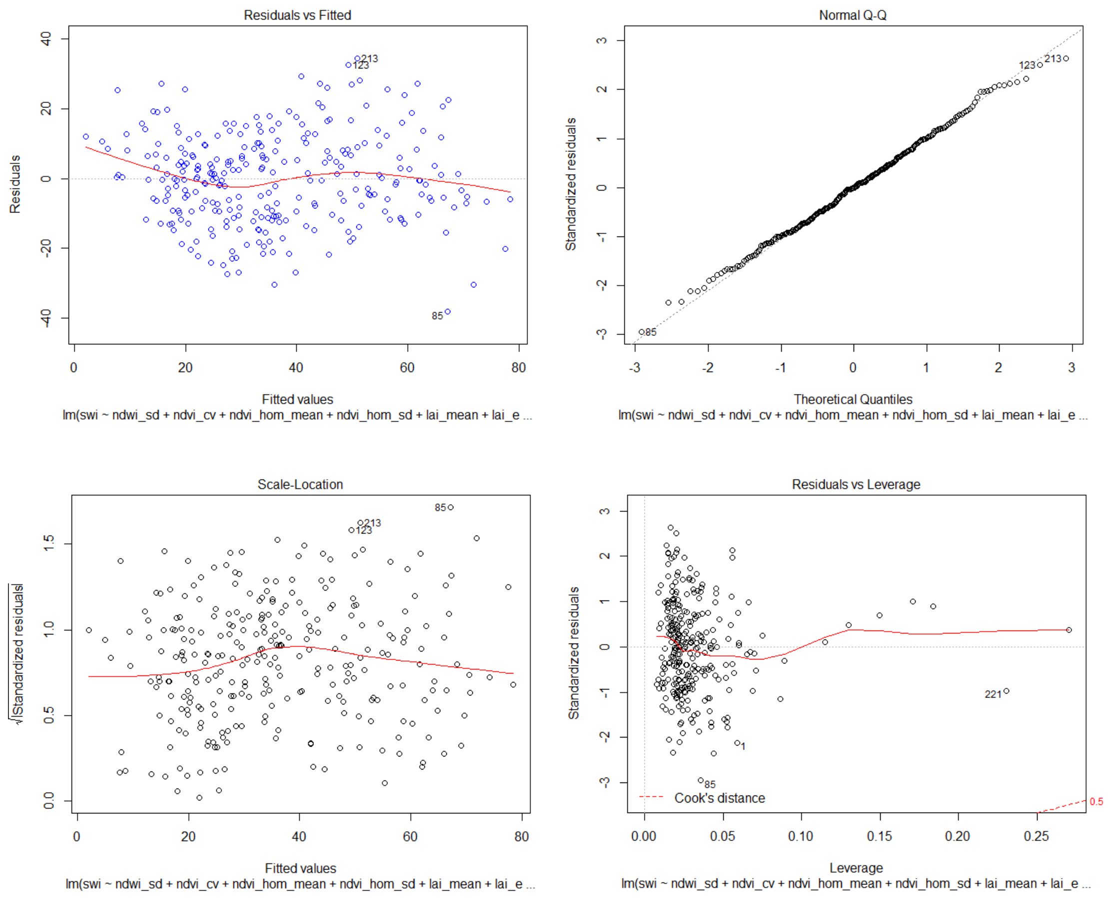


Figure 28. Diagnostic plots for stepAIC model in Table 7.

4.4.2 Subset Regression

With the stepAIC model shown above, the subset regression is used to show how the subsets of predictors influence the model's performance in terms of some defined objective criterion, such as R^2, adjusted R^2, mallow's Cp, AIC and Mean Squared Error of Prediction (MSEP). In Table 8, the 5th model with 5 predictors can explain 60% of the variance in the SWI. These predictors or independent variables are related to the central tendency of FAPAR and LAI, spatial heterogeneity of the sample site's FAPAR and NDVI, as well as the deviation in the textural orderliness of NDVI. The most prominent contributions to explain SWI variance are from the FAPAR and LAI mean. Lastly, inclusion of textural information of NDVI and LAI aids the explained variance but the models are penalized for including additional variables as indicated by the adjusted R^2 (Figure 29).

Subset Model Index	**Predictors**	**R^2**	**MSEP**
1	fapar_mean	0.44	248.10
2	lai_mean, fapar_mean	0.50	225.63
3	ndvi_hom_sd, lai_mean, fapar_mean	0.54	208.25
4	ndvi_cv, lai_mean, fapar_mean, fapar_sd	0.57	194.52
5	ndvi_cv, ndvi_hom_sd, lai_mean, fapar_mean, fapar_sd	0.60	184.25
6	ndwi_sd, ndvi_cv, ndvi_hom_sd, lai_mean, fapar_mean, fapar_sd	0.61	179.83
7	ndwi_sd, ndvi_cv, ndvi_hom_mean ndvi_hom_sd lai_mean fapar_mean fapar_sd	0.61	180.30
8	ndwi_sd, ndvi_cv, ndvi_hom_mean,ndvi_hom_sd,lai_mean,lai_ent_mean,fapar_mean fapar_sd	0.62	180.04

Table 8. Subset models and participating predictors.

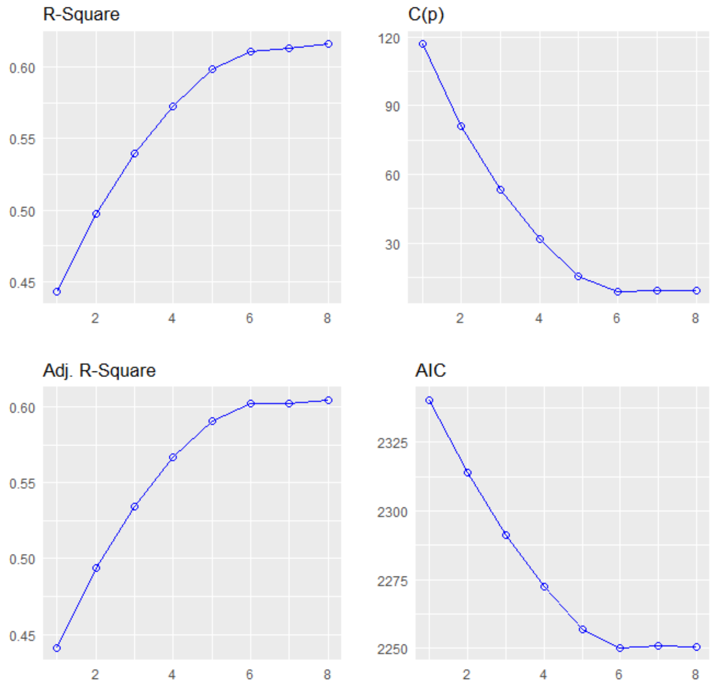


Figure 29. Four common metrics for each subset model.

4.4.3 Seasonal Difference in Regression

In reference to the regression results for all the 280 observations, below demonstrate how the results differ when grouping the wet and dry season observations separately in regression models. This way, how different characteristics of VIs relate to SWI in different seasons can be compared.

In Table 9, one statistic from each characteristic group for VIs is added to run the test regression models. The test model for the dry season showed very low value in R^2 which indicated not much variance in dry season SWI can be explained with the VI statistics; the P-value is also higher comparing to the mixed model and the wet season model. Relating to the scatterplots in Figure 17~Figure 21, dry season observations for all statistics of all four VIs are very scattered, and the correlations are not significant. This could indicate that the dry season soil moisture condition is more complex, and it cannot be directly linked with average, spatial heterogeneity, textural information of the four VIs examined here. Due to the low significance of the dry season model, further regression analysis is not applied to it.

Model ID	R^2	Adjusted R^2	Residual standard error	P-value	# of qualified predictors	Model
Test-wet	0.63	0.59	14.29	< 2.2e-16	15	SWI= – 55.61 – 34.28×ndwi_mean + 235.18×ndwi_sd + 172.15×ndwi_hom_mean – 157.57× ndwi_ent_sd - 157.99×ndvi_mean – 61.58×ndvi_cv + 19.21× ndvi_hom_mean – 221.79×ndvi_hom_sd + 88.11×lai_mean – 26.39×lai_sd + 35.55× lai_ent_mean + 12.89×lai_ent_sd + 34.06× fapar_mean + 8.3× fapar_hom_mean + 81.74×fapar_hom_sd
Test-dry	0.36	0.30	9.02	7.787e-07	10	SWI= 43.62 +10.31×ndwi_mean + 70.32×ndwi_hom_sd - 53.78×ndvi_mean + 28.81× ndvi_cv - 190.39×ndvi_hom_sd - 28.65× lai_cv - 3.88× lai_ent_sd + 75.64×fapar_mean +7.40 ×fapar_ent_mean - 0.28×fapar_ent_sd

Table 9. Test models statistics for wet vs. dry seasons.

With the test model for wet season (Table 9), the StepAIC method generated the following model (Table 10) to optimize the overall AIC. Comparing to the overall model of both seasons (Table 8), the wet season regression model (Table 10) included 7 instead of 8 variables, and with approximately the same R^2, about 62% variance in SWI can be explained with these two models. Different from the overall model, mean of LAI explained the most variance in this model, and the homogeneity of NDVI also add to the wet season model (Table 11). The bigger distinction is in the wet season model, none of the FAPAR statistics are making significant contribution to explain the SWI variance, yet in the overall model, the mean of FAPAR describing central tendency of the canopy's primary productivity of photosynthesis in each site explained about 44% variance in SWI. In terms of the residual plots of the wet season models (Figure 30), the Normal Q-Q plot shows a slight S-shaped, which could indicate the model is light-tailed, but most the points show good alignment with the central straight line.

Model ID	**R^2**	**Adjusted R^2**	**Residual standard error**	**P-value**	**# of Independent variables**	**Model**
Step-AIC-wet	0.62	0.60	14.04	< 2.2e-16	7	SWI= - 56.34 + 127.89×ndwi_sd + 173.17×ndwi_hom_mean – 119.45×ndwi_ent_sd – 103.81×ndvi_mean – 187.19×ndvi_hom_sd + 74.9×lai_mean + 30.39× lai_ent_mean

Table 10. StepAIC model for wet season observations.

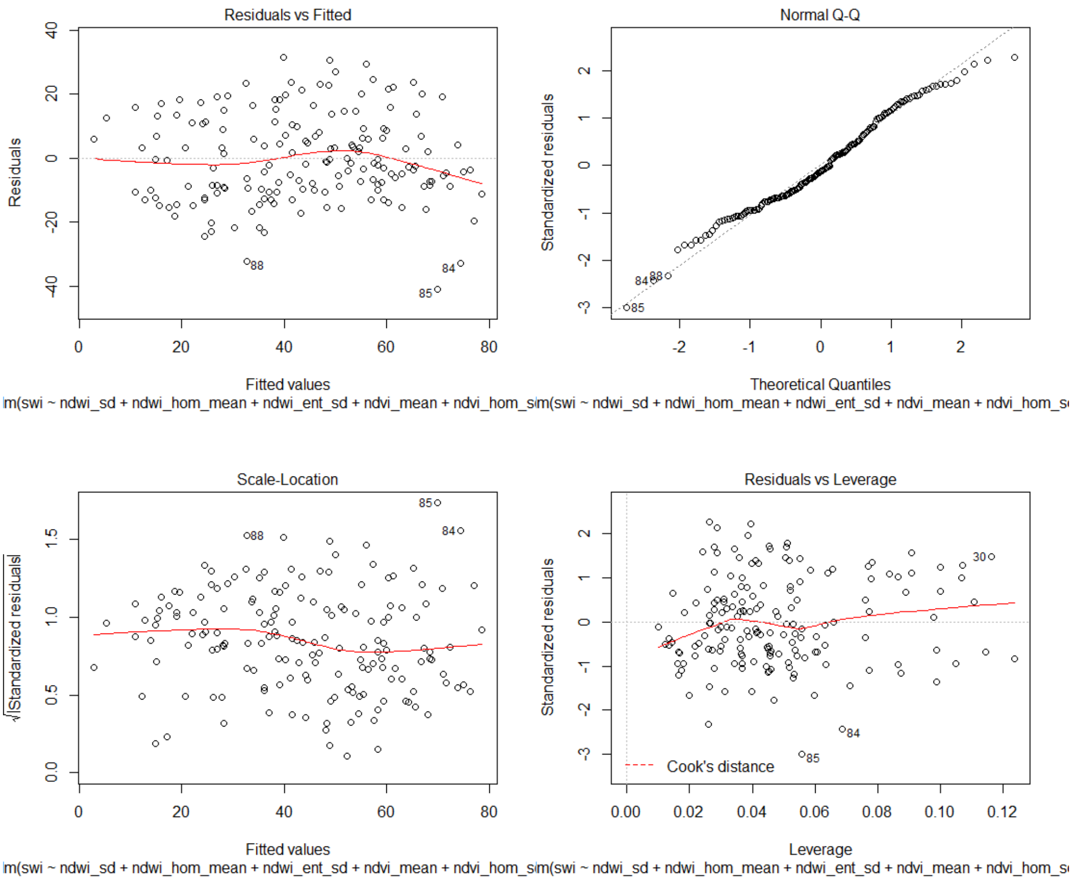


Figure 30. Residual analysis for the wet season stepAIC model.

Subset Model Index	Predictors (wet)	R^2	MSEP
1	lai_mean	0.48	263.31
2	ndvi_hom_sd lai_mean	0.54	236.52
3	ndwi_hom_mean ndwi_ent_sd lai_mean	0.56	228.58
4	ndwi_hom_mean ndvi_mean lai_mean lai_ent_mean	0.59	217.84
5	ndwi_sd ndwi_hom_mean ndwi_ent_sd lai_mean lai_ent_mean	0.60	210.66
6	ndwi_sd ndwi_hom_mean ndwi_ent_sd ndvi_mean lai_mean lai_ent_mean	0.62	207.41
7	ndwi_sd ndwi_hom_mean ndwi_ent_sd ndvi_mean ndvi_hom_sd lai_mean lai_ent_mean	0.62	207.13

Table 11. Subset analysis of the participating predictors for the wet season regression model.

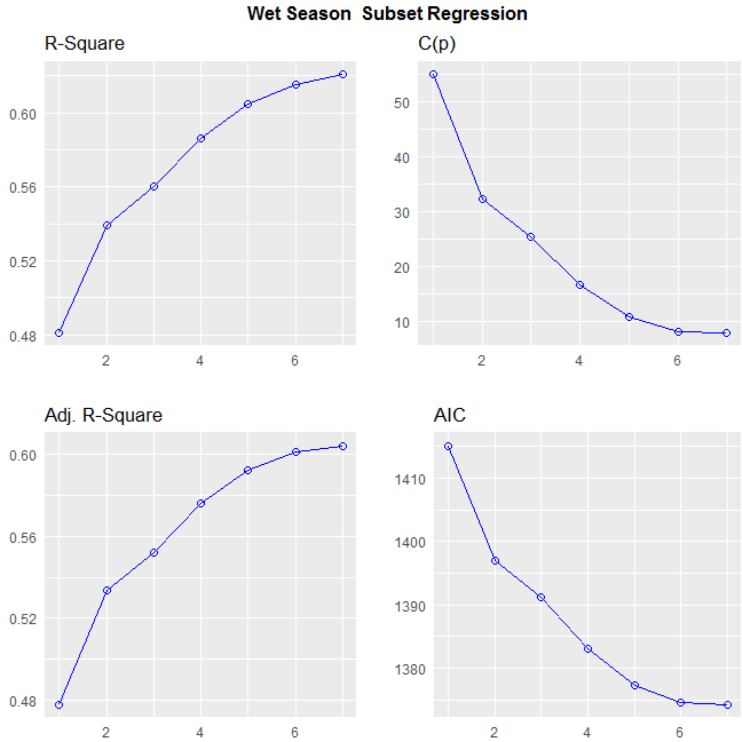


Figure 31. Four common evaluation matrices for each model in the subset regression for wet season observations in Table 11.

4.4.4 Different Dominant Vegetation Types and Regression

Next, the observations are grouped by the dominant vegetation type to examine the relationship between different vegetation information conveyed through the VI statistics and the SWIs at the 28 sample sites. The test models in Table 12 compares the independent variables entered in the regression analysis and the overall performance of the test models. The test model for shrub dominant observations has 2 more qualified independent variables, and the overall R^2 value is higher and the residual error is lower. This model explains 82% variance which is 20% higher than the overall model with all observations. In Table 13, after the automatic variable selection with the stepAIC method, the overall R^2 for both models have stayed the same while the less significant variables are excluded from the models. In the residual analysis for both models, Figure 32 and Figure 33, both models have residuals showing normal distribution; in the Q-Q plots, both align with the 45° line with some offsets at the top and bottom. But in the "Residuals vs Leverage" plot for the DBOF model, one observation's residual is slightly within the Cook's Distance, considering it is one residual from one observation and the sample size, it should not be significant to the overall model; even after removing this one observation, the results remained the same.

Model ID	**R^2**	**Adjusted R^2**	**Residual standard error**	**P-value**	**# of qualified predictors**	**Model**
Test-DBOF	0.61	0.58	14.29	< 2.2e-16	12	SWI= - 11.04 - 55.78× ndwi_mean + 142.41×ndwi_sd - 64.52×ndvi_mean + 56.6× ndvi_cv - 278.76×ndvi_hom_sd + 92.8× lai_mean - 124.48×lai_sd + 2.12×lai_ent_sd +10.68×lai_ent_mean + 128.92× fapar_mean + 7.91×fapar_ent_sd + 4.42×fapar_ent_mean
Test-Shrub	0.82	0.79	8.27	< 2.2e-16	14	SWI= - 4.56 - 18.55×ndwi_mean + 136.28×ndwi_sd + 45.58×ndwi_hom_mean + 248.87×ndvi_mean - 5.34×ndvi_cv + 50.3×ndvi_hom_mean - 248.66×ndvi_hom_sd – 16.2×lai_mean + 74.53× lai_sd - 55.21× lai_ent_sd + 4.33×lai_ent_mean – 156.81×fapar_mean - 312.65× fapar_sd – 49.44× fapar_hom_sd

Table 12. Test models with observations grouped by two dominant vegetation types: DBOF and Shrubs.

Model ID	**R^2**	**Adjusted R^2**	**Residual standard error**	**P-value**	**# of Independent variables**	**Model**
Step-AIC-DBOF	0.61	0.59	14.2	< 2.2e-16	8	SWI= - 19.08 - 62.38× ndwi_mean + 115.84× ndwi_sd + 80.42×ndvi_cv - 271.83× ndvi_hom_sd + 93.98× lai_mean -132.3×lai_sd +11.29× lai_ent_mean + 83.31×fapar_mean
Step-AIC-Shrub	0.82	0.80	8.06	< 2.2e-16	10	SWI=6.81+113.87×ndwi_sd + 49.56×ndwi_hom_mean + 242.58×ndvi_mean + 48.88×ndvi_hom_mean – 275.82×ndvi_hom_sd - 28.53×lai_mean + 85.96×lai_sd – 68.04×lai_ent_sd – 152.72×fapar_mean – 295.69×fapar_sd

Table 13. StepAIC models for DBOF and Shrubs.

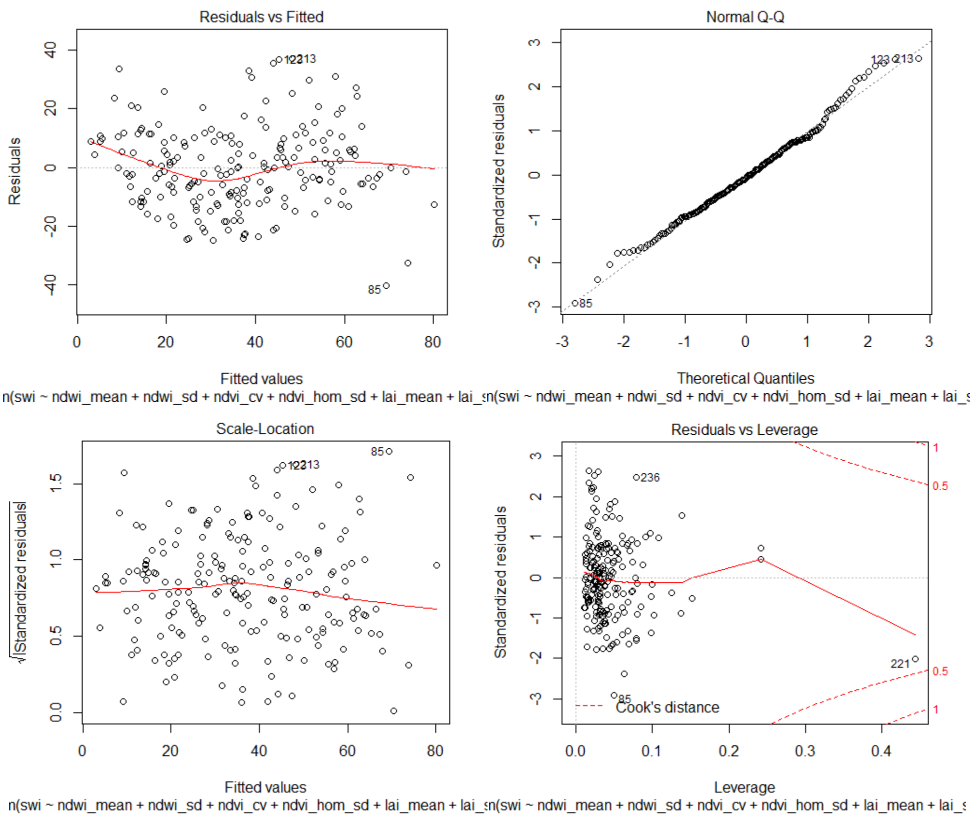


Figure 32. Residual analysis of StepAIC model for DBOF.

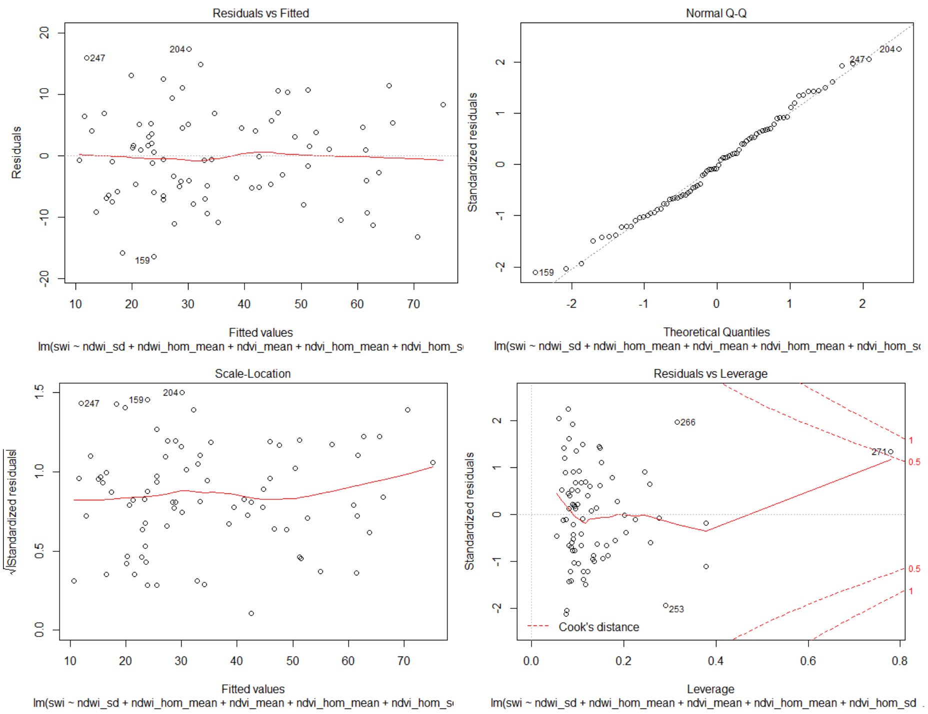


Figure 33. Residual analysis of StepAIC model for Shrubs.

The following stepwise subset regression compares how the addition of different variables influences the variance explained by each subset model. In both models, the FAPAR mean is prominent for explaining about 46 ~ 48% of the variance. The DBOF model shows including of LAI mean and the deviation in NDVI's homogeneity in the 2nd and 3rd steps add to the model, while in the Shrub model, instead of the deviation of NDVI's textural homogeneity, it is firstly the mean of NDVI's second order homogeneity that increases the amount of variance the model explains. But in the 3rd to 6th subset model for the Shrub model, the combination of NDVI mean, NDVI's and LAI's second order homogeneity, and NDWI's and FAPAR's standard deviation increase the R^2 for the subset model, instead of FAPAR mean. One thing to note is the role of NDWI statistics generally are not prominent in these models and in the overall model.

Subset Model Index	**Predictors (DBOF)**	**R^2**	**MSEP**
1	fapar_mean	0.46	269.54
2	lai_mean, fapar_mean	0.51	248.32
3	ndvi_hom_sd, lai_mean, fapar_mean	0.54	230.95
4	ndwi_mean ndvi_hom_sd lai_mean fapar_mean	0.56	227.33
5	ndvi_cv ndvi_hom_sd lai_mean lai_sd fapar_mean	0.59	214.60
6	ndwi_mean ndvi_cv ndvi_hom_sd lai_mean lai_sd fapar_mean	0.59	212.53
7	ndwi_mean ndwi_sd ndvi_cv ndvi_hom_sd lai_mean lai_sd fapar_mean	0.60	212.00
8	ndwi_mean ndwi_sd ndvi_cv ndvi_hom_sd lai_mean lai_sd lai_ent_mean fapar_mean	0.60	211.31

Table 14. Subset models and participating predictors for DBOF observations.

Subset Model Index	**Predictors (Shrubs)**	**R^2**	**MSEP**
1	fapar_mean	0.48	174.01
2	ndvi_hom_mean fapar_mean	0.66	116.44
3	ndvi_mean ndvi_hom_mean lai_ent_sd	0.74	91.80
4	ndvi_mean ndvi_hom_mean ndvi_hom_sd lai_ent_sd	0.78	81.06
5	ndwi_sd ndvi_mean ndvi_hom_mean ndvi_hom_sd lai_ent_sd	0.78	81.60
6	ndwi_sd ndvi_mean ndvi_hom_mean ndvi_hom_sd lai_ent_sd fapar_sd	0.80	77.17
7	ndwi_sd ndvi_mean ndvi_hom_mean ndvi_hom_sd lai_ent_sd fapar_mean fapar_sd	0.80	77.25
8	ndwi_sd ndvi_mean ndvi_hom_mean ndvi_hom_sd lai_sd lai_ent_sd fapar_mean fapar_sd	0.81	76.47
9	ndwi_hom_mean,ndvi_mean ndvi_hom_mean ndvi_hom_sd lai_mean lai_sd lai_ent_sd fapar_mean fapar_sd	0.82	76.22
10	ndwi_sd ndwi_hom_mean ndvi_mean ndvi_hom_mean ndvi_hom_sd lai_mean lai_sd lai_ent_sd fapar_mean fapar_sd	0.82	75.51

Table 15. Subset models and participating predictors for Shrub observations.

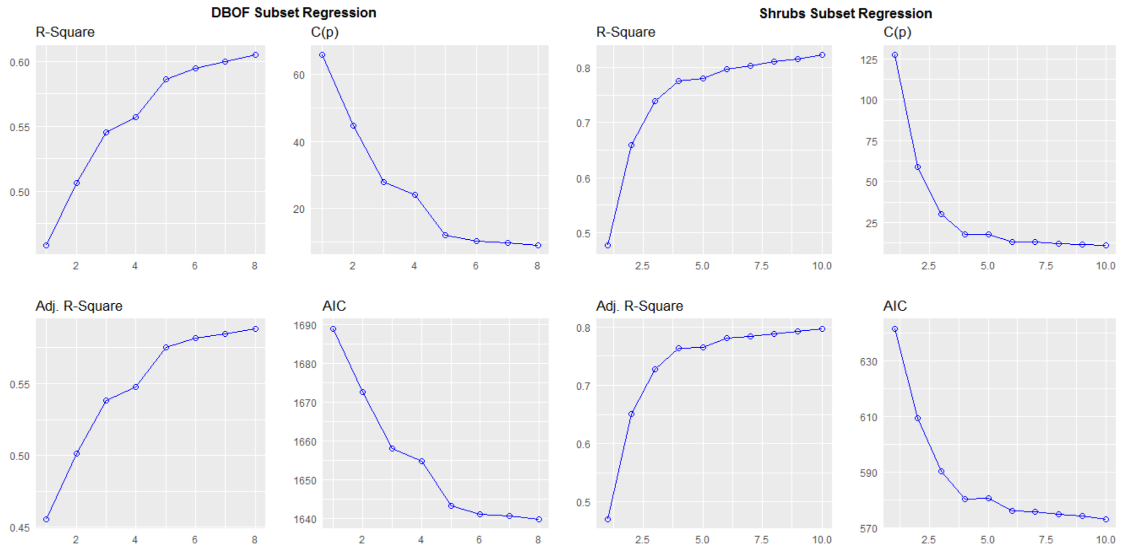


Figure 34. Four common evaluation matrices for each model in the subset regression for DBOF (left) and Shrub (right) dominant observations

4.5 Cartographic Visualization and Implementation

The interactive map platform named "SoilWater[3]" is essentially an interactive map embedded in a webpage. The webpage displays the contextual information including the interaction instructions, data sources, and general information related to the map. The link to the webpage is:

https://mliang8.github.io/SoilWaterCube

4.5.1 Basemaps and Overlays

Two standard raster tilesets are used as basemaps to provide users with geographic context. A simple vector street map and a satellite basemap are available to switch between depending on the contextual background the user preferred.

The overlays or the thematic layers can be toggled on and off allowing users to explore the main variables included in the remote sensing analysis (Figure 35). Some overlays display additional context of the research study, such as the layer "Delta extent". Other layers "Land cover" and "Soil water index" present the key variables accompanied by corresponding legends allowing users to match the color-coded graphics with the data values (Figure 36). The four overlays of correlations show the spatially interpolated correlation strength between each analyzed VI and SWI (Figure 37). These four Kriging interpolation results are not a part of the remote sensing analysis but are created specifically for the cartographic visualization by taking advantage of the concept of “generalization” in cartography and enabling users to establish a more complete understanding about spatial heterogeneity in the Delta. The “Vegetation indices” layer demonstrates the workflow of retrieving VIs for experimental sites.

Moreover, besides displaying each thematic dataset, the overlays also initiate or activate the interactive elements of the map. Toggling on the “Soil water index” layer initiates the slider bar for adjusting the time slice of data displayed on the map. The “Vegetation indices” activates the experimental site graphics’ clicking ability, which in turn initiates the VI panel and temporal slider for comparing all four VIs over time.

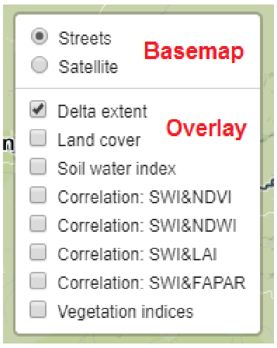


Figure 35. Map basemaps and overlays.

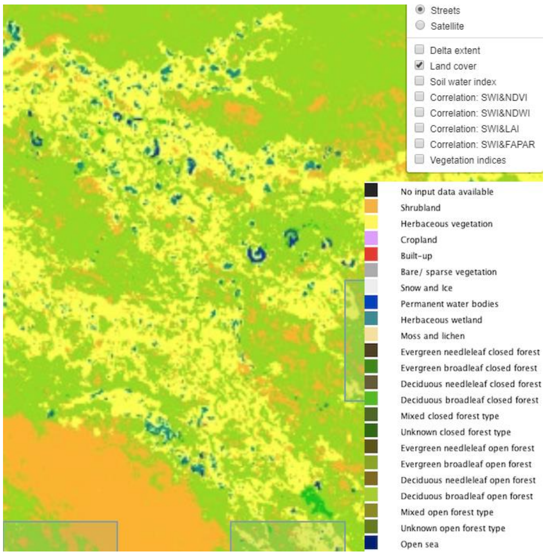


Figure 36. Examples of overlay and matching legend.

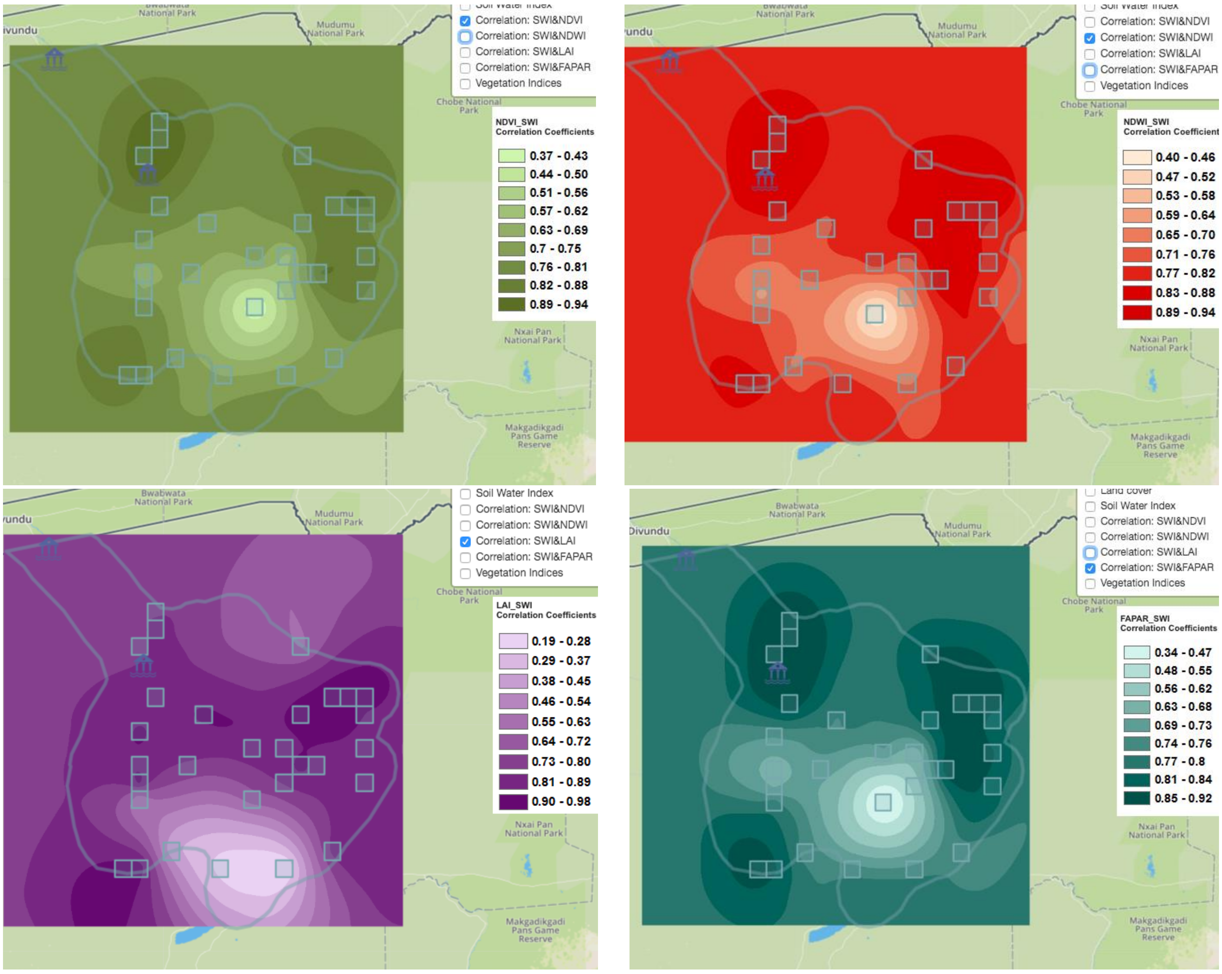


Figure 37. Kriging interpolation of the correlation strength between SWI and each VI at 28 sample sites for the entire Delta. Correlation coefficient is generated from correlating Vis to SWI at each site over 10 dates.

4.5.2 Other Interactive Elements

There are other interactive elements on the map that provide background or contextual information about Okavango Delta. The in-situ water stations icons are always displayed; upon clicking, popups appear on the map showing element name, and an additional UI appears to the left on the webpage allowing further interaction to explore Okavango water level information (Figure 38). The square icons showing the spatial extents of the sample sites are also always displayed on the map. These icons can be clicked to access the pie chart summarizing the clicked sample site's land cover composition (Figure 39). Last but not least, the VI panel is also an interactive element that is initiated when the "Vegetation indices" layer is activated and when one of the highlighted sample sites is clicked. This series of interactions help users to understand the steps of data retrieval (Figure 12). Toggling on/off the overlays generally does not interfere with the intractability of these interactive elements mentioned above, thus, it allows concurrent data display and exploration; excessive overlapping is checked and avoided during implementation to eliminate occlusion of information.

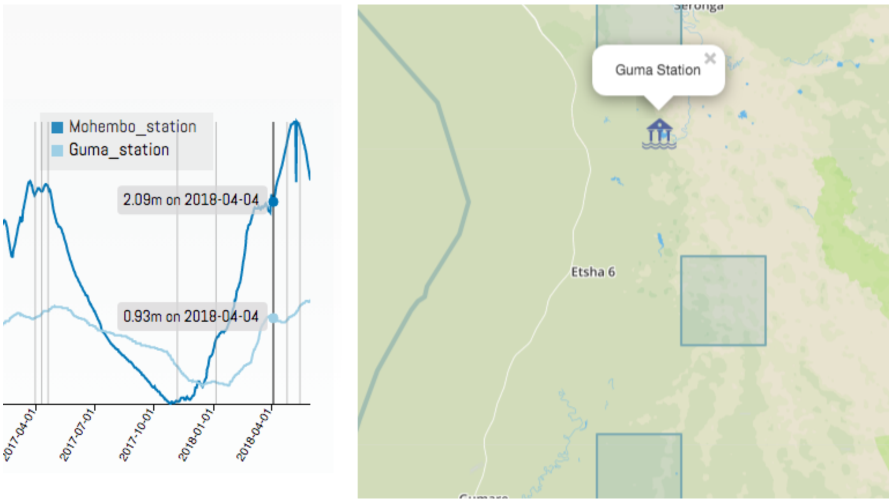


Figure 38. In-situ station icons and their interactions.

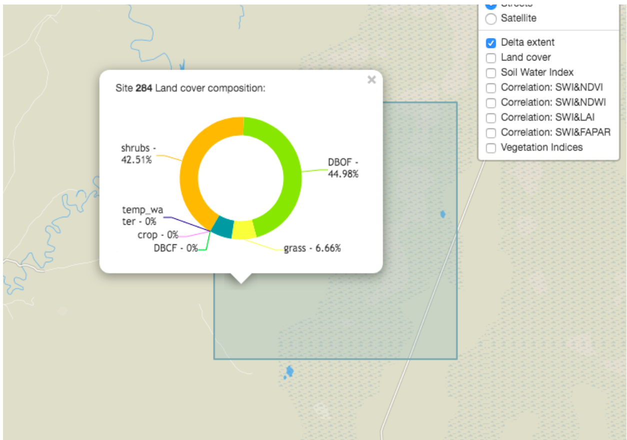


Figure 39. Pie chart display and cover composition display on the click of site icon.

4.5.3 Time Dimension

The time dimension of this visualization reflecting the change of SWI and VIs in real life are realized through temporal sliders. Upon moving the slider handles, the texts appear to indicate the corresponding time stamp, and the map overlay changes for the "Soil water index" layer, while the VI panel updates for the "Vegetation indices" Layer (Figure 40). Both temporal sliders can be displayed together to show concurrent change of SWI and VIs, as long as both layers are activated (Figure 41). This way, users have more freedom to explore and examine the corresponding changes of both variables together.

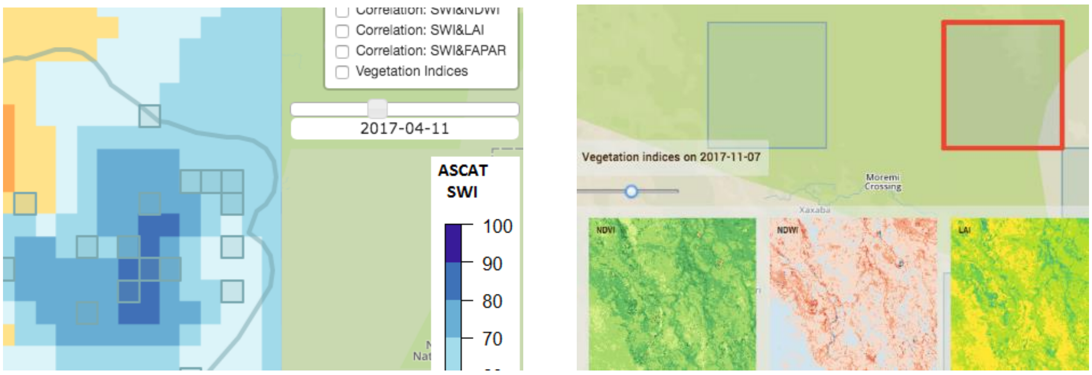


Figure 40. Time slider for SWI (left) and for VI panel (right).

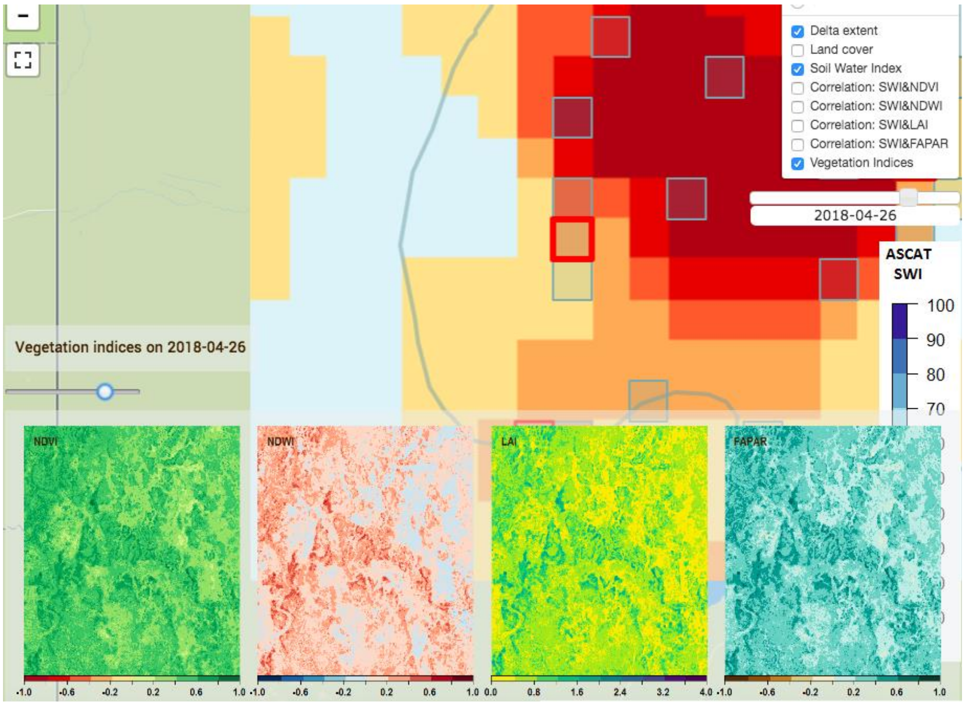


Figure 41. Concurrent display of two overlays enables two time sliders to be manipulated together.

4.5.4 GitHub Repository

Lastly, the finished visualization product is published by taking advantage of GitHub's static web hosting service (Figure 42). All source codes, graphs, and content used in the mapping platform are stored in the public repository under *https://github.com/mliang8/SoilWaterCube* (Figure 43). All content is available to be browsed and downloaded.

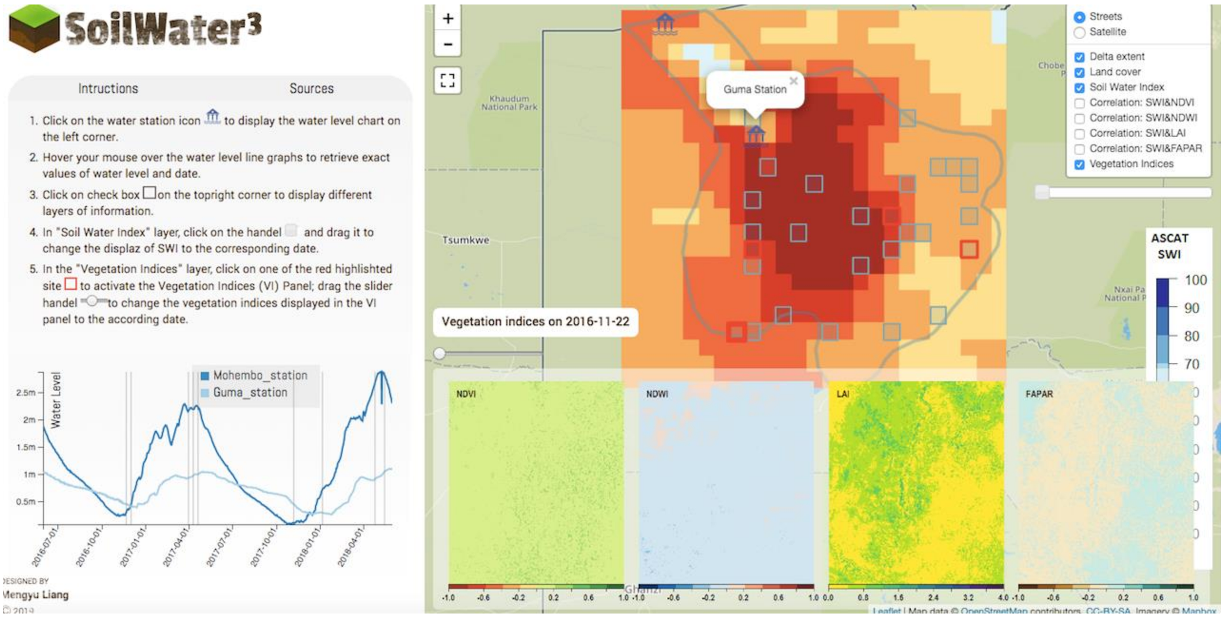


Figure 42. One view of the complete SoilWater³ product.

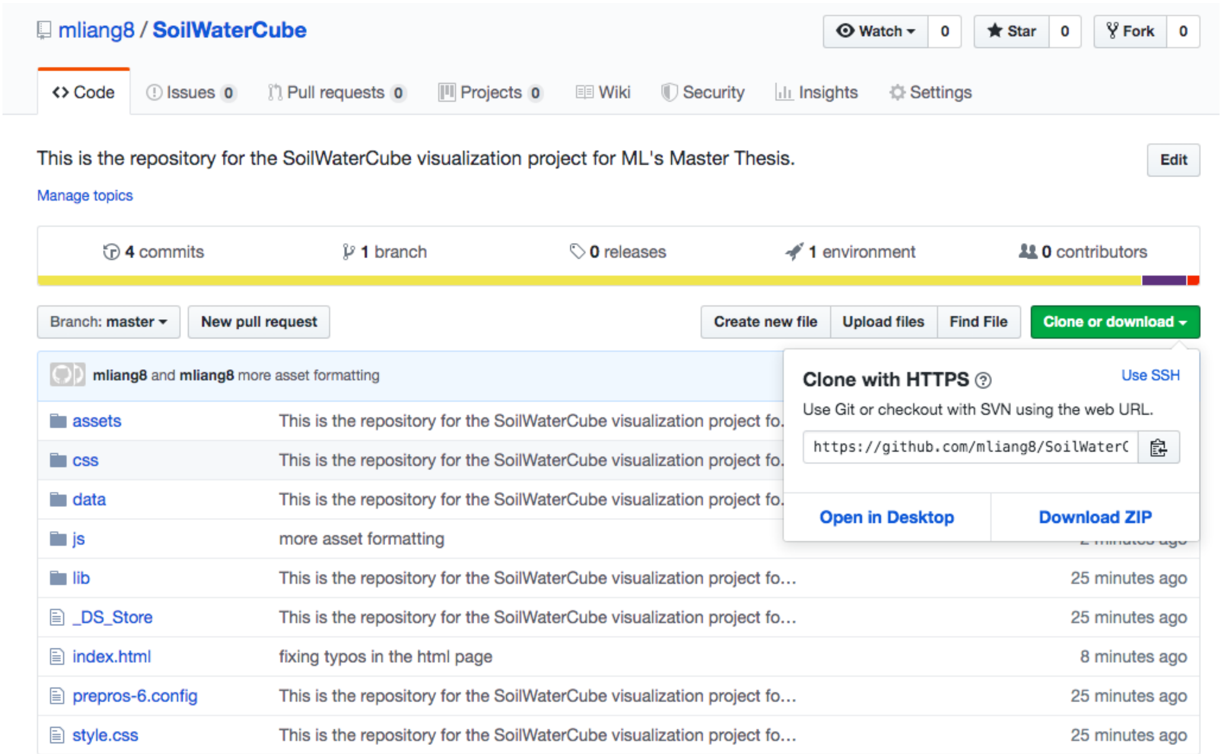


Figure 43. GitHub Repository for the project.

5 Discussion

In this chapter, the results from the analysis will be discussed in relation to the ecological significance and practical implication; the limitation and issues, as well as possible alternative approaches related to the research, will also be examined.

5.1 Summary of key findings

In this research, correlations and regression methods have demonstrated the possibility of using VIs to estimate soil water conditions in the wetland environment. Furthermore, it indicates that optical remote sensing can add information about soil moisture in a finer spatial resolution. Seasonal differences in using vegetation proxies for soil moisture are obvious – in the wet season, vegetation information has a strong linkage to soil moisture while very scattered results are observed for the dry season. Differences in using remote sensing VIs to understand soil moisture also exist for areas with different dominant vegetation but not drastic – in sites with shrubs as dominant vegetation, vegetation proxies performed generally well in estimating soil moisture, and in sites with DBOF, a moderately strong correlation could be found as well. In the visualization product development, several key interactive strategies for visualizing multidimensional spatial-temporal data are adopted allowing the targeted users to explore the data used in this research and to develop visual thinking about the research workflow.

In the subset regression (Table 8, Table 11, Table 14, Table 15), the mean values of FAPAR explained the most variance observed in SWI (approximately 45%), indicating the central tendency in the vegetation's evapotranspiration and photosynthetic primary production capacity are well linked to soil moisture. In the regression model for wet season observation (Table 11), LAI alone explained 48% of the variance while FAPAR mean was not selected in the stepAIC model because of its low correlation strength and significance. Even though the LAI and FAPAR are usually companied biophysical variables that characterize the canopy and photosynthetic activity of plants, in the wet season, LAI, accounting for the amount of foliage in the plant canopy including the understory, has a stronger correlation to SWI. Vegetation's vitality and greenness conveyed through the NDVI also help to explain the variance in SWI but only show weak correlation through its second order texture measures. NDWI, which measures the liquid water content in vegetation and is an important drought indicator, does not contribute greatly in understanding soil moisture variance; this could also be observed in Figure 16, after the peaks, the NDWI decreases slower than the SWI, which could indicate after the high water saturates the plants, the vegetation's water content stays high for a period of time instead of immediately showing a response with the loss of water in soil.

Besides the central tendency described by the means, the texture features included here, the homogeneity and entropy, contributed weakly to moderately to the explanation of SWI variance. The second order texture measures describe the vegetation structure through the tonal relationship of neighboring pixel values, thus, they can describe vegetation's horizontal structure. They are mathematically more complex than the first order statistics discussed above but have advantages in explaining the neighboring pixel values' relations. In the mixed, wet and DBOF regression models, the second order homogeneity deviation for NDVI explained 3% ~ 6% variance, but in the shrub dominate observation model, the second order homogeneity mean for NDVI explained 18% variance. Shrubs comparing to DBOF are slightly more homogenous in vegetation structure. This may be explained as that the variance in SWI relates to NDVI's texture information more when the site has more a homogenous vegetation condition.

The other group of statistic characteristics included in the research is the SD and CV for interpreting first order spatial variation in the four VIs. In the subset regression analysis, these two variables have weak influence in the model's overall explanation of SWI variance, but in the correlation analysis of each variable to the SWI, they show a moderate correlation to SWI (Table 5), with CV generally negatively correlated to SWI and SD positively correlated to SWI. This indicates that variations in the VIs should still be considered and could be relevant for understanding VIs' relations with SWI.

From the cartographic visualization implementation point of view, it can be demonstrated that simple UIs like the slider bars can add important information about the temporal dimension of the data, and will provide additional initiatives for users to perceive the complexities of the topic through visual thinking. Using Leaflet to provide the main UI has the advantages of easy implementation and simple interaction for exploring various thematic datasets. Additionally, this product demonstrates the value-adding role of cartographic visualization in remote sensing analysis by allowing users to interact with the data/results and generate their own insights.

Aligning with other related work, vegetation indicators can reflect soil moisture conditions through the change in the VI. Different characteristics of each investigated VI infer that the heterogeneity in vegetation's vitality, vegetation's evapotranspiration and photosynthetic primary production capacity over the landscape contribute to the explanation of soil water heterogeneity underneath. Because the VIs only reveal the current state of the vegetation and do not immediately reflect when the vegetation is water-stressed, long time series of VIs are more ideal for interpreting their true conditions after some accumulation in plants' water-limited features over time. Time lagged analysis between vegetation and soil moisture can show the real-time soil moisture condition.

5.2 Limitations and Issues Encountered

Some limitations in both the remote sensing analysis and visualization development of this project are inevitable. These limitations occur due to data, time and technical constraints within the scope of the research. Below the different aspects of issues and limitations encountered will be discussed.

In the correlation and multiple regression analysis, statistics used are derived from VIs at sample sites over 10 selected time stamps. Limitations are related to the time stamp selection. These dates are selected firstly to reflect the regional flow dynamics based on the in-situ water level record and secondly to cater to the availability of cloud-free Sentinel-2 images. Therefore, analysis of the temporal relations are limited to seasonal variations by grouping these dates, and evenly spaced time series cannot be constructed. Unevenly spaced time series are plotted to show the corresponding behaviors of the SWI and VIs, but not used to conduct cross-correlation analysis for time delay investigation.

Another constraint is related to the land cover product used. The CGLS 100m Land Cover product provides discrete and cover fraction layers of land cover classes for the year of 2015 in Africa. This discrete land cover classification is used in the research to identify the dominant land cover type of each site, and the cover fraction layer for seasonal inland water assessment to determine the quality of SWI and whether to include a site or not. Even though the land cover product is developed to reflect the land cover distribution in 2015, it is assumed most land cover does not change drastically over 2 to 3 years (the time span of the data analyzed). The exception is, water in the Delta environment is very dynamic, the flooded extent in one location could be different within a year, therefore, using the 2015 land cover to determine a sample site's flood regime may not represent the flood extent in years after 2015. Other datasets, such as the 10-daily water body extent dataset provided by CGLS, can be helpful for understanding the flood regime of the sample sites in the future.

Related to the dominant vegetation type selection, the discrete land cover class with the highest percent area in each site is selected. In the 28 sites (2 excluded for having a high percentage of seasonal inland water coverage), the shrubs and DBOF are the dominant types, and 20 sites have DBOF as dominant types and 8 sites for shrubs. There are also sites (Table 3) that have two competently high vegetation types, such as site 284 with 42.51% shrub and 44.98% DBOF, in these cases the highest is still chosen to be the dominant type. So the signals observed in VIs are not purely representing one type of vegetation. In the correlation analysis in Figure 27, the observations in these two groups do not show a drastic difference in VI when correlating to SWI. This means, the grouping reflects general behaviors of how VIs correlate with SWI in sites with these two prominent vegetation types, rather than a definite indication that sites with a higher percentage of shrubs will always show a higher correlation with or better estimation for SWI. In the heterogeneous landscape of the Okavango Delta, it is unlikely to find one site with purely one vegetation to use as a stereotype. Nevertheless, the distinction made with the dominant vegetation helps to address the landscape heterogeneity.

Related to using the GLCM's second order texture measures, entropy and homogeneity, to describe the vegetation structure in this research, they two features are selected among 8 other common measures available for GLCM because they are the most commonly used measures and are representative of the contrast and regularity characteristics of the image. Other texture measures should also be explored, but many of the measures are strongly correlated to each other, thus, should take into account the collinearity. Different moving window size and moving direction could also be tested further.

An issue encountered when developing the web-based interactive visualization is raster data storage and hosting. GeoServer is tested several times to host the raster data used in the visualization via WMS, but data access request made by remote clients require server-side machine to be active during the time of access. Permanent access is possible but requires the WMS to be hosted by a commercial platform, such as the Amazon Web Service, which generates additional costs. Moreover, the WMS hosted by GeoServer can only be accessed by the remote client via the server-side machine's IP address, and after several rounds of testing, this connection is not very stable, thus it is abandoned. Alternatively, the raster data used are reduced in file size to be hosted locally, but this could result in slower responding time with the product once it is on the web. However, even though hosting raster data via WMS is unsuccessful, other existing WMS can be easily integrated into this platform because it is developed using JavaScript Mapping API. Many raster or vector datasets are distributed using WMS, they can be added onto JavaScript Mapping API, like Leaflet, as a simple overlay in few lines of code (Figure 44).

European Commission

Joint Research Centre

Global Surface Water - Data Access

NEW LAYERS

The new dataset 1984-2018 is available to download

Web Map Services

The Global Surface Water data can also be used within other websites or GIS clients by using what are called 'Web Map Services'. These services provide a d the data and produce cartographic products. They are not suitable for analysis as the data are represented only as RGB images.

Desktop GIS

In ArcGIS for Desktop:

- In the ArcCatalog Window, click on GIS Servers and then double click on Add WMTS Servers
- In the URL box, enter: https://storage.googleapis.com/global-surface-water/downloads_ancillary/WMTS_Global_Surface_WaterV2.xml and click OK
- Expand the 'Global Surface Water on storage.googleapis.com' item and drag any of the layers onto the map

In QGIS:

- In the Manage Layers toolbar, click on Add WMS/WMTS Layer
- Click New and enter a name(Global Surface Water) and URL (https://storage.googleapis.com/global-surface-water/downloads_ancillary/WMTS_Globa
- Click OK and click Connect - select a layer to add to the map

Websites

The Web Map Services can also be used within websites using any Javascript Mapping API that supports tiled layer types. The examples below show you hov transitions, occurrence, change, seasonality, recurrence or extent.

Leaflet:

```
var transitions = new L.tileLayer("https://storage.googleapis.com/global-surface-water/tiles2018/transitions/{z}/{x}/{y}.png",
{ format: "image/png",
  maxZoom: 13,
  errorTileUrl : "https://storage.googleapis.com/global-surface-water/downloads_ancillary/blank.png",
  attribution: "2016 EC JRC/Google" });
map.addLayer(transitions);
```

ArcGIS Javascript API:

Figure 44. Example of a data hub providing demonstrations on how to use the data via WMS on a JavaScript Mapping API.

6 Conclusion and Outlook

This project aligns with the ongoing scientific efforts to explore the relationships between remote sensing vegetation traits and soil moisture and seeks to use vegetation as sensors for soil monitoring. The implementation of the interactive web map demonstrate the potentials that cartographic visualization has in adding values to remote sensing analysis and appeal to a border audience.

The remote sensing analysis in this project makes use of the popular remote sensing products, ASCAT-SWI for soil moisture and the Sentinel-2 for VI retrieval, to demonstrate the relationships between vegetation and soil moisture in the uniquely complex wetland environment. High spatial resolution vegetation traits calculated from the Sentinel-2 data convey information about the spatial heterogeneity within each coarse SWI pixel, and indicate that the spatial heterogeneity of vegetation traits and the variation in vegetation structure are related to the soil moisture conditions at the time when the information is obtained. The grouping techniques, by seasons and by dominant vegetation type, also shed light on how the soil-vegetation relations change under the influence of regional flow dynamics and dominant vegetation patterns. At the end, it can be inferred that the timestamps when vegetation flourishes reflected through high values in VIs signals, the correlations with soil moisture are stronger; low VIs indicating low vegetation vitality show a weak correlation with SWI. In sites with different dominant vegetations, correlations strength does not differ by much, but vegetation structures could possibly make an influential distinction, but this needs further analysis. Therefore, the research exhibits elements in line with state-of-the-art soil monitoring in remote sensing.

The cartographic visualization makes use of mainly the popular Leaflet JavaScript library and HTML/CSS to implement a web-based interactive platform for data exploration. The user groups for the project are targeted at students in remote sensing classrooms and decision-makers who need insights for building in-situ sensor networks. This visualization can be introduced to remote sensing students as a case study and can demonstrate a workflow of remote sensing analysis as well as the multidimensional nature of remote sensing data and natural phenomena. People who are developing in-situ sensor networks for water or soil can use this platform to get an overview of the patterns in the soil and vegetation and identify potentially interesting locations for further investigation. Furthermore, it is also a good indication that cartographic visualization and remote sensing analysis combined can make scientific findings more visible and engaging to a broader audience.

Limitations exist in this research and indicate the need for further scientific efforts. The vegetation indicator method can reflect soil moisture conditions as VI changes, but VI cannot immediately reflect when the vegetation is stressed. Time lagged analysis is not implemented because of the limitation in Sentinel-2 data due to cloud coverage and general temporal resolution. Therefore, long term and high temporal resolution series providing information on the vegetation traits should be developed and analyzed to better uncover the time lag between vegetation dynamics and soil water content. Further improvement of the visualization product could be to add more case studies in other focal area, to implement advanced computation capacities with spatial data accessed from other data hubs via WMS, and to conduct user tests for feedback on usability.

References

Alexandridis, T., Cherif, I., Bilas, G., Almeida, W., Hartanto, I., van Andel, S., . . . Araujo, A. (2016). Spatial and Temporal Distribution of Soil Moisture at the Catchment Scale Using Remotely-Sensed Energy Fluxes. *Water, 8*(1), 32.

Bartalis, Z., Naeimi, V., Hasenauer, S., & Wagner, W. (2008). *ASCAT Soil Moisture Report Series No. 15 ASCAT Soil Moisture Product Handbook.* Institute of Photogrammetry and Remote Sensing, Vienna University of Technology, Austria.

Bauer-Marschallinger, B., Paulik, C., Hochstöger, S., Mistelbauer, T., Modanesi, S., Ciabatta, L., . . . Wagner, W. (2018). Soil Moisture from Fusion of Scatterometer and SAR: Closing the Scale Gap with Temporal Filtering. *Remote Sensing, 10*(7), 1030.

Calvet, J.-C., Wigneron, J.-P., Walker, J., Karbou, F., Chanzy, A., & Albergel, C. (2011). Sensitivity of Passive Microwave Observations to Soil Moisture and Vegetation Water Content: L-Band to W-Band. *IEEE Transactions on Geoscience and Remote Sensing, 49*(4), 1190-1199.

Couclelis, H. (1998). Worlds of Information: The Geographic Metaphor in the Visualization of Complex Information. *Cartography and Geographic Information Systems, 25*(4), 209-220.

Dirmeyer, P., Gao, X., Zhao, M., Guo, Z., Oki, T., Hanasaki, N., . . . Hanasaki, N. (2006). GSWP-2: Multimodel Analysis and Implications for Our Perception of the Land Surface. *Bulletin of the American Meteorological Society, 87*(10), 1381-1398.

Dorigo, W., Gruber, A., De Jeu, R., Wagner, W., Stacke, T., Loew, A., . . . Kidd, R. (2015). Evaluation of the ESA CCI soil moisture product using ground-based observations. *Remote Sensing of Environment, 162*, 380-395.

Dorigo, W., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Drusch, M., . . . Jackson, T. (2011). The International Soil Moisture Network: a data hosting facility for global in situ soil moisture measurements. *Hydrology and Earth System Sciences Discussions, 8*(1), 1609-1663.

Elachi, C., & Van Zyl, J. (2006). *Introduction to the physics and techniques of remote sensing.* Wiley-Interscience.

Esri. (2001). *ArcGIS ® 9 Using ArcGIS ® Geostatistical Analyst.*

Flanagan, D. (2006). *JavaScript : the definitive guide.* O'Reilly.

Fuchs, R., & Hauser, H. (2009). Visualization of Multi-Variate Scientific Data. *COMPUTER GRAPHICS forum, 28*(6), 1670-1690.

Fulakeza, M., Druyan, L., & Krishnamurti, T. (2002). A simple soil moisture scheme for regional climate simulations in the tropics. *Meteorology and Atmospheric Physics, 79*(1-2), 105-126.

Gao, B.-c. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment, 58*(3), 257-266.

Hall-Beyer, M. (2017). *GLCM Texture: A Tutorial v. 3.0 March 2017.* University of Calgary.

Jackson, T., Cosh, M., Zhao, T., Starks, P., Bosch, D., Seyfried, M., . . . Leroux, D. (2012). Validation of Soil Moisture and Ocean Salinity (SMOS) Soil Moisture Over Watershed Networks in the U.S. *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, 50*(5).

Johnson, A. (1962). *Methods of measuring soil moisture in the field.*

Klinke, R., Kuechly, H., Frick, A., Förster, M., Schmidt, T., Holtgrave, A.-K., . . . Neumann, C. (2018). Indicator-Based Soil Moisture Monitoring of Wetlands by Utilizing Sentinel and Landsat Remote Sensing Data. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science, 86*(2), 71-84.

Köbben, B., & Kraak, M.-J. (1999). WEBCARTOGRAPHY: DISSEMINATION OF SPATIAL DATAON THE WEB. *2nd AGILE conference on Geographic Information Science, At Roma (Italy).* Roma, Italy.

Legates, D., Mahmood, R., Levia, D., DeLiberty, T., Quiring, S., Houser, C., & Nelson, F. (2011). Soil moisture: A central and unifying theme in physical geography. *Progress in Physical Geography: Earth and Environment, 35*(1), 65-86.

Lu, D., & Batistella, M. (2005). Exploring TM image texture and its relationships with biomass estimation in Rondônia, Brazilian Amazon. *Acta Amazonica, 35*(2), 249-257.

MacEachren, A., & Kraak, M.-J. (2001). Research Challenges in Geovisualization. *Cartography and Geographic Information Science, 28*(1), 3-12.

Mahmood, R. (1996). Scale issues in soil moisture modelling: problems and prospects. *Progress in Physical Geography: Earth and Environment, 20*(3), 273-291.

Ochsner, T., Cosh, M., Cuenca, R., Dorigo, W., Draper, C., Hagimoto, Y., Kerr, Y., Njoku, E., Small, E., Zreda, M. (2013). State of the Art in Large-Scale Soil Moisture Monitoring. *Soil Science Society of America*(77), 1888–1919.

Oki, T., & Kanae, S. (2006). Global Hydrological Cycles and World Water Resources. *Science, 313*(5790), 1068-1072.

Plewe, B., & Brandon. (1997). *GIS online : information retrieval, mapping, and the Internet.* OnWord Press.

Robinson, A., Demšar, U., Moore, A., Buckley, A., Jiang, B., Field, K., . . . Sluter, C. (2017). Geospatial big data and cartography: research challenges and opportunities for making maps that matter. *International Journal of Cartography, 3*(sup1), 32-60.

Romano, N. (2014). Soil moisture at local scale: Measurements and simulations. *Journal of Hydrology, 516*, 6-20.

Roth, R. (2017). User Interface and User Experience (UI/UX) Design. *Geographic Information Science & Technology Body of Knowledge, 2017*(Q2).

Roth, R., Donohue, R., Sack, C., Wallace, T., & Buckingham, T. (2015). A Process for Keeping Pace with Evolving Web Mapping Technologies. *Cartographic Perspectives*(78), 25-52.

Sanchez, N., Martinez-Fernandez, J., Scaini, A., & Perez-Gutierrez, C. (2012). Validation of the SMOS L2 Soil Moisture Data in the REMEDHUS Network (Spain). *IEEE Transactions on Geoscience and Remote Sensing, 50*(5), 1602-1611.

Seneviratne, S., Corti, T., Davin, E., Hirschi, M., Jaeger, E., Lehner, I., . . . Teuling, A. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews, 99*(3-4), 125-161.

Steiner, A., Pal, J., Giorgi, F., Dickinson, R., & Chameides, W. (2005). The coupling of the Common Land Model (CLM0) to a regional climate model (RegCM). *Theoretical and Applied Climatology, 82*(3-4), 225-243.

Torres-Rua, A., Ticlavilca, A., Bachour, R., & McKee, M. (2016). Estimation of Surface Soil Moisture in Irrigated Lands by Assimilation of Landsat Vegetation Indices, Surface Energy Balance Products, and Relevance Vector Machines. *Water, 8*(4), 167.

Venables, W., & Ripley, B. (2002). *Modern Applied Statistics with S Fourth edition* (4 Ausg.). Springer.

Wagner, W., Lemoine, G., & Rott, H. (1999). A method for estimating soil moisture from ERS Scatterometer and soil data. *Remote Sensing of Environment, 70*(2), 191-207.

Weiss, M., & Baret, F. (2016). *S2ToolBox Level 2 products: LAI, FAPAR, FCOVER.*

Wood, E., Pidgeon, A., Radeloff, V., & Keuler, N. (2012). Image texture as a remotely sensed measure of vegetation structure. *Remote Sensing of Environment,* 121, 516-526.