Masters Program in Geospatial Technologies

Landsat 8 satellite data-based estimation of soil moisture in McMurdo Dry Valleys, Antarctica

Raül Raga Gonzalez

Dissertation submitted in partial fulfilment of the requirements for the Degree of *Master of Science in Geospatial Technologies*

Landsat 8 satellite data-based estimation of soil moisture in the McMurdo Dry Valleys, Antarctica

Raül Raga Gonzalez

al398189@uji.es
rragagon@uni-muenster.de

Dissertation supervised by

Maite Lezama Valdes, PhD Candidate
Institute for Geoinformatics
University of Münster

Prof. Dr. Hanna Meyer
Institute for Geoinformatics
University of Münster

Prof. Dr. Joel Dinis Baptista Ferreira da Silva
Information Management School
Universidade NOVA de Lisboa

February 2021

UNIVERSITAT JAUME I

NOVA IMS Information Management School

ifgi Institute for Geoinformatics University of Münster

Declaration of Academic Integrity

I hereby confirm that this thesis on *Landsat 8 satellite data-based estimation of soil moisture in the McMurdo Dry Valleys, Antarctica* is solely my own work and that I have used no sources or aids other than the ones stated. All passages in my thesis for which other sources, including electronic media, have been used, be it direct quotes or content references, have been acknowledged as such and the sources cited.

_____

Münster, 24th February of 2021

I agree to have my thesis checked in order to rule out potential similarities with other works and to have my thesis stored in a database for this purpose.

_____

Münster, 24th February of 2021

Acknowledgements

I would first like to thank the support of my thesis supervisors Maite Lezama Valdes and Hanna Meyer for their very valuable comments, help and guidance during the performance of this research. Also, thanks to Joel Dinis Baptista Ferreira da Silva for the comments provided to improve my study.

I must express my very profound gratitude to my parents and brother for giving me with unfailing support and continuous encouragement throughout this master and process of researching and writing this thesis. Also, I would really like to show my sincere thankfulness to my partner for being so patience during whole master and allow me being sat for so many hours in front of my computer performing my tasks and work.

Finally, a huge mention to my special classmates and friends (Anu, David S., David P. and Poshan) who I have been running this path with from 2019 to now.

Thank you.

Contents

Chapter 1. Introduction

Chapter 2. Theoretical background

Chapter 3. Methodology

Chapter 4. Results

Appendix

List of Figures

List of Tables

List of Acronyms

L8 - Landsat 8

McMurdo Dry Valleys - MDV

NSDSIs - Normalized shortwave-infrared difference bare soil moisture indices

SWIR - Shortwave-infrared

TIRS - Thermal Infrared

LRM - Linear regression model

MLR - Multiple Linear Regression Model

Keywords

Soil moisture, McMurdo Dry Valleys, Antarctica, Landsat 8, climate stations, estimation, prediction, regression model.

Thesis Outline

Chapter 1 provides the motivation, theoretical framework and the related literature to this research as introduction, as well as software and PC specifications used. In Chapter 2, the theoretical background of the methods is described and also how to evaluate them.

Chapter 3 presents the study area and data in detail, methodology used in this thesis, data pre-processing and its result, and the experimental design applied to the data. In Chapter 4, the results of each method used in this the research are shown.

Finally, Chapter 5 argues and discusses the findings throughout the thesis and Chapter 6 comments the limitations and suggests recommendations for further research.

Chapter 1

Introduction

1.1 Motivation, theoretical framework and literature review

Soil moisture is the total amount of water present in the upper 10 cm of soil and it represents the water in land surface which resides in the pores of the soil which is not in river, lakes or groundwater and which depends of the weather conditions, soil type and associated vegetation, among others. Soil moisture assessments are important to understand the hydrological cycles and biophysical processes caused by global climate changes (Finn et al., 2011). Usually, soil moisture has been mapped with airborne microwave radiometers (Klemas et al., 2014) to measure the water retained in the spaces between soil particles. Its importance is due to the microorganism metabolic activity, regulation of the soil temperature and carriage of nutrients, among others. Soil moisture typically takes the form of small ice crystals, vapour, or small parts of liquid water in cold desert soils (Campbell & Claridge, 1982).

Antarctic soils are composed by basically no organic and very low moisture content (Campbell and Claridge, 1987). Antarctica is a sensitive area to balance the global climate and its changes and its soil ecosystems are strongly regulated by variables of the abiotic environment and due to this, a research measures the incidence and spatial occurrence of the layer freezing to know how regional climate change could affect the energy exchange of this layer and its invertebrate communities (Wlostowski et al., 2017). Also, knowing how the dynamic of the surface varies in polar regions is transcendent to predict the impact of climate change in global sea-level rise in the future (Quincey & Luckman, 2009).

The soils present in McMurdo Dry Valleys are a central component of the polar desert ecosystem which are very susceptible to human activities (Campbell & Claridge, 2013). Over ten years of monitoring, Seybold et al. (2010) found a very low water content without any increasing or decreasing tend in McMurdo Sound Region soils.

A research conducted by Levy et al. (2014) was performed to determine if remote sensing techniques could be used to assess the conditions of soil moisture in the McMurdo Dry Valleys, Antarctica. A spectrometer to measure the wetted samples collected in this area under natural illumination conditions was used in order to evaluate their reflectance in the laboratory. The results suggested airborne hyperspectral imagery as adequate to generate soil moisture maps for the McMurdo Dry Valleys due to the measurement in the laboratory of the soil moisture values from the samples taken in the study area and their reflectance in the spectra at 1.4 μm and 1.9 μm. Moreover, in another research (Tian & Philpot, 2015), three soil samples with different properties were taken and measured using a spectrometer ASD FieldSpec® Proin in the laboratory with a spectral range of 350–2500 nm to know the relationship between the surface soil water content and SWIR bands reflectance. From saturated to dry water soil content, the bands present at 1440 nm and 1930 nm were shown very susceptive of these changes.

Also, Sadeghi et al. (2015) performed a research to create a linear physically-based model for remote sensing of soil moisture using SWIR bands and verifies these bands as the most appropriate to detect the surface water content in the solar domain (350 – 2500 nm) and the accuracy shown in the band 7 of Landsat and MODIS satellites (SWIR – 2210 nm) exhibits an optimistic method to detect soil moisture through laboratory-measured spectral reflectance data of different soils datasets from Lobell & Asner (2002) and Whiting et al. (2004). To see the result in different kind of soils, a study by Lobell & Asner (2002) measures the reflectance in four soils taken from temperate and arid ecosystems with different characteristics where the connection of SWIR reflectance and water saturation degree shows the most useful remote sensing relationship, mostly when moisture values are over 20%.

In other paper, authors argue that NSMI (Normalized Soil Moisture Index), an index based in the reflectance of the SWIR bands ($\sigma[1800\ nm]-\sigma[2119\ nm]\ /\ \sigma[1800\ nm]+\sigma[2119\ nm]$) was successfully measured in the laboratory to quantify surface soil moisture from a high resolution hyperspectral airborne sensor data (HyMap) and field soil samples taken from a mine of lignite in Germany from low to absent vegetation density and sandy/clayey composition (Haubrock et al., 2008). Also, Jibo et al (2019) studied the application of three

normalized shortwave-infrared difference bare soil moisture indices (NSDSIs) based on the water absorption difference between shortwave-infrared (SWIR) bands to estimate the soil moisture content in bare soils with an R2 around 0.8. In this study, Sentinel 2 data was acquired and compared to field measurements, also field samples were collected and measured their reflectance in a laboratory using an ASD FieldSpec 3 spectrometer. Authors concluded that values to estimate soil moisture got from NSDSIs work well when the values of these three proposed indices have a range between 0–50%.

In this research a new approach is going to be implemented using Landsat 8 (L8) imagery and climate stations series analysis data as a base to estimate soil moisture in McMurdo Dry Valleys (MDV), Antarctica.

1.2 Problem Statement

The problem statement of this research is the use of remote sensing imagery to detect soil moisture in a cold desert as McMurdo Dry Valleys, Antarctica.

The results of this study examining the significance and relationship between Landsat 8 satellite imagery and climate stations data on the estimation of the soil moisture in the cold desert of McMurdo Dry Valleys, Antarctica.

1.3 Aim and Research Question

Through the comparison of satellite imagery which provides information in a wider spatial context and time series data from climate stations located in the study area, the aim of this research is testing the potential of the Landsat 8 and climate stations time series data to predict soil moisture.

Therefore, this study wants to answer the research question about which relationship patterns follow Landsat 8 imagery data and climate stations data to estimate soil moisture in McMurdo Dry Valleys, Antarctica.

1.4 Software and PC specification

The system of the laptop used during this research has a processor Intel® Core™ i7-7500U CPU @ 2.70GHz 2.90 GHz, a RAM memory of 8 GB and a system type of 64-bit Operating System, x64-based processor.

R (version 4.0.3) is the programming language used to build whole code in the integrated development environment software RStudio Version 1.3.1093. Also, the software ArcGIS Pro version 2.5 is used to locate the climate stations in the study area and for creating maps.

Chapter 2

Theoretical background

2.1 NSDSI indices

Three normalized shortwave-infrared difference bare soil moisture indices (NSDSIs) based on the water absorption difference between shortwave-infrared bands (SWIR) were developed to estimate the soil moisture content (Jibo et al., 2019) using next formulas:

$$NSDSI1 = (SWIR1-SWIR2) / SWIR1$$

$$NSDSI2 = (SWIR1-SWIR2) / SWIR2$$

$$NSDSI3 = (SWIR1-SWIR2) / (SWIR1+SWIR2)$$

2.2 Statistical Analysis - Graphics

For visualizing quantitative data, statistical graphics are used. The most common plot is the scatterplot for data analysis when is necessary to understand the nature of the association between two variables.

2.3 Statistical Analysis - Linear Regression Model and Multiple Linear Regression Model

Linear regression model (LRM) is a statistical procedure for predicting the dependent variable from an independent variable, measuring thus the relationship between them (Kumari, 2018):

$$Y = \beta_0 + \beta_1 X$$

Multiple Linear Regression Model (MLR) is the statistical method to predict the values of a dependent variable from a set of independent variable values (Sinharay, 2010) as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ...+ \beta_i X_i$$

Where (in both formulas) Y is the dependent variable, X the independent(s) variable, β_0 is the intercept (the predicted value of Y when X is 0) and β_i is the regression coefficient (how much we expect Y to change as X increases).

2.4 Evaluation of Regression Models

R-squared (R2), or coefficient of determination, p-values and Root Mean Square Error (RMSE) values are calculated to evaluate the fit of each LRM and MLR models, the significance of the relationships between independent and dependent variables and the validity of the models, respectively.

R-squared (R2) or coefficient of determination is the proportion of the variation in the dependent variable (Y) which is described by the independent variable in the model (X) (Peng et al., 2002) and it ranges between 0 and 1. Apart of this, and assuming that the null hypothesis is right, the p-value is the likelihood of producing outcomes at least as extreme as the results of a statistical hypothesis test obtained. Getting low p-value means better evidence for the alternative hypothesis is available (Beers, 2021). The Root Mean Square Error (RMSE) is a standard way to measure the error of a model in predicting quantitative data (Moody, 2021) as next:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

Where $\hat{y}$ is the predicted values, y are the values measured and n is the total of observations measured.

Chapter 3

Methodology

3.1 Study Area

The study area for this research are the McMurdo Dry Valleys (Figure 1), one of the coldest and driest extreme deserts on Earth which is located along the western coast of the Ross Sea in east of Antarctica (77-78°S 160-164°E), a continent where about 98% of the continent is covered by ice. MDV cover a total area of 22700 km² and its ice-free area has 4500 km². (Levy, 2013). The mean air temperature measured between 1986 and 2017 in the MDV varied between -14.7°C and -29.6°C (Obryk et al., 2020).

McMurdo Dry Valleys are ice-free because of the presence of the Transantarctic Mountains which stop the ice from the polar plateau and avoid it goes to these valleys. This cold desert creates an important arid environment as the evaporation in the area is over the usual snowfall of 1cm approx. annually (Bromley, 1986). This snow is mostly transformed to gas and the rest is melted and infiltrated into the soil (Gooseff et al. 2006, Fountain et al. 2009).

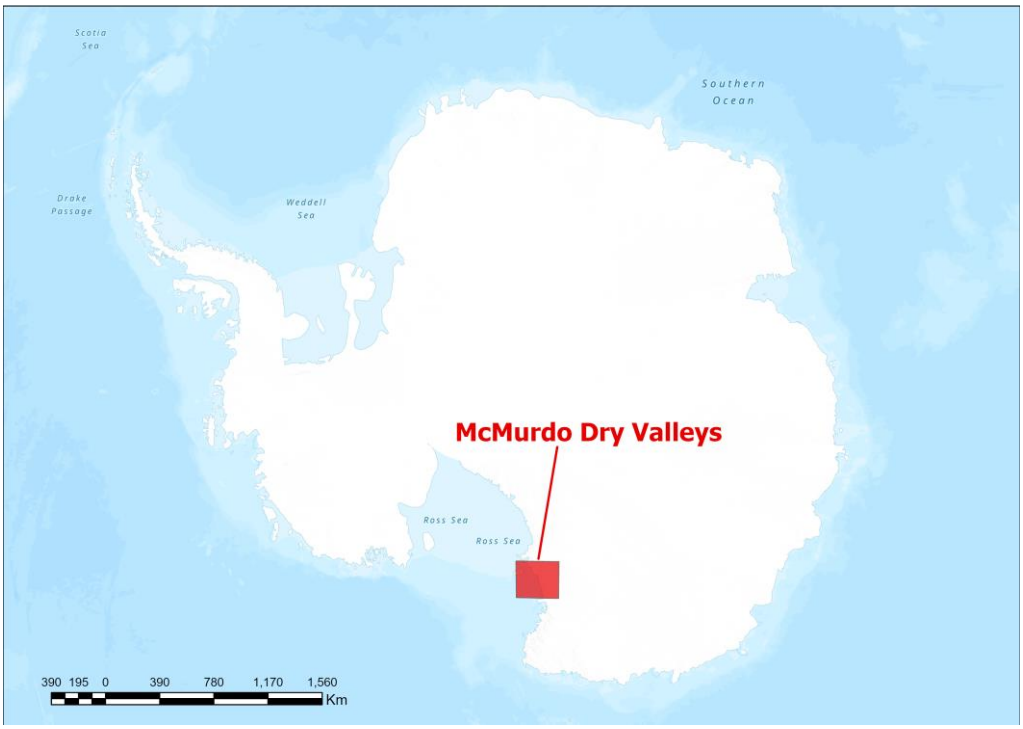

Fig.1. Location of the McMurdo Dry Valleys in Antarctica.

3.2 Data

3.2.1 Satellite data

Landsat 8 is the satellite chosen to perform this research as a useful and important amount of data from April 2013 to December 2019 can be obtained free of charge to be analysed. It is provided by USGS (United States Geological Survey) and was ordered and downloaded in its website (https://espa.cr.usgs.gov/). This data is Landsat 8 Level-2 Data Products at a 30 and 100-meter spatial resolution which contains Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) Surface Reflectance (Figure 2).

These products include, according to the USGS, an approximation of surface spectral reflectance as determined at ground level in the absence of atmospheric dispersion or absorption. At a 30-meter spatial resolution, the Surface Reflectance products are produced at the Earth Resources Observation and Science (EROS) Center. To build Level-2 data items, the EROS Science Processing Architecture (ESPA) on-demand interface corrects satellite images for atmospheric effects.

	Bands	Wavelength (micrometers)	Resolution (meters)
Landsat 8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) **Launched February 11, 2013**	Band 1 - Coastal aerosol	0.43 - 0.45	30
	Band 2 - Blue	0.45 - 0.51	30
	Band 3 - Green	0.53 - 0.59	30
	Band 4 - Red	0.64 - 0.67	30
	Band 5 - Near Infrared (NIR)	0.85 - 0.88	30
	Band 6 - SWIR 1	1.57 - 1.65	30
	Band 7 - SWIR 2	2.11 - 2.29	30
	Band 8 - Panchromatic	0.50 - 0.68	15
	Band 9 - Cirrus	1.36 - 1.38	30
	Band 10 - Thermal Infrared (TIRS) 1	10.60 - 11.19	100
	Band 11 - Thermal Infrared (TIRS) 2	11.50 - 12.51	100

Fig. 2. Landsat 8 bands wavelength and resolution.

3.2.2 Climate Stations data

The Antarctica Soil Climate Research Stations data is managed and owned by NRCS - USDA (Natural Resources Conservation Services – United States Department of Agriculture), which is involved in a project originally lead by Landcare Research (New Zealand) with a total of nine climate stations (Bull Pass, Bull Pass East, Don Juan Pond,

Granite Harbour, Marble Point, Minna Bluff, Mt. Fleming, Scott Base and Victoria Valley). However, two of them (Minna Bluff and Mt. Fleming) are not used in our further analysis as both get the soil moisture values in a depth (over 3cm) which is not suitable for this study.

Therefore, data from seven climate stations from 2013 to 2019 is downloaded from https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/climate/. Soil moisture data is measured every day and hourly with the hydra-probe sensors from the company 'Stevens Water Monitoring Systems' - Portland (USA), which are installed at various depths in the active layer of the soil. In this research, just the shallowest soil moisture values are considered which are the measurements taken between at 2 and 3 cm depth. The reason of considering these soil moisture values is due to L8 imagery data, which are compared with, does not penetrate deeper layers of the soil.

The climate stations involved in this study are located in ice-free areas of the McMurdo Dry Valleys (Antarctica) to get the soil moisture values for further analysis (Figure 3).

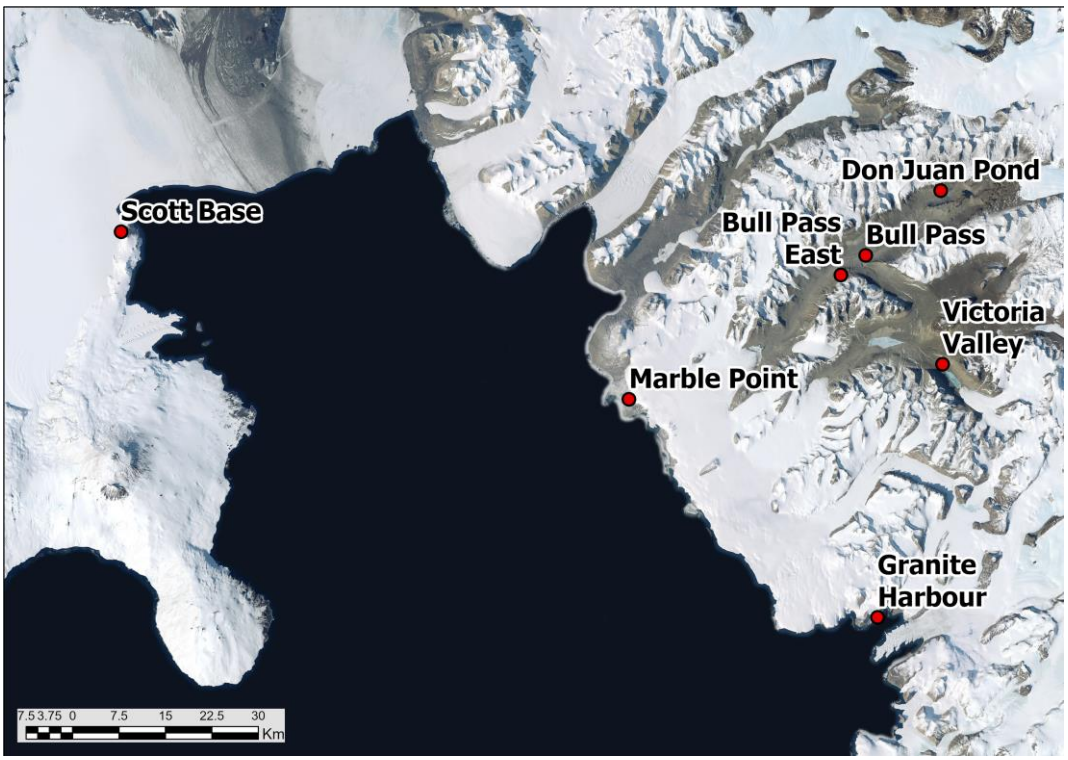

Fig. 3. Location of the Soil Climate Research Stations (in red dots) in McMurdo Dry Valleys, Antarctica.

3.3 Methods

3.3.1 Methodology workflow overview

An overview of the methodology workflow is displayed in Figure 4 and the steps followed toward to get the results.

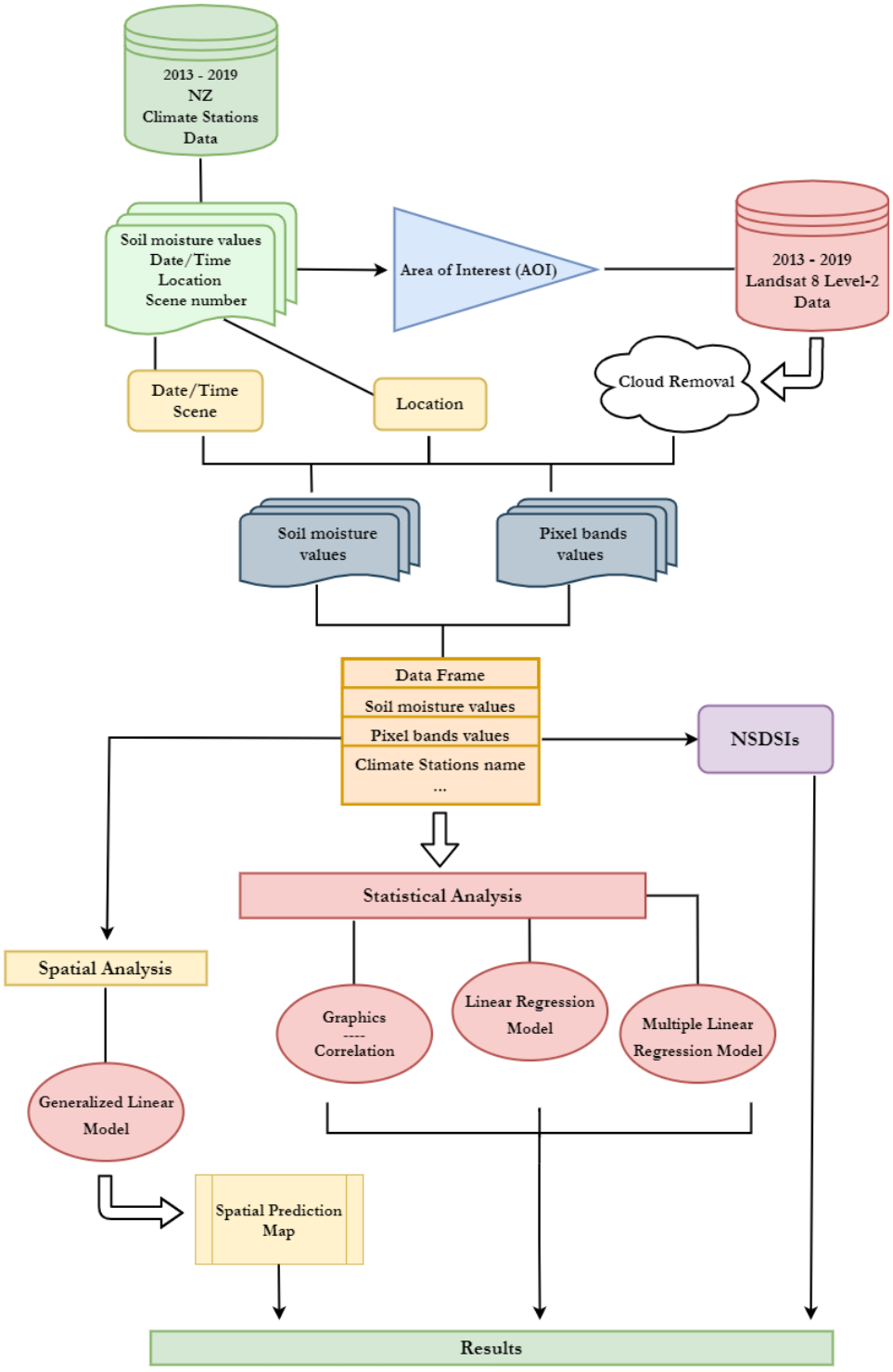

Fig. 4. Methodology workflow overview.

3.3.2 Data pre-processing

3.3.2.1 Climate Stations data

This data is pre-processed in order to be prepared it for further analysis, selecting just what is important for this research including date, hour, soil moisture value between 2 and 3 cm and name of the climate station. Between two and three measurements of soil moisture values were taken in five of the seven climate stations each day at the same time, therefore in these cases a new column was created in which the mean of these measurements was calculated which will be used for further analysis. Soil moisture values are measured in water fraction volume (m^3/m^3).

Also, all climate stations take their data in NZ standard time except for the Victoria Valley station which takes in NZ savings time which was corrected as the other ones for data harmonization.

3.3.2.2 Satellite data

A total of 180 images were downloaded for the years compressed between 2013 and 2019 (Table 1). L8 images used for this study area are only available from October to February coinciding with the austral summer. Imagery cannot be captured properly by L8 satellite during the remaining months as Antarctica is under constant darkness. Each L8 image downloaded from 2013 to 2019 has 9 different images inside them related to band 1 (Coastal Aerosol), band 2 (Blue), band 3 (Green), band 4 (Red), band 5 (Near Infrared - NIR), band 6 (SWIR-1), band 7 (SWIR-2), band 10 (Thermal Infrared) and band 11 (Thermal Infrared). These images have a cloud cover less than 10% and each of these images come with an associated file with *_pixel_qa.tif* extension.

Year	Images
2013	18
2014	36
2015	18
2016	30
2017	26
2018	22
2019	30
Total	180

Table 1. Total of images downloaded for each year.

The clouds present in these images are defined through different pixel values known for these features. To remove them, an atmospheric correction process is carried out which iterates individually in each image downloaded. During this process, TIRS bands (band 10 and band 11) are resampled to 30x30 meters of resolution to have all bands in same resolution for further statistical analysis.

This process runs as follows: if any of the pixel values related to clouds are found in the file with *_pixel_qa.tif* extension, this will be automatically removed from the image. After this, a new image without clouds for each image of each band is created.

Also, each Landsat 8 image comes with an .xml document which contains information about when (date and hour) the image was taken in UTM time. This time was transformed to New Zealand time to match with the climate stations date and time.

From each of the new images created without clouds during the atmospheric correction process, the values of the pixels of each band images are extracted in the same location where the climate stations are located when date and time (in New Zealand time) of each of these images and climate stations data matched at o'clock times during any day.

3.3.2.3 Pre-processing result

The result of the processes performed above for each year between 2013 and 2019 is a final data frame containing all the information needed for further analysis between these years (scene number, date/time, climate station name, soil moisture values and pixel values) with a total of 737 rows/observations. Some of these rows of this final data frame obtained are removed for a better data analysis performance in further analysis process as is explained in the corresponding section of this research when this occurs.

3.4 Experimental design

3.4.1 NSDSI indices

The NSDSI indices are applied to the data using SWIR bands as is required by these indices as have been mentioned using next formulas:

$$NSDSI1 = (SWIR1-SWIR2) / SWIR1$$
$$NSDSI2 = (SWIR1-SWIR2) / SWIR2$$
$$NSDSI3 = (SWIR1-SWIR2) / (SWIR1+SWIR2)$$

Where SWIR1 is band 6 and SWIR2 is band 7 of Landsat 8 satellite data.

Values obtained are compared to the original soil moisture values measured to check the suitability of the application of these indices to estimate soil moisture in the study area.

3.4.2 Statistical Analysis - Graphics

To find if there is any correlation between soil moisture values measured by the climate stations and L8 pixel values in same location, we make plots and calculate their coefficients of determination (R2) between these two variables for each band of L8 in each image downloaded. These bands are: band 1 (Coastal Aerosol), band 2 (Blue), band 3 (Green), band 4 (Red), band 5 (Near Infrared - NIR), band 6 (SWIR-1), band 7 (SWIR-2), band 10 (TIRS) and band 11 (TI).

This graphic procedure and obtaining R2 is also performed for soil moisture estimation values from NSDSI indices, mentioned above, and the original measured soil moisture values by the climate stations.

3.4.3 Statistical Analysis - Linear Regression Model and Multiple Linear Regression Model

Linear Regression Model is applied to predict the dependent variable which is soil moisture (Y) based on the independent variable (X) which is a Thermal Infrared (TIRS) band of Landsat 8 (band 10 or band 11) as follows:

$$Y = \beta_0 + \beta_1 X$$

Multiple Linear Regression Model is also used to explain the relationship between the two TIRS bands of Landsat 8 (bands 10 and 11) as independent variables (X) and soil moisture values as dependent variable (Y) as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

Apart of this MLR model mentioned, more covariates information about location, elevation and slope of the climate stations are added to the MLR models to know their significance into it. Two of the climate stations (Granite Harbour and Victoria Valley) do not present the slope information which are stated as 0 to carry on the research.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ...+ \beta_i X_i$$

3.4.4 Spatial Analysis – Spatial Prediction

Based on the Central Limit Theorem which declares that the distribution of the samples approximates to a normal distribution when the data increases the size assuming all samples are equal in size and independently of the population distribution pattern (Kwak and Kim, 2017) and also taking into account the continuous satellite data, a spatial soil moisture prediction map in a large ice-free area in MDV is performed.

In this section, soil moisture values are spatially predicted in one of the biggest ice-free areas in McMurdo Dry Valleys which contains four of the climate stations involved in this research (Bull Pass, Bull Pass East, Don Juan Pond and Victoria Valley) as can be seen in Figure 5. To this purpose, a Landsat 8 image obtained in 26th December 2019 which contains these climate stations mentioned is selected to be used.

The goal of the spatial soil moisture prediction map process is the estimation of the soil moisture values at the locations where no measurements have been made using the best regression model obtained. This is performed using the climate stations data and pixels values of the L8 image selected within the ice-free area on interest drawn as mentioned.

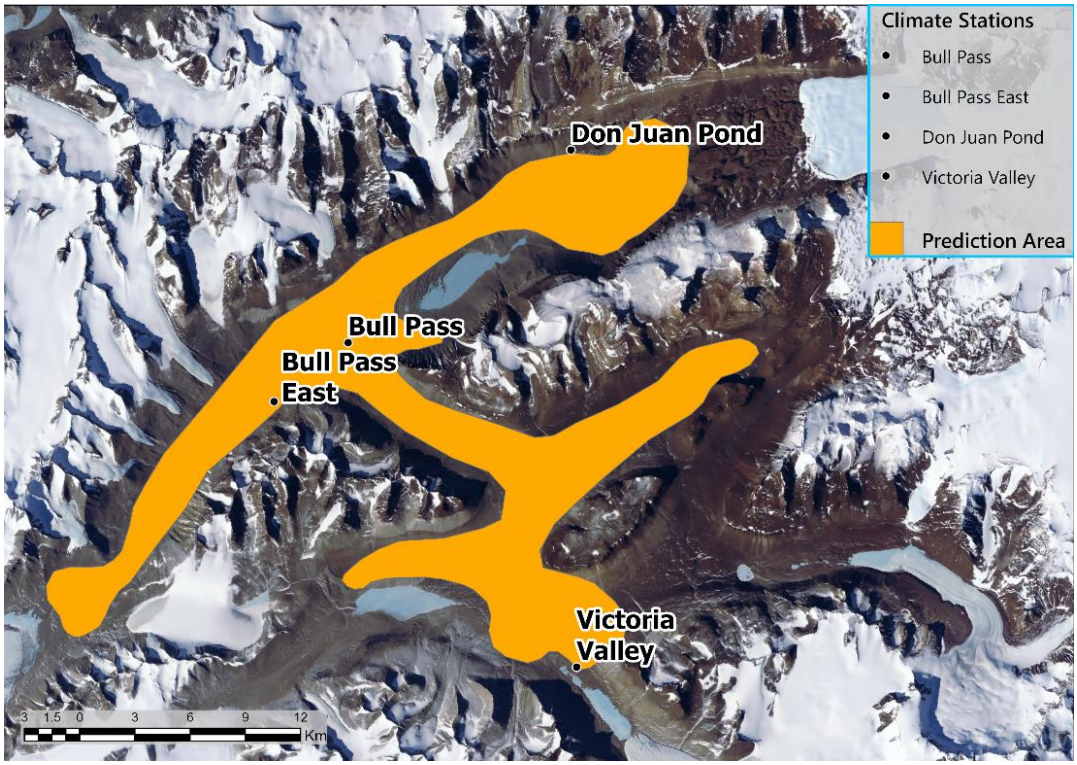

Fig. 5. Area chosen for spatial soil moisture prediction.

Chapter 4

Results

4.1 NSDSI indices

First, the final data frame obtained is reduced until containing soil moisture values measured from 0 to 0,205 (m3/m3) to apply the NSDSI indices. This range contains almost 92% of soil moisture values measured, therefore the data frame used to calculate NSDSIs has 668 lines/observations.

Based on NSDSIs formulas, estimated soil moisture values are obtained and are compared to soil moisture values measured in all climate stations to check the suitability of the application of these indices to estimate soil moisture in the study area. The coefficient of determination (R2) of these comparisons between measured and predicted soil moisture values can be seen in the first column of Table 2. Also, each climate station is checked individually to find in which one of them these NSDSI indices have the best performing. The best climate station to estimate soil moisture applying NSDSI indices is Scott Base and its R2 obtained throughout the comparisons between measured and predicted soil moisture values can be find in the second column of Table 2.

NSDSI indices	R2 (All Climate Stations)	R2 (Scott Base)
NSDSI1	0.019	0.088
NSDSI2	0.032	0.075
NSDSI3	0.026	0.085

Table 2. R2 values obtained in NSDSIs computation.

As Table 2 shows, for all climate stations data, NSDSI2 is the best performing index, however and looking individually in each climate station, Scott Base is the one which presents the highest R2 values compared with other climate stations with its highest value in the NSDSI1 index (0.088).

4.2 Statistical Analysis - Graphics

Same data frame which contains 668 lines/observations mentioned in last section (4.1) is used to plot and obtaining the R2 between the pixel values of Landsat 8 bands and soil moisture values measured by all climate stations. R2 obtained can be seen in first column of the Table 3.

Also, Don Juan Pond climate station is removed as it has all its soil moisture values measured as 0. Therefore, a data frame with six climate stations and a total of 614 lines/observation are used to obtain the coefficient of determination (R2) which can be seen in second column of Table 3.

Landsat 8 bands	R2 (All Climate Stations)	R2 (All except Don Juan Pond)
Band 1 (Coastal Aerosol)	0.0098	0.0041
Band 2 (Blue)	0.0064	0.0021
Band 3 (Green)	0.0032	0.00067
Band 4 (Red)	0.0019	0.00029
Band 5 (NIR)	0.0018	0.00028
Band 6 (SWIR1)	0.00069	0.00084
Band 7 (SWIR2)	0.0054	0.0042
Band 10 (Thermal)	0.031	0.032
Band 10 (Thermal)	0.026	0.028

Table 3. R2 values obtained in graphs calculation between soil moisture values and pixel values of each band of Landsat 8.

TIRS bands (bands 10 and 11) obtained the highest results in terms of R2. Considering this results and the literature review presented which mentions SWIR as most successful bands to detect soil moisture content, R2 is also calculated for each climate station individually for SWIR and TIRS bands of L8 to test which of these bands have a higher correlation detecting soil moisture content. Highest R2 for SWIR and TIRS bands and its climate stations related are shown in Table 4.

Landsat 8 bands	R2 and Climate Station
Band 6 (SWIR1)	0.2 (Granite Harbour)
Band 7 (SWIR2)	0.22 (Granite Harbour)
Band 10 (Thermal)	0.43 (Scott Base)
Band 11 (Thermal)	0.41 (Scott Base)

Table 4. Highest R2 values obtained in graphs computation for each climate station between soil moisture values, and SWIR and TIRS bands of Landsat 8.

Granite Harbour and Scott Base climate stations exhibit the highest R2 value for SWIR and TIRS bands, respectively. Granite Harbour individual data frame contains 87 lines/observations and Scott Base has 68 lines/observations.

4.3 Statistical Analysis – Linear Regression Model and Multiple Linear Regression Model

The data frame used in sections 4.1 and 4.2 which contains 668 lines/observations is also used to run LRM and MLR models. A total of 24 models are performed (see Appendix) to check the covariance between Landsat 8 bands, soil moisture values, and location, elevation and slope of the climate stations. Also, R2 values are obtained which show the correlation between the covariates. However, only 5 models are displayed in Table 5. As can be seen, these models which present the highest R2 are the ones which contain the covariates location, elevation and slope.

	Covariates								Evaluation
Models	Soil moisture	Band 1 + 2 + 3 + 4 + 5	Band 6 + 7	Band 10 + 11	All bands	Location	Elevation	Slope	R2
MLR4	x	x				x	x	x	0.2939
MLR6	x	x	x			x	x	x	0.2973
MLR10	x			x		x	x	x	0.2818
MLR13	x				x				0.2446
MLR14	x				x	x	x	x	0.3187

Table 5. Covariates used and highest R2 values obtained in regression models computation.

Better performing models are the ones which include all bands (pixel values) information, soil moisture data measured from all climate stations and all covariates data. Also, MLR13 model has a R2 value of 0.2446 not far from others, being the model without location, elevation and slope information which has the highest goodness-of-fit in our data.

4.4 Spatial Analysis – Spatial Prediction

The data frame used keep being the same as sections 4.1, 4.2 and 4.3. To estimate soil moisture in the ice-free area of interest drawn, this data frame is split into two smaller datasets, one contains 75% (training dataset) of the data frame used chosen randomly and other has the remaining 25% of it (test dataset). Training dataset is taken for fitting the model and test dataset is used for validation. A new model (NM) is built using all bands data and soil moisture values measured using the training dataset which obtains a R2 of 0.2466. To validate the test dataset, NM is used for this purpose creating a new column of predicted soil moisture values. These values are compared with the soil moisture values measured with a R2 of 0.198 which means 19.8% of the values predicted are well fitted in the model. Also, a prediction soil moisture map (Figure 6) is created using the NM described before and the bands of the Landsat 8 image mentioned in section 3.4.4.

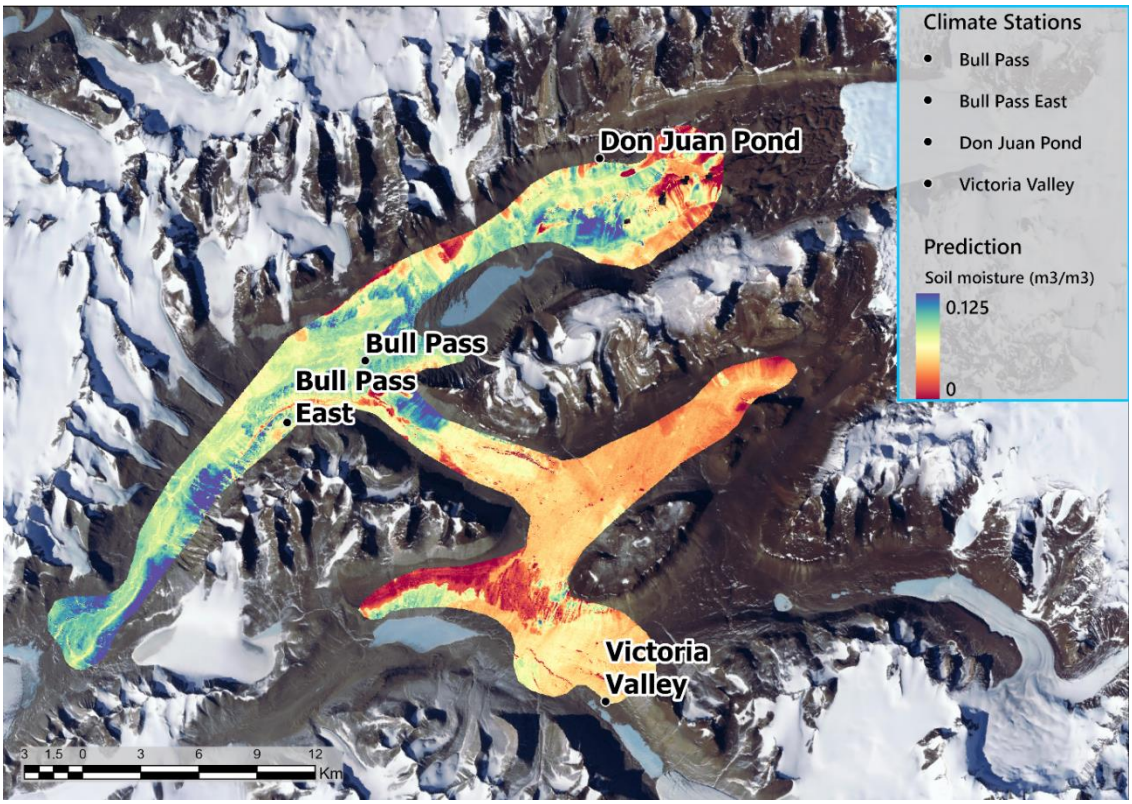

Fig. 6. Prediction soil moisture map using one L8 image from 2019.

The interpolation is made to estimate the soil water content at locations where no measurements have been taken by the climate stations in McMurdo Dry Valleys. This prediction result can be improved with the addition of all covariates (location, elevation and slope) information to the model used to create a more accurate prediction soil moisture map through the soil moisture values predicted.

The MLR14 model and the prediction soil moisture map model present a not much difference RMSE value which tells us that their amount of error in terms of comparing predicted and known values is similar (Table 6). However, as the difference between soil moisture values measured is really low this can play an important role. As is mentioned, all covariates information would be relevant in the prediction soil moisture map model as can be seen comparing both RMSE values as MLR14 (with covariates information) has a smaller error between predicted and measured soil moisture values. The significance of each covariate in each model is shown in Table 6. Elevation and slope covariates have importance in the MLR14 model so they would be transcendent in any prediction model.

	Models	
	MLR14	**Prediction Map Model**
RMSE	0.0239	0.0264
Covariates	**p-values (significance)**	
Band 1 (Coastal Aerosol)	***	***
Band 2 (Blue)	***	**
Band 3 (Green)	.	.
Band 4 (Red)	.	***
Band 5 (NIR)		***
Band 6 (SWIR1)		***
Band 7 (SWIR2)	*	***
Band 10 (Thermal)		
Band 11 (Thermal)		
Location		No data
Elevation	*	No data
Slope	**	No data
***	0 to 0.001	
**	0.001 to 0.01	
*	0.01 to 0.05	
.	0.05 to 0.1	
Blank	No significance	

Table 6. RMSE and p-values obtained in MLR14 and Prediction Map Model.

Chapter 5

Conclusions and Discussion

In this research a prediction soil moisture model has been applied to a remote and inaccessible region as the cold desert McMurdo Dry Valleys which demonstrates the power of remotely sensed data. This model can be applied without the need to move to the area of interest to predict soil moisture in an environmentally sustainable way with minimum economical costs to easily analyse large areas. This model could be applicable to other regions of the Dry Valleys, but caution is necessary as it would have to be used always in ice-free areas.

The significance of the relationship patterns between Landsat 8 imagery and climate stations soil moisture time series data have been also evaluated concluding that the data of these two covariates are not strongly related and that the addition of more information of other covariates would improve the results to predict soil moisture in McMurdo Dry Valleys, Antarctica. Despite what the literature review pointing out SWIR bands as the most useful to detect soil moisture, the spectral reflectance of the TIRS bands has been found as the most related with the soil moisture data got from the seven climate stations used in this research. The TIRS bands of L8 satellite have been used in others studies (Wicki & Parlow, 2017; Zubair & Iqbal, 2015) with vegetation as a covariate, factor that is not possible to be included in our models as our study area lacks of it. Our models revealed the importance of all covariates data applied into them in this research, which should be included in case these models wanted to be enhanced as has been proved the increasing of the goodness-of-fit using them. This strength observed highlights the sensitivity of the model depending the information contributed. The incorporation of these covariates would also provide a substantial benefit to the prediction soil moisture map, being a preliminary study of soil moisture estimation in this area, the result gotten a satisfactory seeing the peculiarity of the area and the data collected.

In this research just NSDSI indices (Jibo et al., 2019) can be applied as they were developed to predict soil moisture in bare surfaces as our study area is. However, and opposite to this

research mentioned in which soil moisture values reach until 80%, our soil moisture values measured range from 0 to 20% which could be the reason as these NSDSI indices are not relevant to McMurdo Dry Valleys. This argument is supported by Lobell & Asner (2002) as they assumed that a soil water saturation degree over 20% can be easier detected and reveal a stronger relationship with SWIR bands.

It is well known that soil reflectance is affected not only by soil moisture, but also by many other factors as soil characteristics, vegetation cover, topography, hydrology, atmospheric and weather conditions, soil particle size and sensor noise, among others (Ben-Dor & Irons, 1999) which prevent direct observations of soil responses in terms of spectral information (Muller & Decamps, 2001). Therefore, as more covariates information is included, more accurate performance model is assumed.

McMurdo Dry Valleys are an extreme, arid and cold desert where just 1cm (approx.) of snow falls annually (Bromley, 1986), however in terms of environmental factors, many more should be taken into account to assess the soil moisture in the area. Knowing the depth of the permafrost located in the soil layers would have to be studied which would be highly complex and difficult to quantify. Also, which areas of the Valleys receive the melt snow from the mountains and the topography would increase the knowledge of the area to evaluate the soil moisture.

Looking each climate station individually, Granite Harbour and Scott Base reveal more relationship between soil moisture values and the pixel values of SWIR and TIRS bands, respectively. Comparing to others, is outstanding the R2 values of 0.43 and 0.41 in Scott Base for TIRS bands (band 10 and 11). Water retention in this climate station can be explained because of its soil composition as it has a higher concentration of clay in the topsoil between 0 and 3 cm depth, this is around six times more than the others climate stations (NRCS - USDA, n.d.). This is because the soil's capacity to maintain water is closely related to particle size (Leeper & Uren, 1993) as water molecules keep more adhered to small particles (as clays) than thick particles (as sands).

Chapter 6

Limitations and Further Research

McMurdo Dry Valleys are a very special area as it holds very low soil moisture values without any clear tend observed in its values (Seybold et al., 2010). This desert does not have homogeneity and its area contains different environments as land covered by snow, lakes and rivers and bare soils, all of them with different topography which contributes negatively to the detection of a continuous feature as soil moisture.

Landsat 8 satellite imagery possess 30 meters and 100 meters (for TIRS bands) resolution, however the soil moisture is captured by sensors in the climate stations in an exact location point. Due to the spatial resolution, soil moisture is difficult to be detected by the pixel values of the images giving this a measurement error as their pixel size is much bigger than the point soil moisture measurement mentioned. Due to the nature of the study area, L8 satellite imagery can only be acquired within a short time window (November to February) because of the Antarctica has six months of darkness during its Austral winter. Also, the images downloaded have a high probability to contain clouds which could cover some climate stations making it impossible to compare the pixel value from the image with the climate stations data affected.

In terms of data, some climate stations can have erroneous measurements as has been seen in several outliers collected by the sensors which have to be considered for a correct analysis. Moreover, due to Landsat 8 available data just climate stations data from 2013 to 2019 could be used in this research. Important variables as location and topography data (for elevation and slope covariates) are not added to the model which might improve the results of the spatial soil moisture prediction map.

As soil moisture is dependent on other environmental factors, we recommend that for further studies models used in predicting soil moisture in McMurdo Dry Valleys should include and analyse relevant covariates such as soil composition, hydrology and topography, among others. The addition of these factors as covariates might help the

model in improving the prediction of the water soil content. Similarly, the temporal autocorrelation of the soil moisture is not considered in this research i.e. the dependence of the soil moisture in a time could be able to account for the temporal trends in the soil moisture. Apart of this, applying more advanced statistical models might improve the accuracy of the prediction of the soil moisture using the covariates mentioned.

In our study we only consider images from Landsat 8, trying other imagery from different satellites with other capabilities as higher resolution or/and different sensors (as microwave in radar satellites) would increase the results in our study area. In this research soil moisture data from 2013 to 2019 measured between 0 and 3 cm has been used, however soil moisture data taken more depth can be added and analysed to check their relationship with the pixel values from the satellite imagery in further research. Additionally, more soil moisture data can be incorporated to be evaluated with the satellite imagery, depending of the availability of these images. The methods used in this study could also be tested in other warm and cold deserts, or bare-soil areas to assess if soil moisture can be predicted in other region of the world with the appropriate data.

Finally, the spatial soil moisture prediction map is executed using a single Landsat 8 image, this gives us a static view for an exact date which represents a moment in time, therefore the model used in this prediction, or new models developed, can be applied to other satellite scenes collected.

References

Beers, B. (2021). What P-Value Tells Us. Retrieved 18 February 2021, from https://www.investopedia.com/terms/p/p-value.asp

Ben-Dor, Eyal & Irons, J. & Epema, G.F. (1999). Soil Reflectance.

Bromley, A.M. 1986. Precipitation in the Wright Valley. N.Z. Antarct. Rec. 6:60–68 Special Supplement.

Campbell, I.B., and Claridge, G.G.C., 1987. Antarctica: Soils, weathering processes and environment. Developments in Soil Science 16. Elsevier, New York.

Campbell, I.B. and Claridge, G.G.C., 1982. The influence of moisture on the development of soils of the cold deserts of Antarctica. Geoderma, 28: 221-238.

Campbell, Iain & Claridge, G. & Campbell, Dave & Balks, Megan. (2013). The Soil Environment of the Mcmurdo Dry Valleys, Antarctica. 10.1029/AR072p0297.

Finn, Michael & Lewis, Mark & Bosch, David & Giraldo, Mario & Yamamoto, Kristina & Sullivan, D.G. & Kincaid, Russell & Luna, Ronaldo & Allam, Gopala & Kvien, Craig & Williams, Michael. (2011). Remote Sensing of Soil Moisture Using Airborne Hyperspectral Data. GIScience & Remote Sensing. 48. 522-540. 10.2747/1548-1603.48.4.522.

Fountain, Andrew & Nylen, Thomas & Monaghan, Andrew & Basagic, Hassan & Bromwich, David. (2009). Snow in the McMurdo Dry Valleys, Antarctica, Int. International Journal of Climatology. 30. 633 - 642. 10.1002/joc.1933.

Gooseff, Michael & Lyons, W. & Mcknight, Diane & Vaughn, Bruce & Fountain, Andrew & Dowling, C.. (2006). A Stable Isotopic Investigation of a Polar Desert Hydrologic System, McMurdo Dry Valleys, Antarctica. Arctic Antarctic and Alpine Research - ARCT ANTARCT ALP RES. 38. 60-71. 10.1657/1523-0430(2006)038[0060:ASIIOA]2.0.CO;2.

Haubrock, Sören-Nils & Chabrillat, Sabine & Kuhnert, Matthias & Hostert, Patrick & Kaufmann, Hermann. (2008). Surface soil moisture quantification and validation based on hyperspectral data and field measurements. Journal of Applied Remote Sensing. 2. 10.1117/1.3059191.

Jibo, Yue & Tian, Jia & Tian, Qingjiu & Xu, Kaijian & Xu, Nianxu. (2019). Development of soil moisture indices from differences in water absorption between shortwave-infrared bands. ISPRS Journal of Photogrammetry and Remote Sensing. 154. 216-230. 10.1016/j.isprsjprs.2019.06.012.

Klemas, Victor & Finkl, Charles & Kabbara, Nijad. (2014). Remote Sensing of Soil Moisture: An Overview in Relation to Coastal Soils. Journal of Coastal Research. 296. 685-696. 10.2112/JCOASTRES-D-13-00072.1.

Kumari, Khushbu & Yadav, Suniti. (2018). Linear regression analysis study. Journal of the Practice of Cardiovascular Sciences. 4. 33. 10.4103/jpcs.jpcs_8_18.

Kwak, Sang & Kim, Jong. (2017). Central limit theorem: The cornerstone of modern statistics. Korean Journal of Anesthesiology. 70. 144. 10.4097/kjae.2017.70.2.144.

Leeper, G.W. and Uren, N.C. (1993) Soil Science, an Introduction. 5th Edition, Melbourne University Press, Melbourne.

Levy, J. (2013). How big are the McMurdo Dry Valleys? Estimating ice-free area using Landsat image data. *Antarctic Science, 25*(1), 119-120. Doi:10.1017/S0954102012000727.

Levy, Joseph & Nolin, Anne & Fountain, Andrew & Head, James. (2014). Hyperspectral measurements of wet, dry and saline soils from the McMurdo Dry Valleys: Soil moisture properties from remote sensing. Antarctic Science. 26. 565-572. 10.1017/S0954102013000977.

Lillis, D. (2021). Generalized Linear Models (GLMs) in R, Part 4: Options, Link Functions, and Interpretation – The Analysis Factor. Retrieved 8 February 2021, from https://www.theanalysisfactor.com/generalized-linear-models-glm-r-part4/

Lobell, D.B. & Asner, Gregory. (2002). Moisture Effects on Soil Reflectance. Soil Science Society of America Journal. 66. 722-727. 10.2136/sssaj2002.7220.

Moody, J. (2021). What does RMSE really mean?. Retrieved 15 February 2021, from https://towardsdatascience.com/what-does-rmse-really-mean-806b65f2e48e

Muller, Etienne & Decamps, Henri. (2001). Modeling soil moisture - Reflectance. Remote Sensing of Environment. 76. 173-180. 10.1016/S0034-4257(00)00198-X.

Natural Resources Conservation Services – United States Department of Agriculture. (n.d.). Antarctica Soil Climate Research Stations. Retrieved 14 February 2021, https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/climate/?cid=nrcs142p2_053772

Obryk, M. K., Doran, P. T., Fountain, A. G., Myers, M., & McKay, C. P. (2020). Climate from the McMurdo Dry Valleys, Antarctica, 1986 – 2017: surface air temperature trends and redefined summer season. Journal of Geophysical Research: Atmospheres. Doi:10.1029/2019jd032180.

Peng, Joanne & Lee, Kuk & Ingersoll, Gary. (2002). An Introduction to Logistic Regression Analysis and Reporting. Journal of Educational Research – J EDUC RES. 96. 3-14. 10.1080/00220670209598786.

Quincey, Duncan & Luckman, A. (2009). Progress in satellite remote sensing of ice sheets. Progress in Physical Geography. 33. 10.1177/0309133309346883.

Sadeghi, Morteza & Jones, Scott & Philpot, William. (2015). A Linear Physically-Based Model for Remote Sensing of Soil Moisture using Short Wave Infrared Bands. Remote Sensing of Environment. 164. 10.1016/j.rse.2015.04.007.

Seybold, Cathy & Balks, Megan & Harms, D.S. (2010). Characterization of active layer water contents in the McMurdo Sound region, Antarctica. Antarctic Science. 22. 633 – 645. 10.1017/S0954102010000696.

Sinharay, S. (2010) An Overview of Statistics in Education. In: Peterson, P., et al., Eds., International Encyclopedia of Education, 3rd Edition, Elsevier Ltd., Amsterdam, 1-11. https://doi.org/10.1016/B978-0-08-044894-7.01719-X

Tian, Jia & Philpot, William. (2015). Relationship between surface soil water content, evaporation rate, and water absorption band depths in SWIR reflectance spectra. Remote Sensing of Environment. 169. 280-289. 10.1016/j.rse.2015.08.007.

Whiting, Michael & Li, Lin & Ustin, Susan. (2004). Predicting Water Content Using Gaussian Model on Soil Spectra. Remote Sensing of Environment. 89. 535-552. 10.1016/j.rse.2003.11.009.

Wicki, Andreas & Parlow, E.. (2017). Multiple Regression Analysis for Unmixing of Surface Temperature Data in an Urban Environment. Remote Sensing. 9. 684. 10.3390/rs9070684.

Wlostowski, A. & Gooseff, M. & Adams, Byron. (2017). Soil Moisture Controls the Thermal Habitat of Active Layer Soils in the McMurdo Dry Valleys, Antarctica. Journal of Geophysical Research: Biogeosciences. 123. 10.1002/2017jg004018.

Zubair, Syed & Iqbal, Javed. (2015). Estimation of soil moisture using multispectral and FTIR techniques. The Egyptian Journal of Remote Sensing and Space Science. 18. 10.1016/j.ejrs.2015.10.001.

APPENDIX

NSDSI indices - Graphics

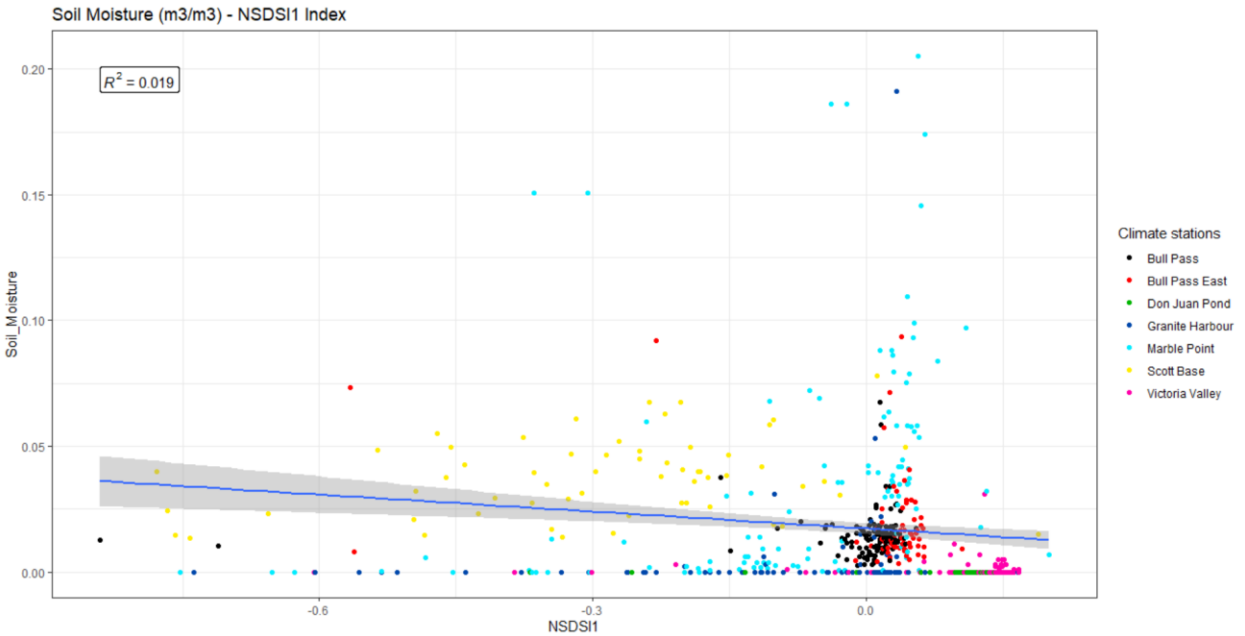

Fig. 1. Scatterplot of NSDSI1 index and soil moisture values measured in all climate stations.

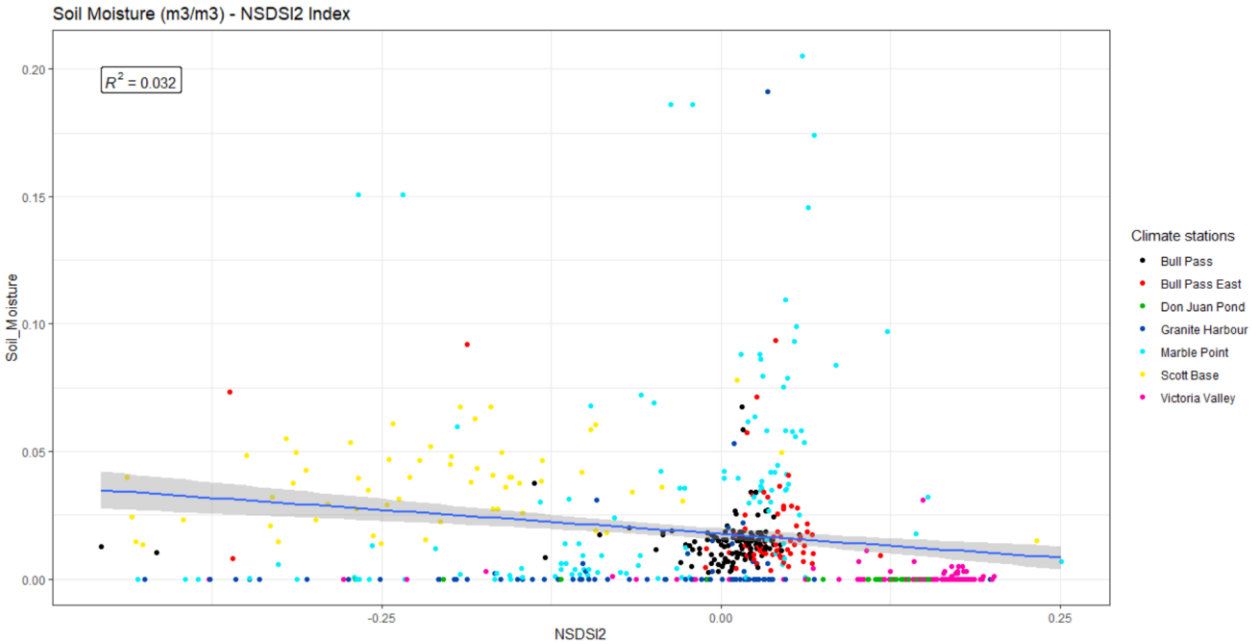

Fig. 2. Scatterplot of NSDSI2 index and soil moisture values measured in all climate stations.

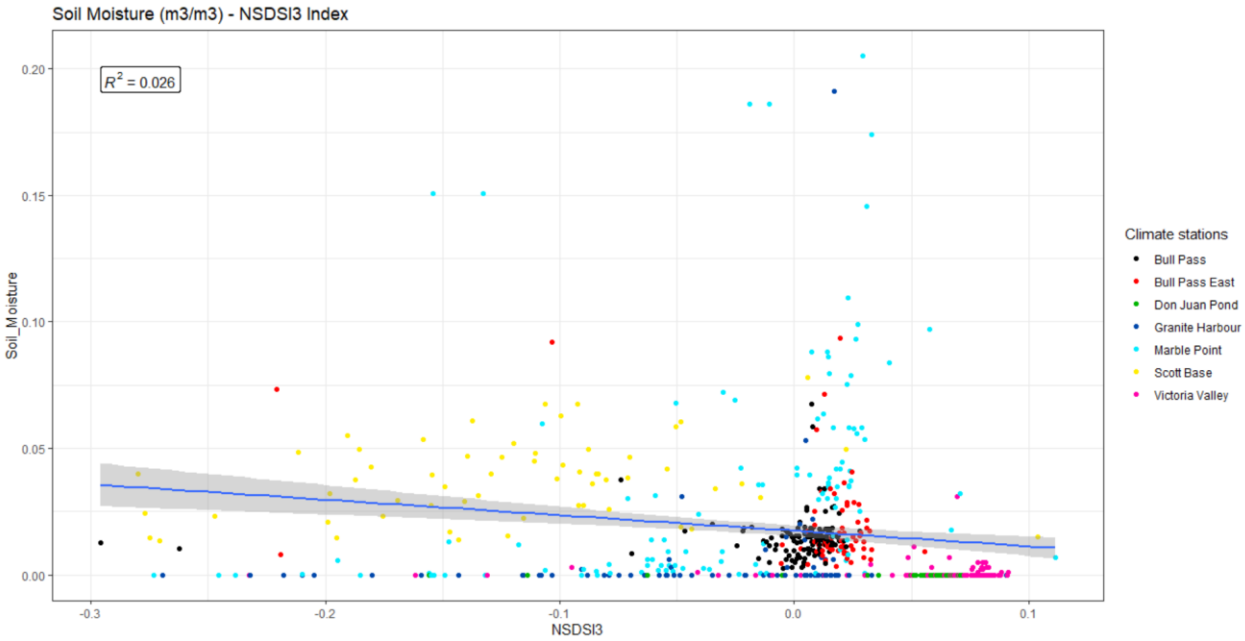

Fig. 3. Scatterplot of NSDSI3 index and soil moisture values measured in all climate stations.

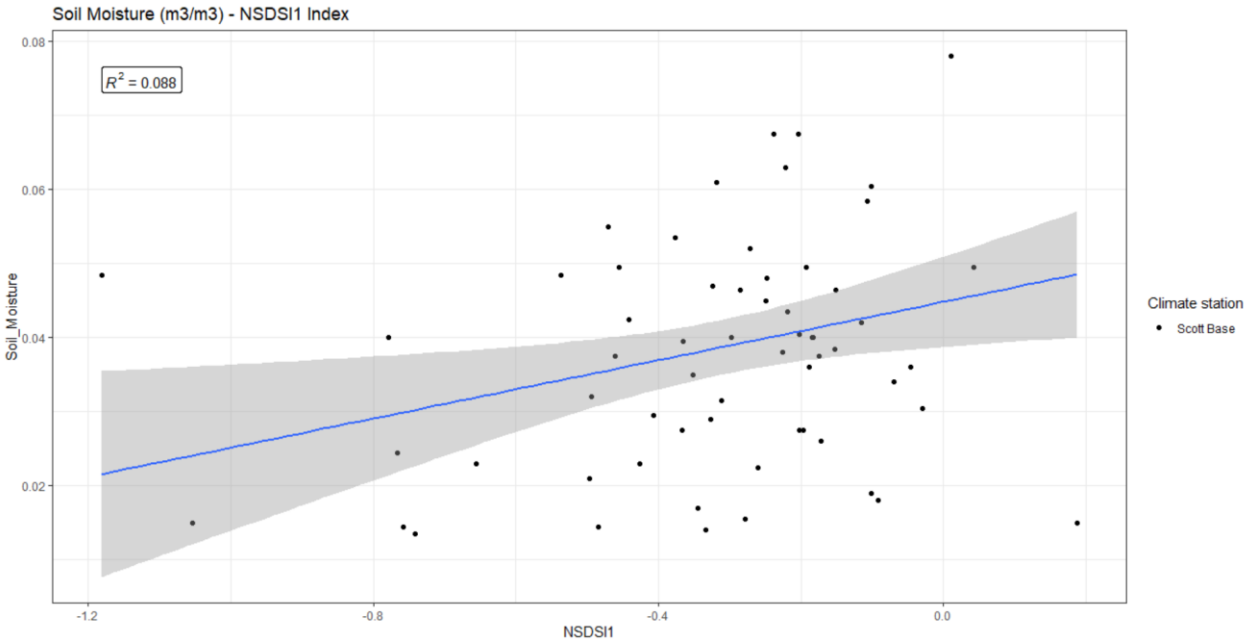

Fig. 4. Scatterplot of NSDSI1 index and soil moisture values measured in Scott Base climate station.

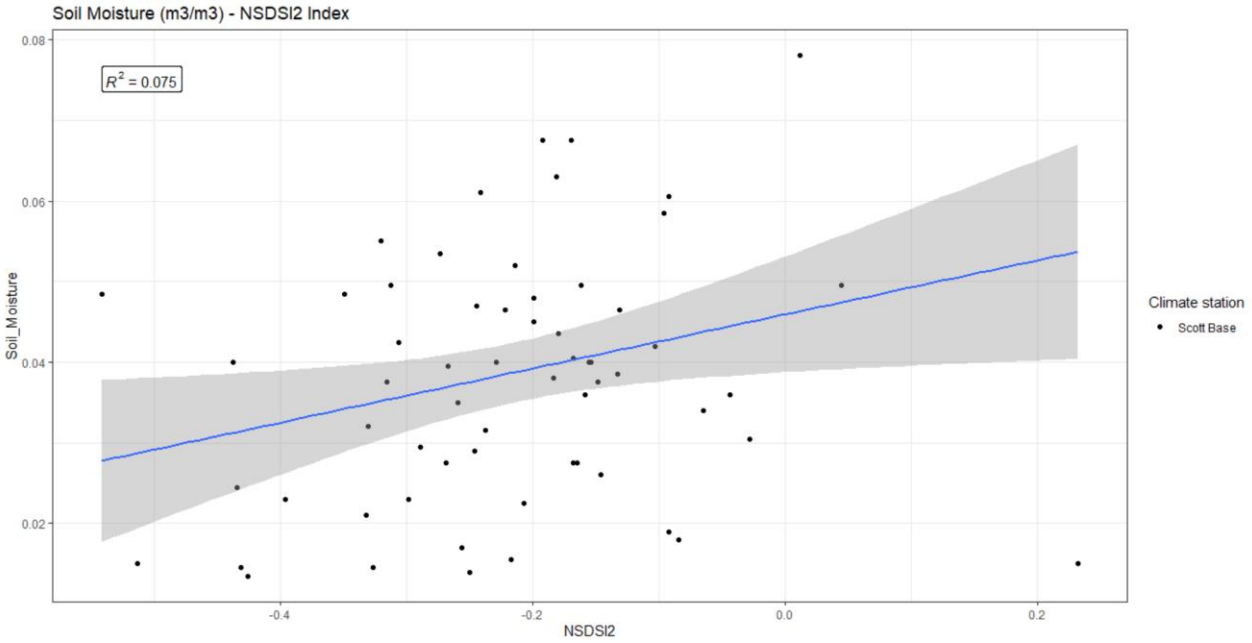

Fig. 5. Scatterplot of NSDSI2 index and soil moisture values measured in Scott Base climate station.

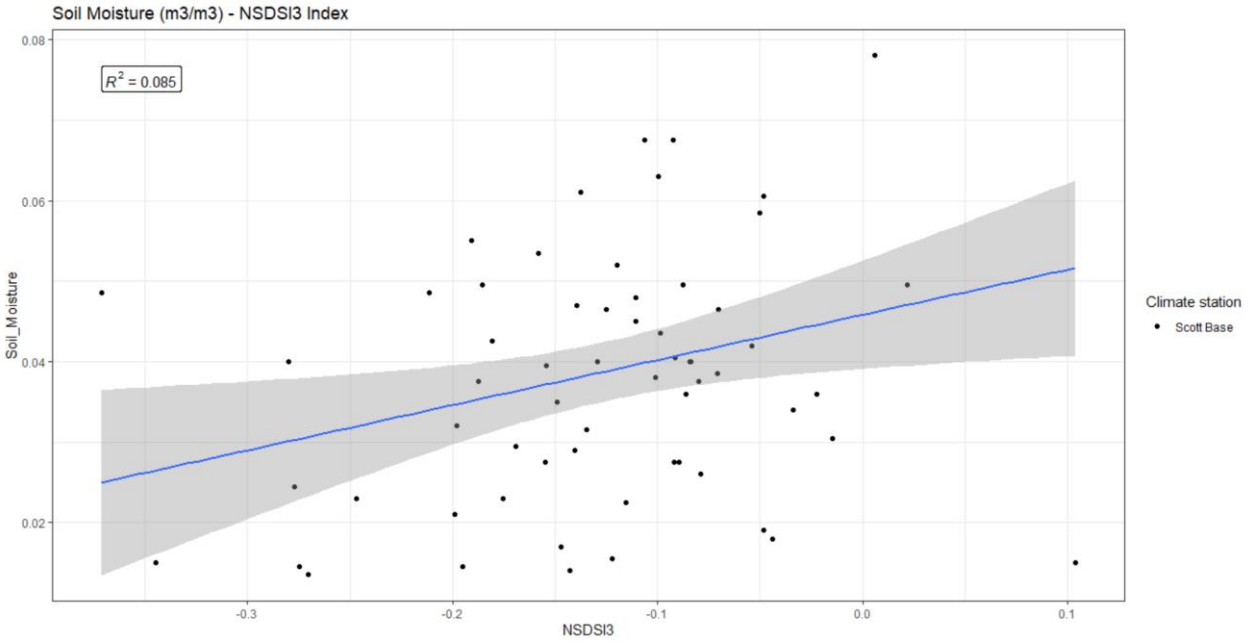

Fig. 6. Scatterplot of NSDSI1 index and soil moisture values measured in Scott Base climate station.

Statistical Analysis - Graphics
Soil moisture values and all climate stations involved

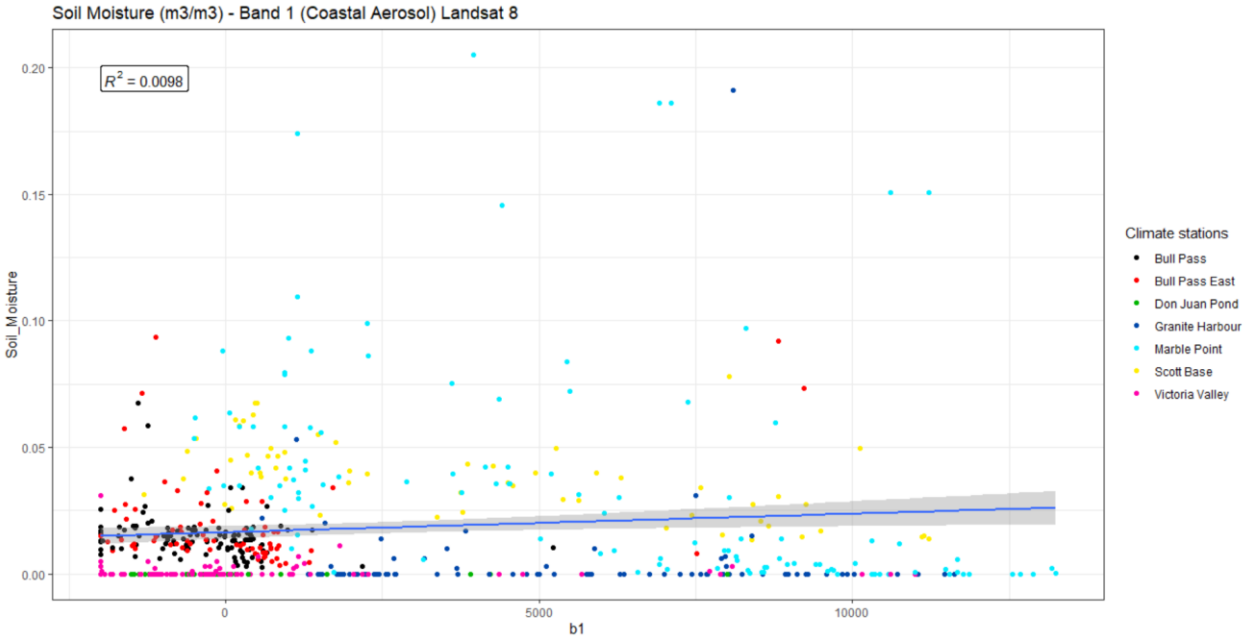

Fig. 7. Scatterplot of L8 Costal Aerosol band and soil moisture values measured in all climate stations.

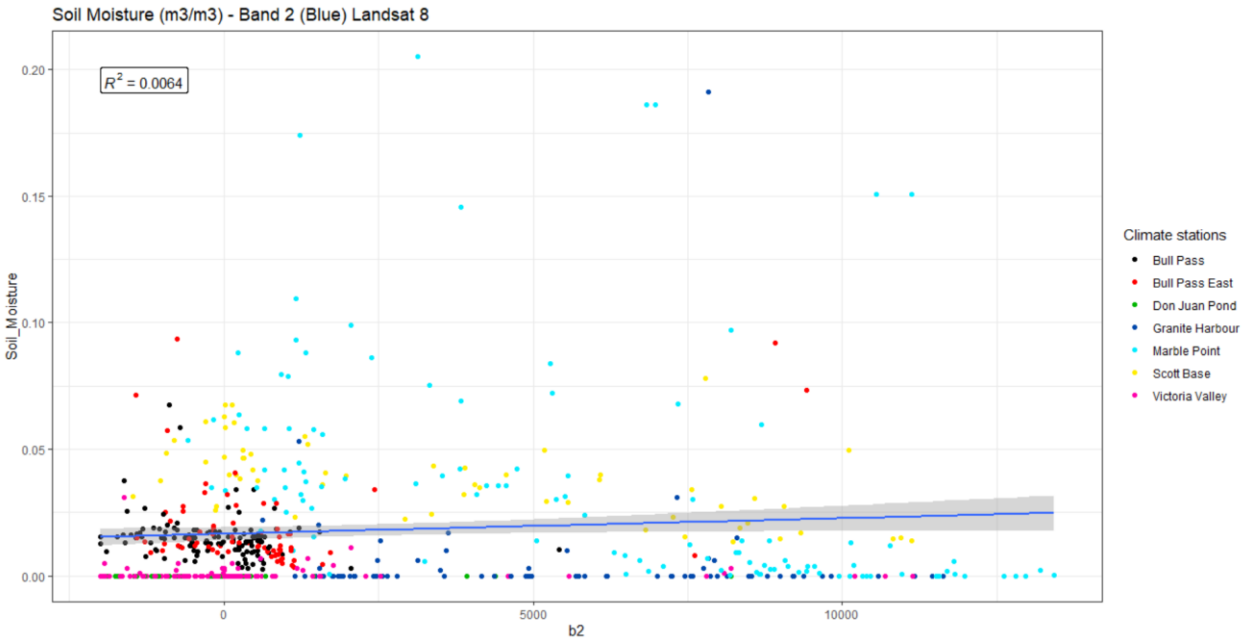

Fig. 8. Scatterplot of L8 Blue band and soil moisture values measured in all climate stations.

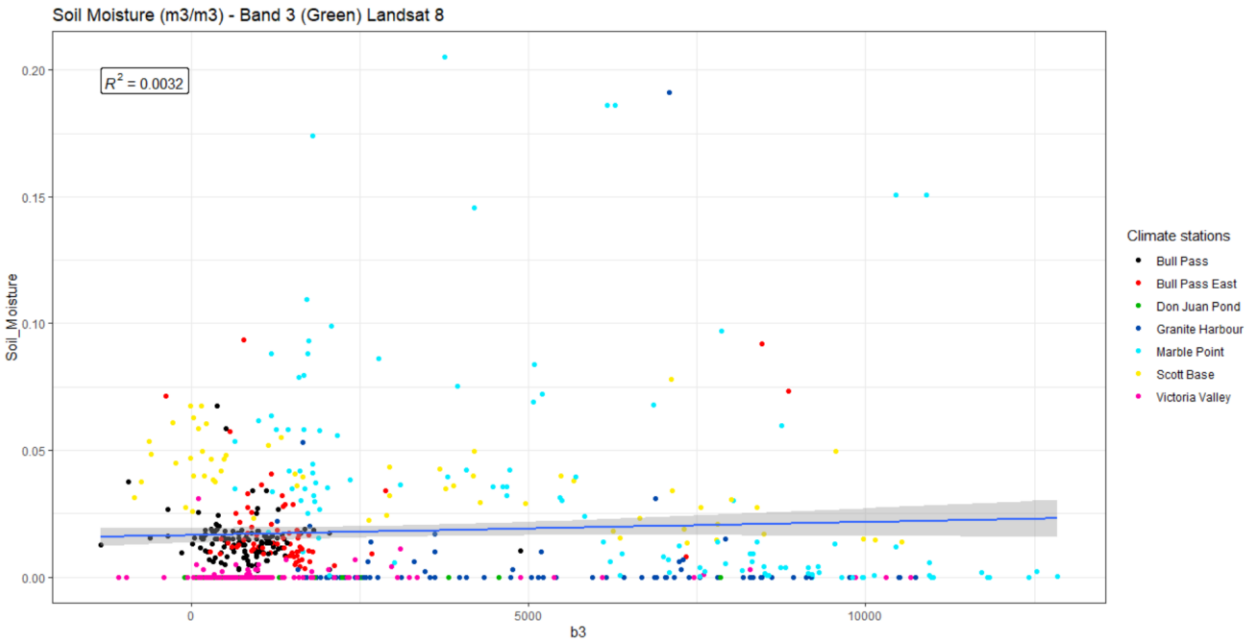

Fig. 9. Scatterplot of L8 Green band and soil moisture values measured in all climate stations.

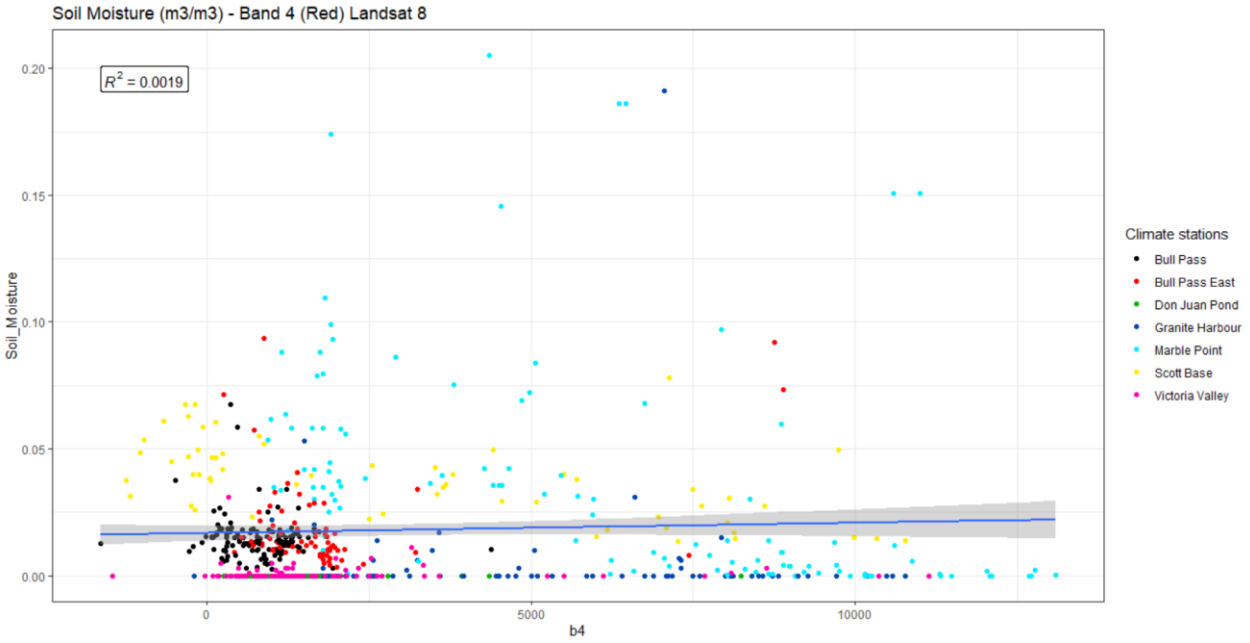

Fig. 10. Scatterplot of L8 Red band and soil moisture values measured in all climate stations.

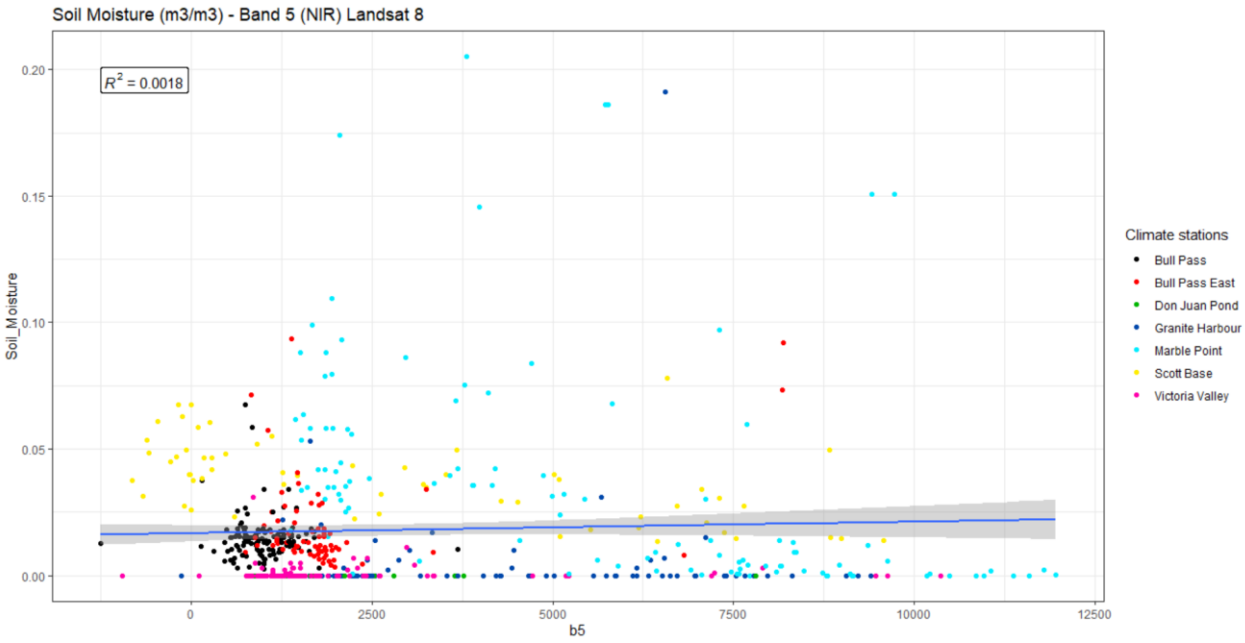

Fig. 11. Scatterplot of L8 NIR band and soil moisture values measured in all climate stations.

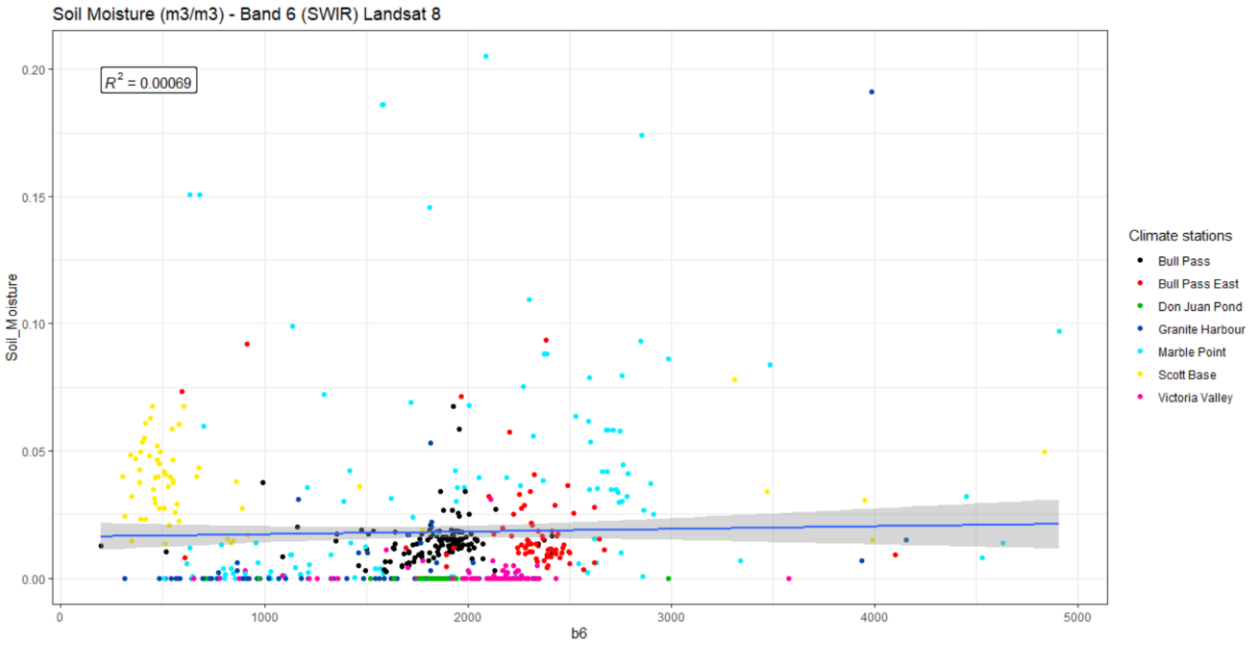

Fig. 12. Scatterplot of L8 SWIR 1 band and soil moisture values measured in all climate stations.

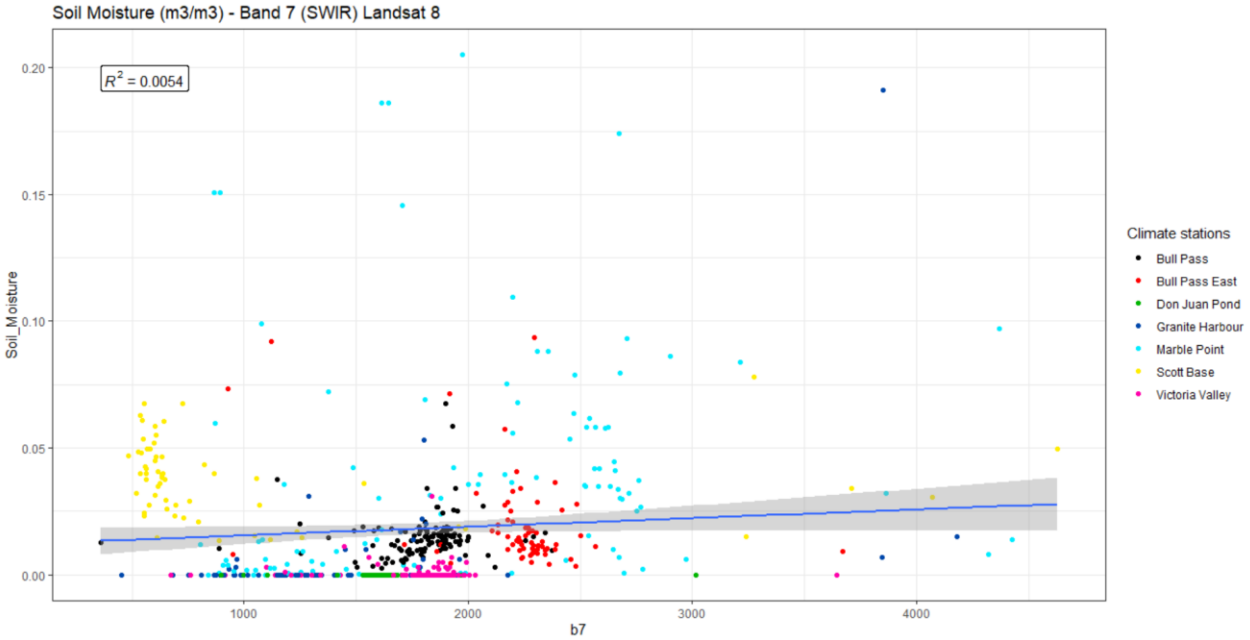

Fig. 13. Scatterplot of L8 SWIR 2 band and soil moisture values measured in all climate stations.

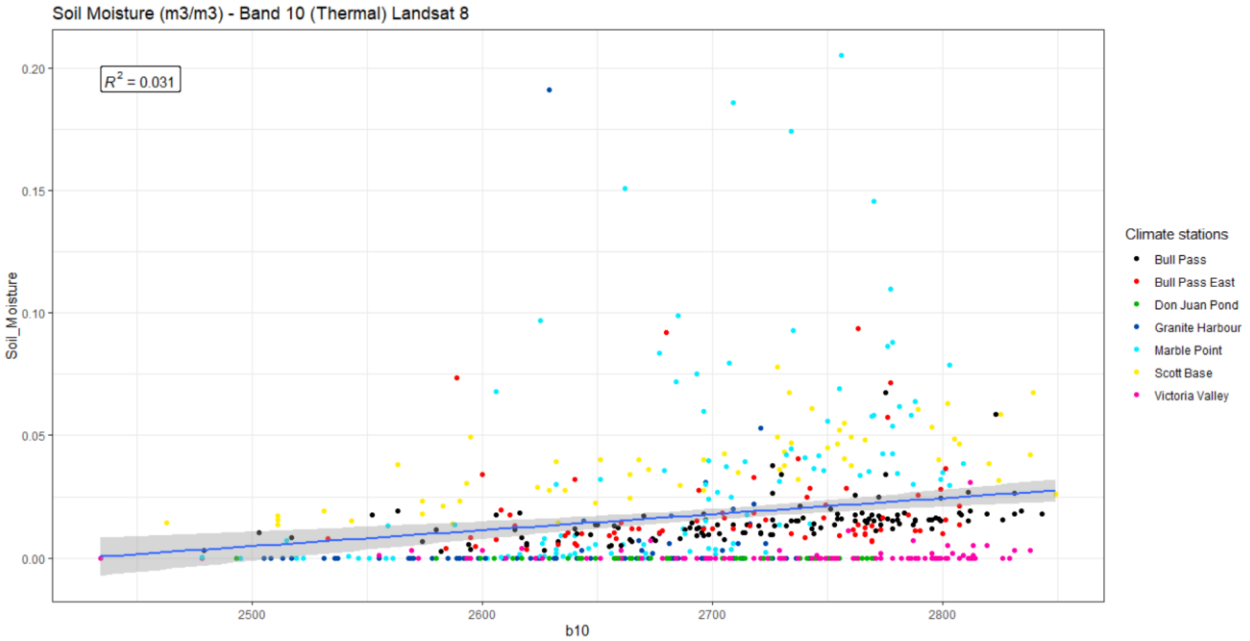

Fig. 14. Scatterplot of L8 TIRS 1 band and soil moisture values measured in all climate stations.

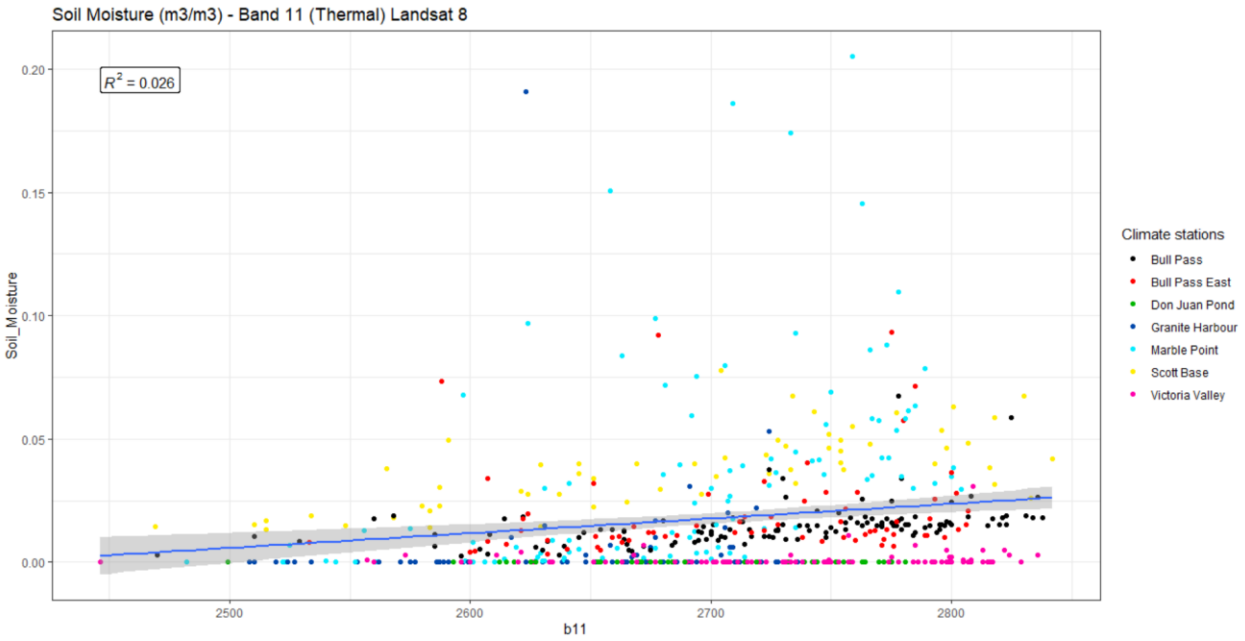

Fig. 15. Scatterplot of L8 TIRS 2 band and soil moisture values measured in all climate stations.

Soil moisture values and all climate stations involved except Don Juan Pond

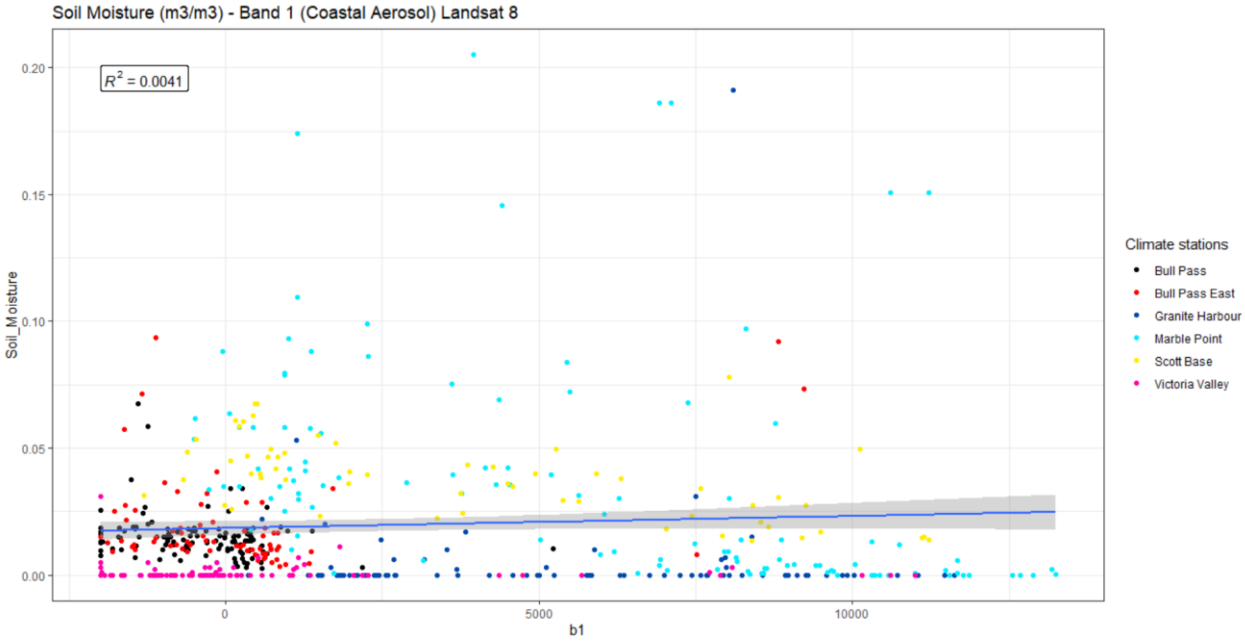

Fig. 16. Scatterplot of L8 Coastal Aerosol band and soil moisture values measured in all climate stations except Don Juan Pond.

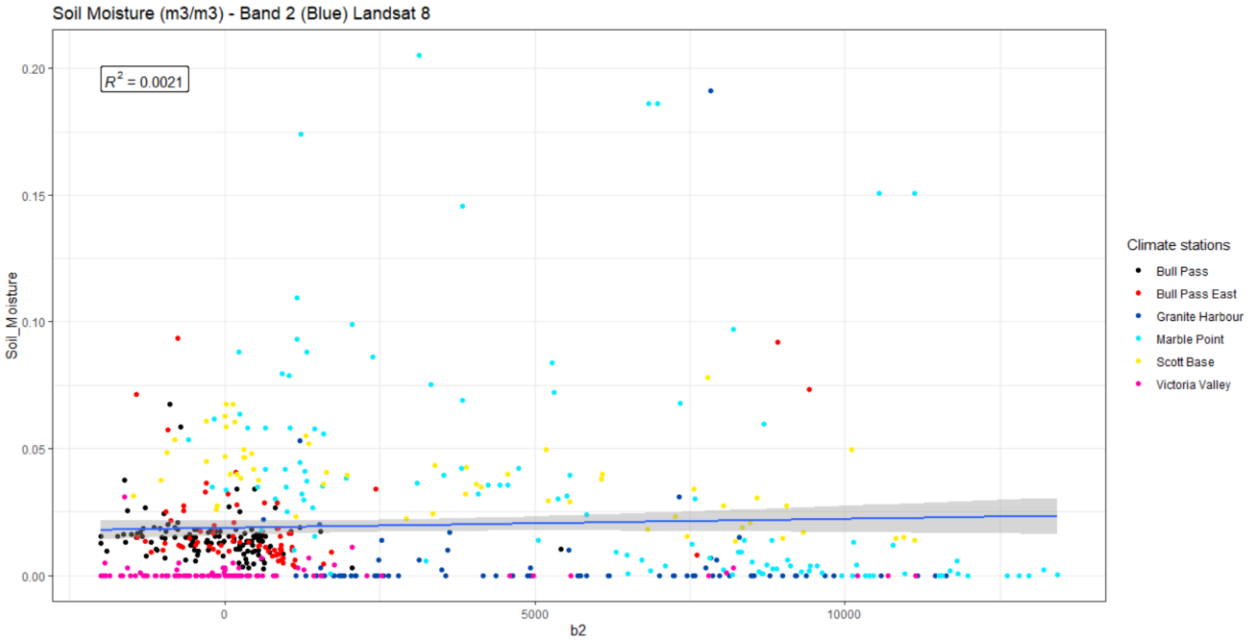

Fig. 17. Scatterplot of L8 Blue band and soil moisture values measured in all climate stations except Don Juan Pond.

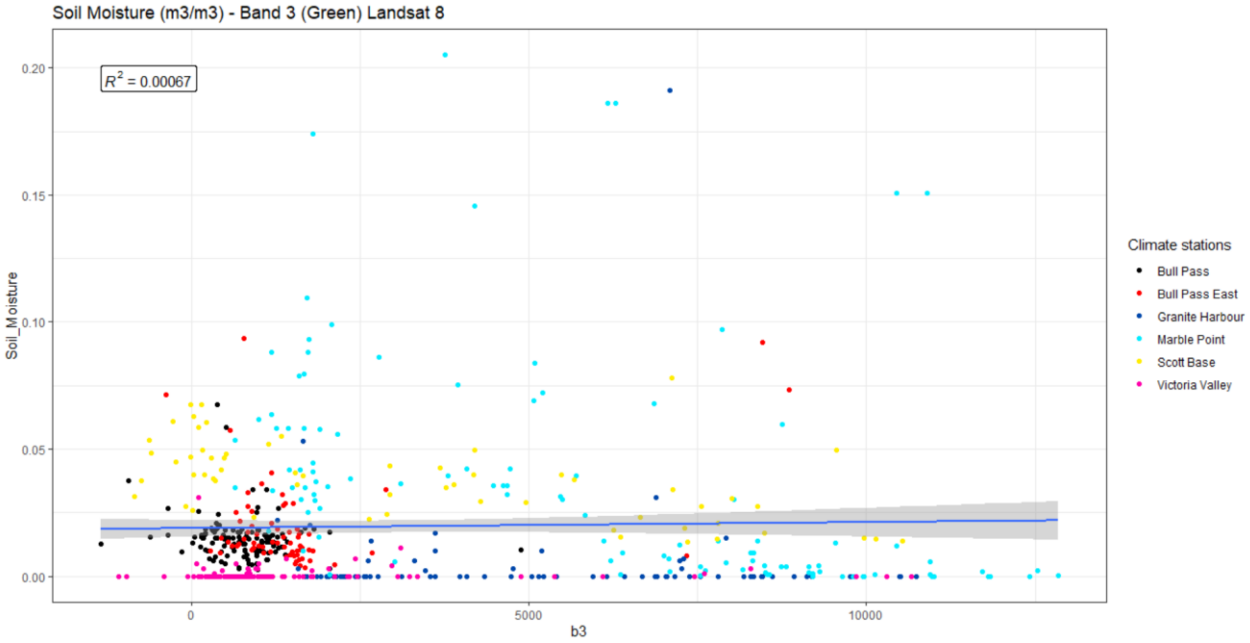

Fig. 18. Scatterplot of L8 Green band and soil moisture values measured in all climate stations except Don Juan Pond.

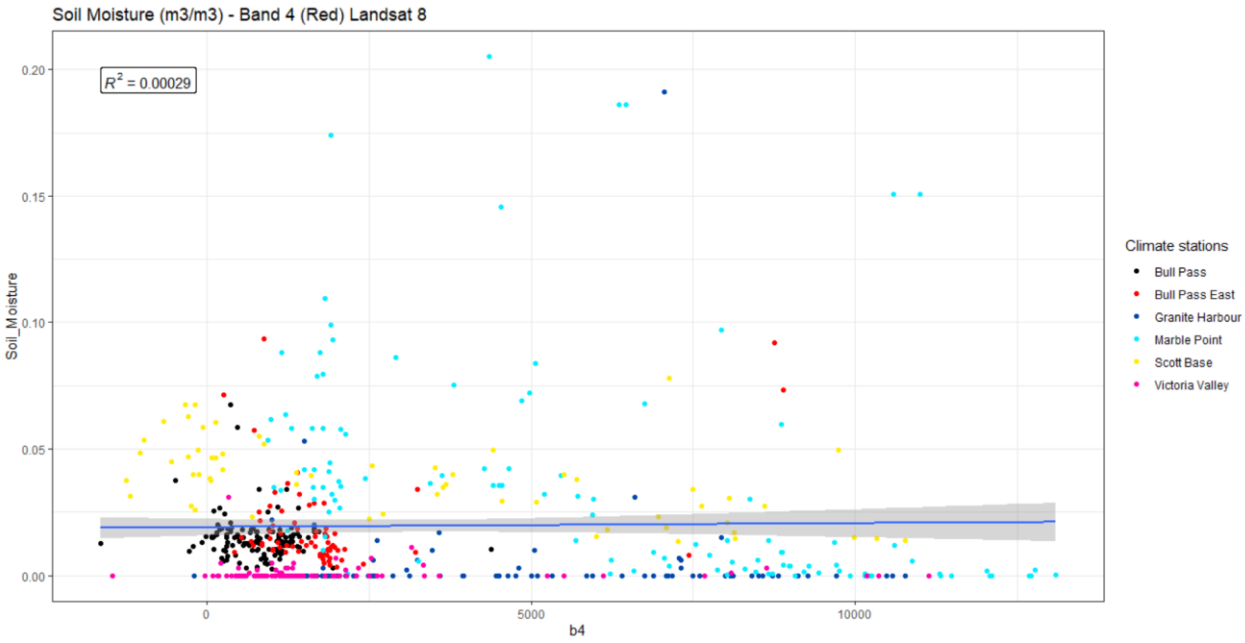

Fig. 19. Scatterplot of L8 Red band and soil moisture values measured in all climate stations except Don Juan Pond.

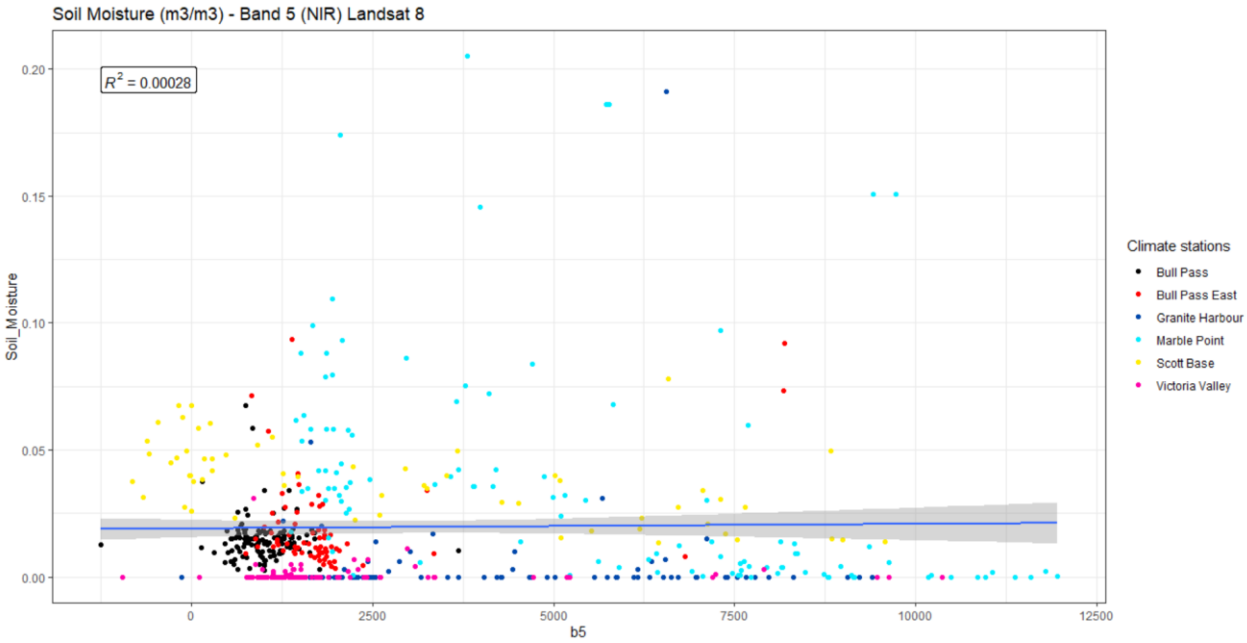

Fig. 20. Scatterplot of L8 NIR band and soil moisture values measured in all climate stations except Don Juan Pond.

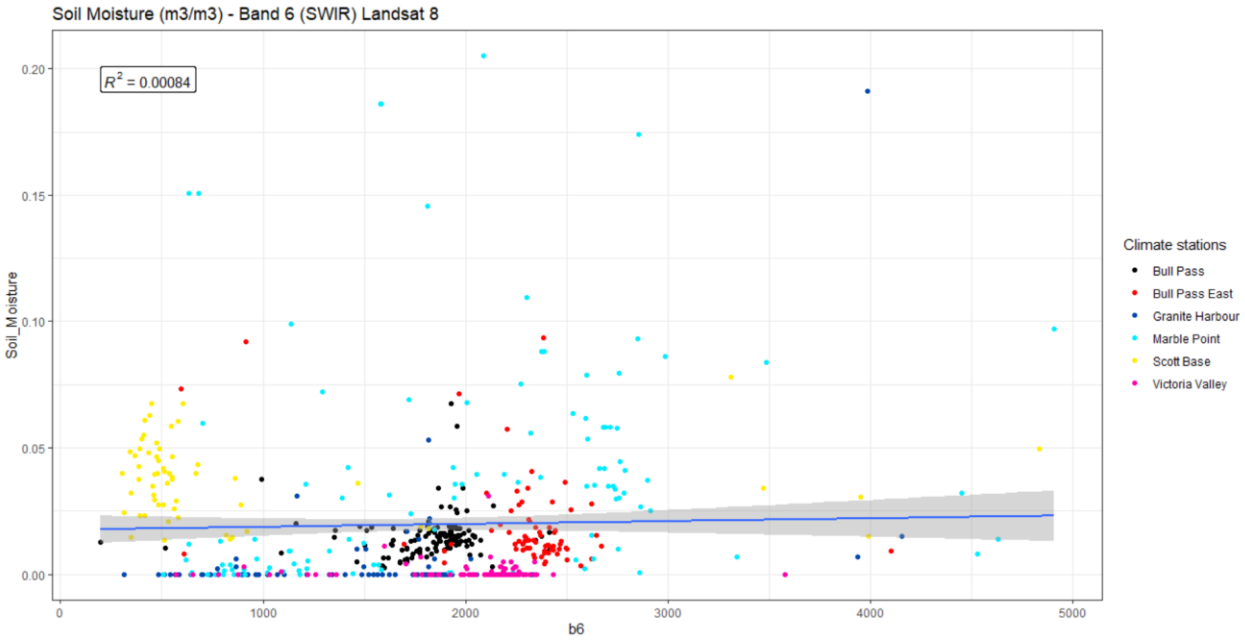

Fig. 21. Scatterplot of L8 SWIR 1 band and soil moisture values measured in all climate stations except Don Juan Pond.

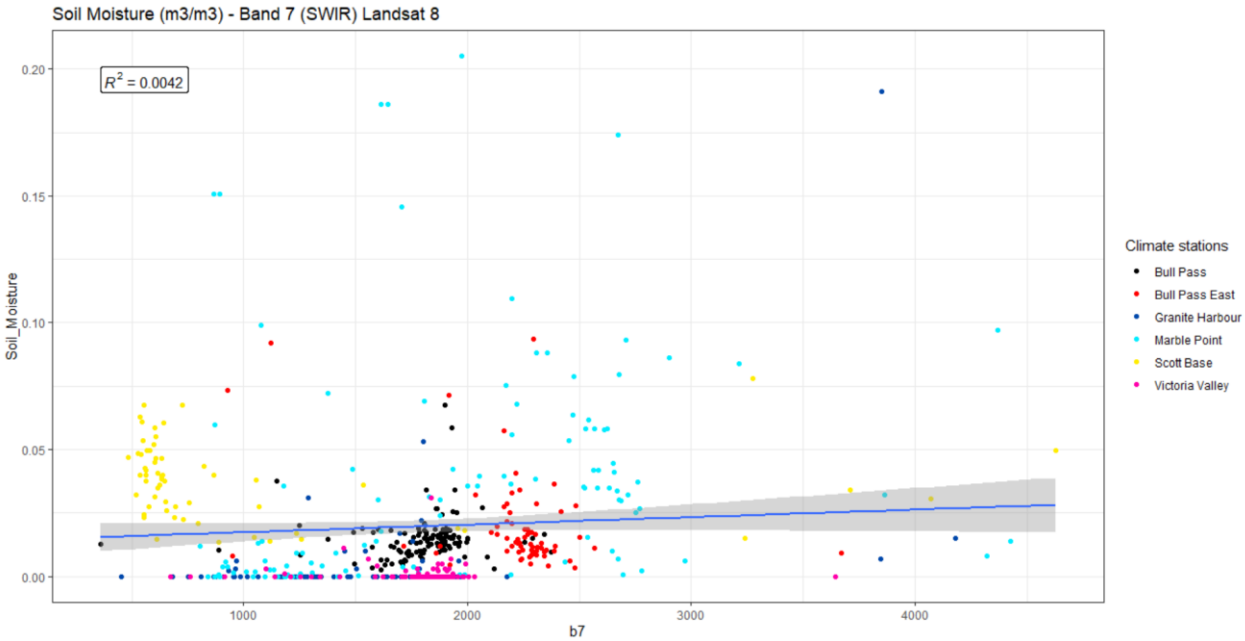

Fig. 22. Scatterplot of L8 SWIR 2 band and soil moisture values measured in all climate stations except Don Juan Pond.

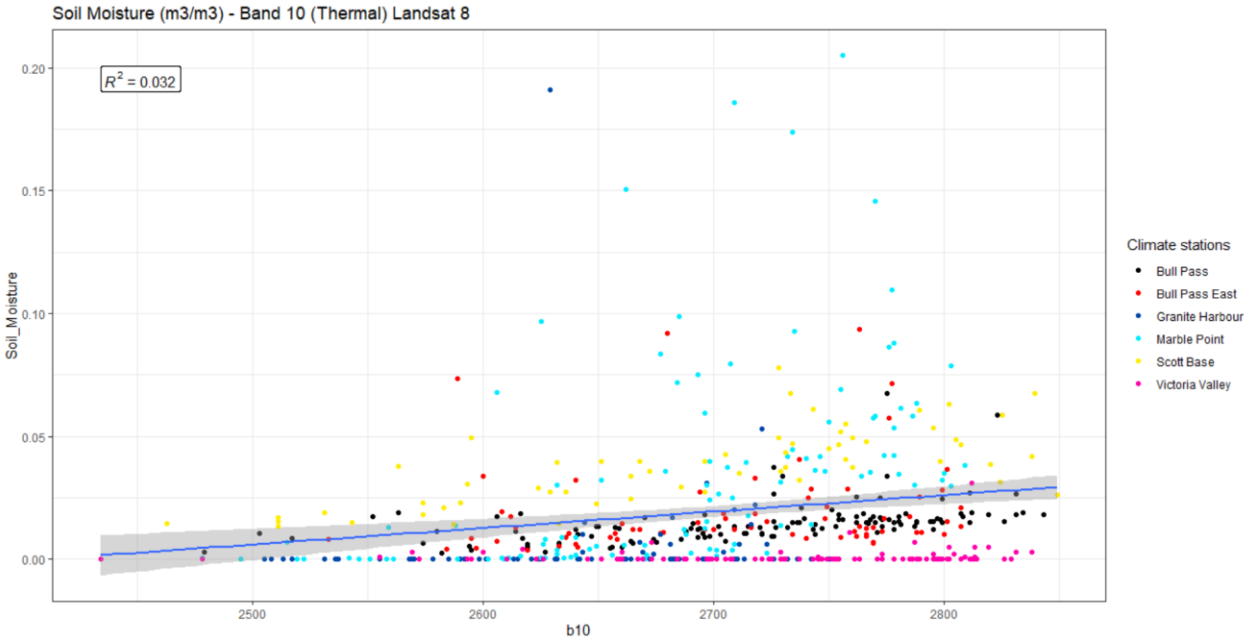

Fig. 23. Scatterplot of L8 TIRS 1 band and soil moisture values measured in all climate stations except Don Juan Pond.

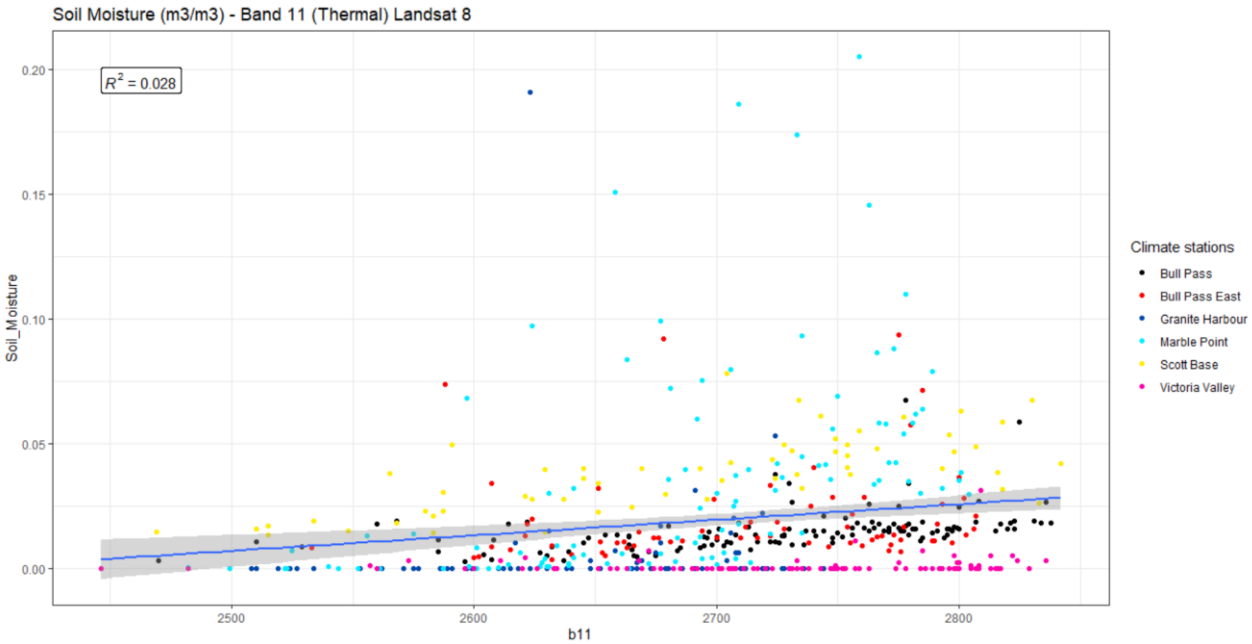

Fig. 24. Scatterplot of L8 TIRS 2 band and soil moisture values measured in all climate stations except Don Juan Pond.

Best performing scatterplots found for L8 SWIR 1, SWIR 2, TIRS 1 and TIRS 2 bands and soil moisture values looking into the climate stations individually

SWIR bands - Granite Harbour

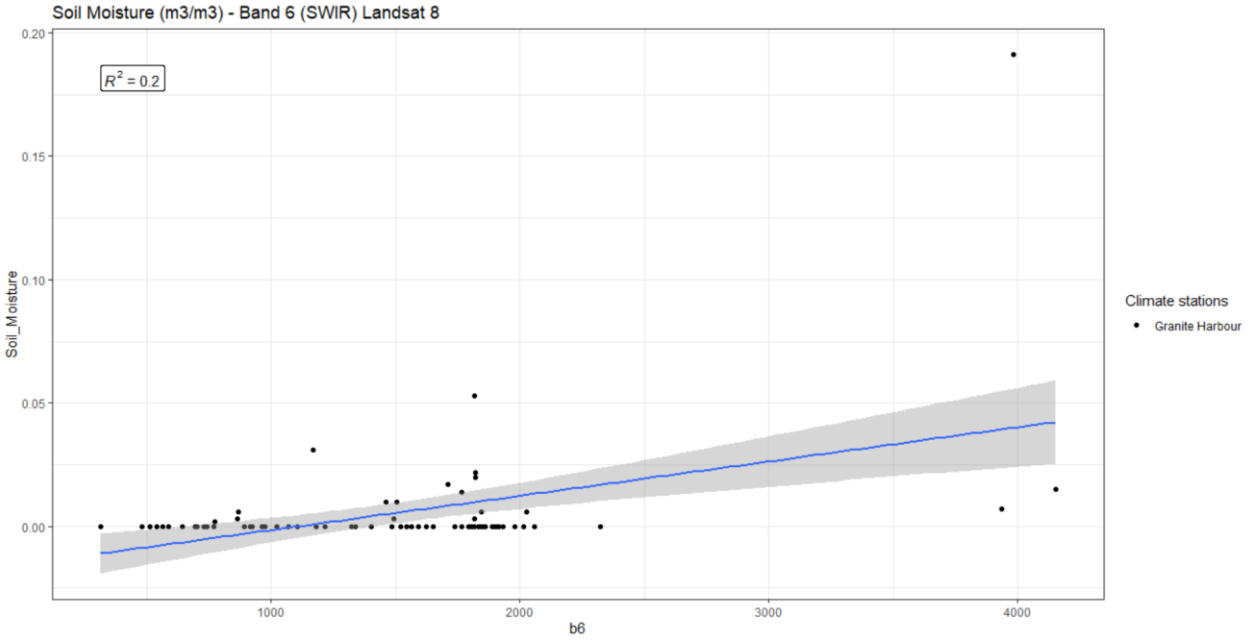

Fig. 25. Scatterplot of L8 SWIR 1 band and soil moisture values measured in Granite Harbour climate station.

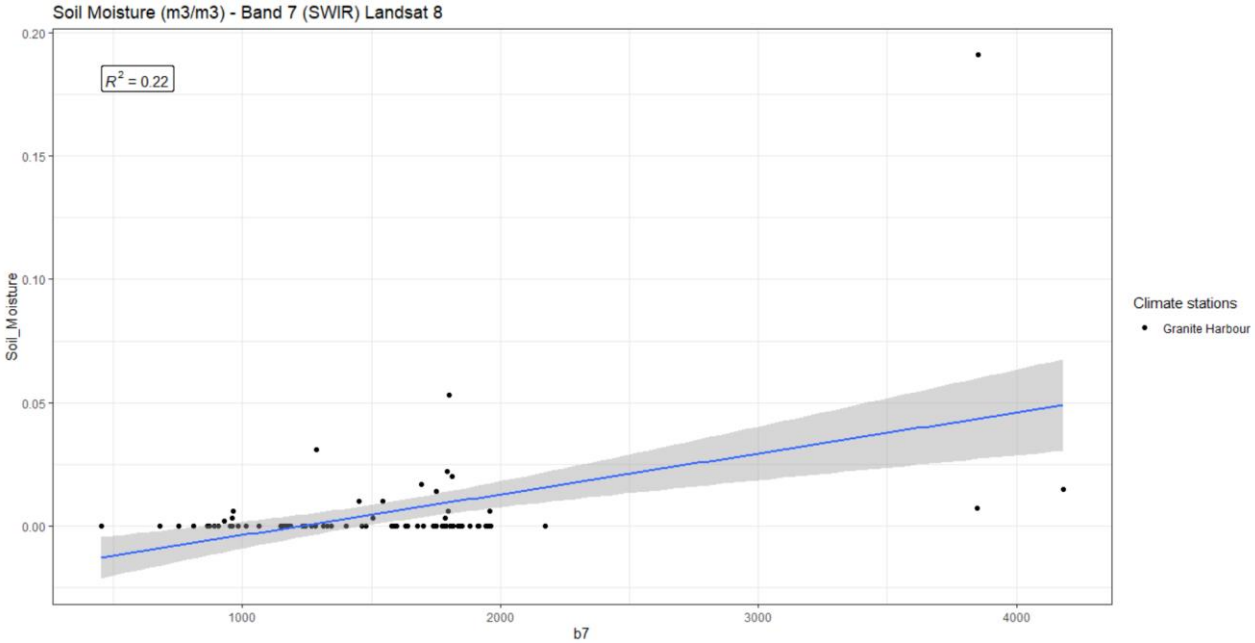

Fig. 26. Scatterplot of L8 SWIR 2 band and soil moisture values measured in Granite Harbour climate station.

Thermal Infrared bands - Scott Base

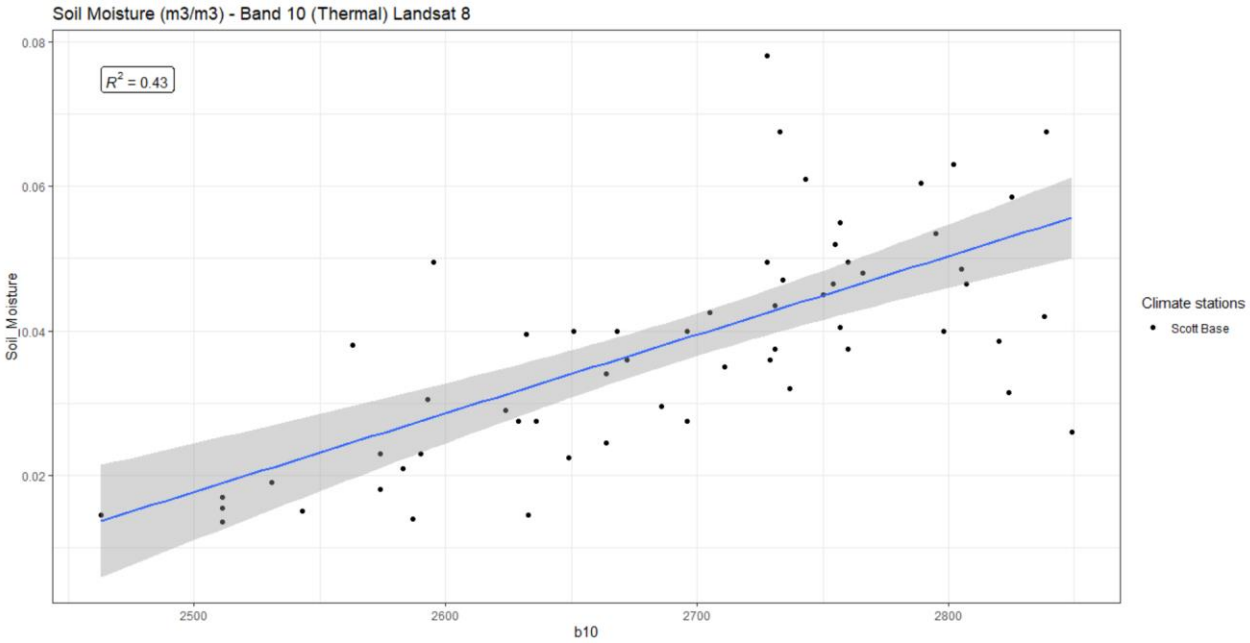

Fig. 27. Scatterplot of L8 TIRS 1 band and soil moisture values measured in Scott Base climate station.

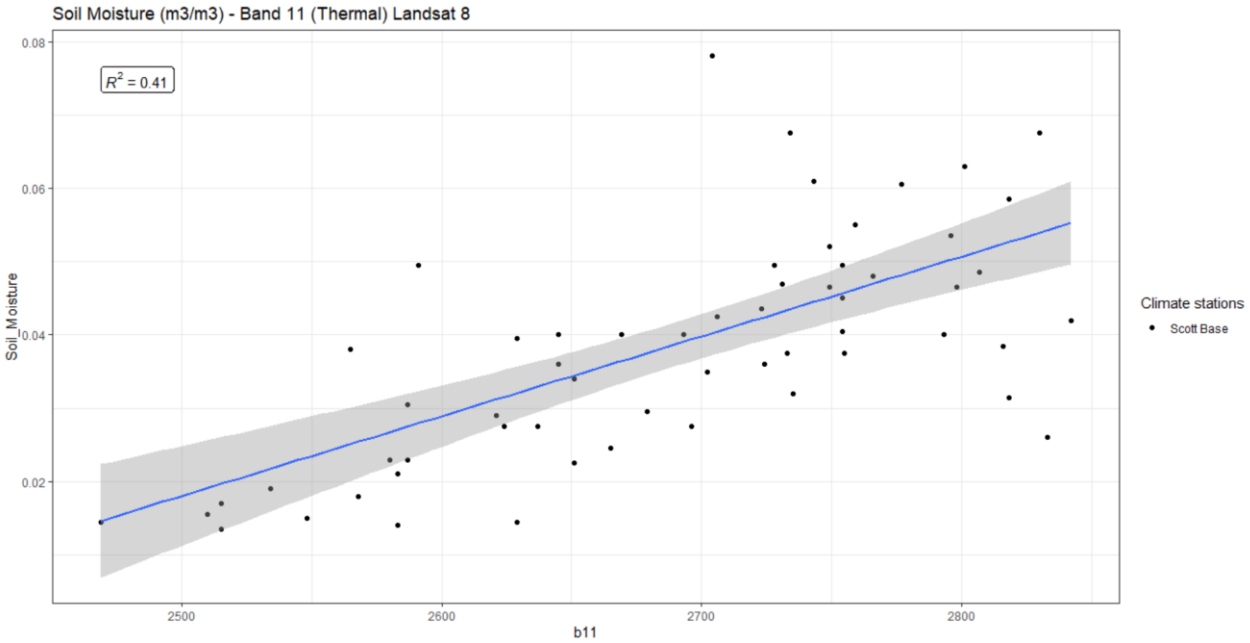

Fig. 28. Scatterplot of L8 TIRS 2 band and soil moisture values measured in Scott Base climate station.

3. Statistical Analysis – Linear Regression Model and Multiple Linear Regression

Models	Covariates														Evaluation
	Soil moisture	Band 1	Band 2	Band 3	Band 4	Band 5	Band 6	Band 7	Band 10	Band 11	All bands	Location	Elevation	Slope	R2
LRM1	x	x													0.009762
LRM2	x		x												0.006354
LRM3	x			x											0.003152
LRM4	x				x										0.001936
LRM5	x					x									0.001795
LRM6	x						x								0.0006869
LRM7	x							x							0.005356
LRM8	x								x						0.0286
LRM9	x									x					0.02557
MLR1	x	x	x	x											0.08924
MLR2	x	x	x	x								x	x	x	0.2913
MLR3	x	x	x	x	x	x									0.1168
MLR4	x	x	x	x	x	x						x	x	x	0.2939
MLR5	x	x	x	x	x	x	x	x							0.169
MLR6	x	x	x	x	x	x	x	x				x	x	x	0.2973
MLR7	x						x	x							0.05398
MLR8	x						x	x				x	x	x	0.2336
MLR9	x								x	x					0.06994
MLR10	x								x	x		x	x	x	0.2818
MLR11	x						x	x	x	x					0.149
MLR12	x						x	x	x	x		x	x	x	0.2856
MLR13	x										x				0.2446
MLR14	x										x	x	x	x	0.3187

Table 1. Total of LRM and MLR models performed in Statistical Analysis

4. Code

Spatial Analysis - Spatial Prediction

```
library(raster)
library(rgdal)
library(sp)
library(mapview)
library(satellite)
library(stringr)
library(lubridate)
library(sf)
library(rgeos)
library(magrittr)
library(dplyr)
options(stringsAsFactors = FALSE)
library(timechange)
library(XML)
library(xml2)
library(magrittr)
library(tmap)
library(methods)
library(devtools)
library(getSpatialData)
library(readxl)
library(espa.tools)
library(readr)
library(corrr)
library(ggpubr)
library(ggplot2)
library(tidyr)
library(car)
library(randomForest)
library(splines)
library(stats)
library(graphics)
```

```
library(caret)
library(spdep)

        ######################################################################
        ################## SOIL MOISTURE PREDICTION ##########################
        ######################################################################

# Read the data frame
df_final_all_climateStations <- read.csv("path_to_your_dataframe", head=T)

# 75% of total data used for training
df_75 <- floor(0.75*nrow(df_final_all_climateStations))

# Training data selected randomly
random_75 <- sample(seq_len(nrow(df_final_all_climateStations)), size = df_75)

training_df <- df_final_all_climateStations[random_75, ]
test_df <- df_final_all_climateStations[-random_75, ]

# Multiple Linear Regression Model
MLR_model_ALL <- lm(Soil_Moisture ~ b1 + b2 + b3 + b4 + b5 + b6 +
            b7 + b10 + b11, data = training_df)
summary(MLR_model_ALL)

# Get the RMSE of the MLR used
RMSE_predict <- sqrt(mean(MLR_model_ALL$residuals^2))


                ##########################################
                ################ VALIDATION ##############
                ##########################################

# Validate the test data (25%)
validation <- predict(MLR_model_ALL, newdata=test_df)
```

```
# Making data frame from the test data
test_df_1 <- cbind(test_df, validation)


                    ##############################################
                    ############ PREDICTING IN RASTER ############
                    ##############################################

# Read band images
b1 <- raster('D:/Thesis_WWU/Predictions_Raul/clip_band1_predict_AOI.tif')
b2 <- raster('D:/Thesis_WWU/Predictions_Raul/clip_band2_predict_AOI.tif')
b3 <- raster('D:/Thesis_WWU/Predictions_Raul/clip_band3_predict_AOI.tif')
b4 <- raster('D:/Thesis_WWU/Predictions_Raul/clip_band4_predict_AOI.tif')
b5 <- raster('D:/Thesis_WWU/Predictions_Raul/clip_band5_predict_AOI.tif')
b6 <- raster('D:/Thesis_WWU/Predictions_Raul/clip_band6_predict_AOI.tif')
b7 <- raster('D:/Thesis_WWU/Predictions_Raul/clip_band7_predict_AOI.tif')
b10 <- raster('D:/Thesis_WWU/Predictions_Raul/clip_band10_predict_AOI.tif')
b11 <- raster('D:/Thesis_WWU/Predictions_Raul/clip_band11_predict_AOI.tif')

# Create a data frame of all bands
b1_df <- as.data.frame(b1)
b2_df <- as.data.frame(b2)
b3_df <- as.data.frame(b3)
b4_df <- as.data.frame(b4)
b5_df <- as.data.frame(b5)
b6_df <- as.data.frame(b6)
b7_df <- as.data.frame(b7)
b10_df <- as.data.frame(b10)
b11_df <- as.data.frame(b11)

# Merge all bands in a data frame
df_raster <- cbind(b1_df, b2_df, b3_df, b4_df, b5_df, b6_df, b7_df, b10_df,b11_df)
```

```
# Add columns to match the model
df_raster$Soil_Moisture <- NA

df_raster$b1 <- df_raster$clip_band1_predict_AOI
df_raster$b2 <- df_raster$clip_band2_predict_AOI
df_raster$b3 <- df_raster$clip_band3_predict_AOI
df_raster$b4 <- df_raster$clip_band4_predict_AOI
df_raster$b5 <- df_raster$clip_band5_predict_AOI
df_raster$b6 <- df_raster$clip_band6_predict_AOI
df_raster$b7 <- df_raster$clip_band7_predict_AOI
df_raster$b10 <- df_raster$clip_band10_predict_AOI
df_raster$b11 <- df_raster$clip_band11_predict_AOI

# Predict from the raster data frame
prediction <- predict(MLR_model_ALL, newdata=df_raster)

# Doing a data frame from all data + prediction
test_df_1 <- cbind(df_raster, prediction)

# Create the raster from prediction
nc <- ncol(b10)
nr <- nrow(b10)

coords <- coordinates(b10)
lon <- coords[,1]
lat <- coords[,2]

crs <- crs(b10)

# Create a prediction raster data frame
prediction_df <- data.frame('lng' = lon,
                            'lat' = lat,
                            'predictions' = prediction)
```

```
# Create the prediction raster
prediction_raster <- rasterFromXYZ(prediction_df)

# Plot the raster soil moisture prediction
plot(prediction_raster)

# Save the raster soil moisture prediction performed
raster_prediction_tif <- writeRaster(prediction_raster,'your_path',options=c('TFW=YES'),
overwrite=TRUE)
```