EVALUATION OF SOIL MOISTURE ESTIMATES FROM SATELLITE-BASED AND REANALYSIS PRODUCTS OVER TWO NETWORK REGIONS

Kivana Naidoo

Submitted in fulfilment of the
requirements for the degree of MSc in Hydrology

Centre for Water Resources Research
School of Agriculture, Earth and Environmental Sciences
University of KwaZulu-Natal
Pietermaritzburg
21 July 2022

Supervisor:	Ms KT Chetty
Co-supervisor:	Dr S Gokool

In Memory of my Beloved Sister,

Shamentha Padayachee

ABSTRACT

The soil is an important variable of the hydrological cycle. It plays a key role in the distribution of water and energy fluxes between the surface and atmosphere. Soil moisture data can be used to develop early warning systems for flood and drought monitoring, improve weather and climate forecasting and provide an indication of crop water requirements. Therefore, the regular monitoring of this variable can prove to be beneficial to various management applications. One of the main issues associated with estimating soil moisture is to adequately account for its spatial and temporal variability as it is influenced by factors such as climate, topography, soil properties and land cover. There are different methods available to derive soil moisture estimations such as *in-situ*, remote sensing and modelling-based approaches. *In-situ* methods generally produce reliable soil moisture estimates, however, are only suitable for small scale studies. Alternatively, remote sensing and modelled reanalysis methods can provide soil moisture estimates over a large spatial extent, however, they are generally limited by their coarse spatial resolutions and may not be suitable for localised applications. Therefore, the aim of this study was to implement and evaluate a downscaling technique across two regions (South Africa and USA) to ultimately produce finer scale soil moisture and address the scale mismatch between *in-situ* methods and coarse resolution products. This procedure was facilitated by two data processing platforms, Google Earth Engine (GEE) and R, which showed significant potential for data processing and analysis. Additionally, satellite-based and reanalysis products were also evaluated to determine which of these methods are more suitable for soil moisture estimation. The soil moisture products and the downscaled products were validated against the CRNS instrument, which was particularly chosen for its performance at an intermediate spatial resolution. The SMAP_25 km product performed best at the Two Streams site and was selected for downscaling, whilst the CFSV2 product performed best at the Mead CSP3 and York Benny catchments and was chosen to be downscaled at both these sites. The results from the study indicated that the downscaled products for the Two Streams and Mead CSP3 sites performed better than the original products when compared to the CRNS data. The data acquired for the York Benny site revealed that the downscaled product performed similarly to the CFSV2 product. Therefore, downscaling does not always result in an improved outcome. However, from the results acquired for the Two Streams and Mead CSP3 study sites, it is evident that downscaling shows significant potential in producing better soil moisture estimates, which could be used to improve planning and management operations for various purposes.

DECLARATION PLAGIARISM

I, Kivana Naidoo, declare that:

(a) The research reported in this document, except where otherwise indicated, is my original work.

(b) This document has not been submitted for any degree or examination at any other university.

(c) This document does not contain text, data, figures, pictures, graphs or tables from another document, unless it is specifically acknowledged as being sourced the original document. Where other sources have been quoted, then:

   (i) their words have been paraphrased/re-written, and the general information attributed to them has been referenced, and

   (ii) where their exact words have been used, their writing has been placed inside quotation marks, and referenced.

(d) Where I have reproduced a publication, of which I am an author, co-author or editor, I have indicated, in detail, which part of the publication was actually written by myself alone and I have fully referenced such publications.

(e) This document does not contain text, graphics or tables copied and pasted from the Internet, unless they are specifically acknowledged and the source is detailed, both in the document and in the References section.

Signed: [redacted] …..

Kivana Naidoo

Supervisor: ……[redacted]……………Type text here

Ms Kershani Tinisha Chetty

Co-supervisor: …[redacted] ………

Dr Shaeden Gokool

ACKNOWLEDGEMENTS

First and foremost, I would like to thank God Almighty for granting me the opportunity to undertake research at a Masters level. Thank you for giving me strength and continuous guidance throughout the duration of this study.

This Masters Research Project titled "Evaluation of Soil Moisture Estimates from Satellite-based and Reanalysis Products over Two Network Regions" has been funded by the National Research Foundation (NRF). I wish to thank the aforementioned institution for providing me financial assistance to conduct research pertaining to this study.

I would also like to convey my sincere gratitude to the following people:

Ms KT Chetty, my supervisor. Thank you for the continuous guidance and encouragement throughout the duration of my postgraduate studies. I truly appreciate you for being so understanding and for the invaluable knowledge and assistance that you have given me all these years. Your experience and dedication have indeed inspired me to achieve my dreams.

Dr S Gokool, my co-supervisor. Thank you for your endless support and assistance during this study. This endeavour would not have been possible without the time and dedication that you have given towards this project. I am extremely grateful for all that I have learnt from you.

Dr T Vather. Thank you for the assistance and advice that you have given me towards this project and for the provision of validation data for the Two Streams study site.

Professor TE Franz. I would like to express my gratitude for the assistance and provision of validation data for the Mead and York Benny study sites.

Thank you to all to my colleagues and staff members at the Centre for Water Resources Research. I am so appreciative of all the knowledge and support I have acquired from you all.

I would like to sincerely thank my family, fiancé and friends for all the love, encouragement and support they have given me. This accomplishment would not have been possible without you all. To my parents, thank you for moulding me into the person that I am today and for always supporting me in my endeavours. I am truly blessed to have you both by my side.

To my late sister, I am so grateful for all the love and support you have given me. I wish to keep making you proud through every step of this journey. You will forever hold a special place in my heart.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

H_2	:	Hydrogen
N_2	:	Nitrogen
O_2	:	Oxygen
r	:	Correlation Coefficient
R^2	:	Coefficient of Determination

LIST OF ABBREVIATIONS

AMSR-2	:	Advanced Microwave Scanning Radiometer
ANN	:	Artificial Neural Network
ASCAT	:	Advanced Scatterometer
BAYE	:	Bayesian
BRT	:	Boosted Regression Trees
BEC	:	Barcelona Expert Center
BT	:	Brightness Temperature
CARET	:	Classification And REgression Training
CART	:	Classification and Regression Trees
CCI SM	:	Climate Change Initiative Soil Moisture
CFSV2	:	National Centers for Environmental Prediction (NCEP) Climate Forecast System
COSMOS	:	The Cosmic Ray Soil Moisture Observing System
CRNS	:	Cosmic Ray Neutron Sensor
Dispatch	:	Disaggregation based on Physical and Theoretical Scale Change
ECMWF	:	European Centre for Medium-Range Weather Forecasts
ERA5	:	Reanalysis 5
ESA	:	European Space Agency
GEE	:	Google Earth Engine
GIS	:	Geographic Information Systems
GLDAS2	:	The Global Land Data Assimilation System version 2
GLM	:	Generalized Linear Model
HRUS	:	Hydrological Response Units
JAXA	:	Japan Aerospace Exploration Agency
KLM	:	Keyhole Markup Language
KNN	:	K Nearest Neighbour
LANDSAT	:	Land Remote-Sensing Satellite (System)
LST	:	Land Surface Temperature
MAE	:	Mean Absolute Error
MLR	:	Multiple Linear Regression
MODIS	:	Moderate Resolution Imaging Spectroradiometer

NASA	:	National Aeronautics and Space Administration
NCEP/NCAR	:	National Centers for Environmental Predication/ The National Center for Atmospheric Research
NDVI	:	Normalized Difference Vegetation Index
NIR	:	Near Infrared Band
OPTRAM	:	Optical Trapezoid Model
RFF	:	Ranger Random Forest
RFI	:	Radio-Frequency Interference
RMSE	:	Root Mean Square Error
ROI	:	Region of Interest
RPart	:	Recursive Partitioning and Regression Trees
SMAP	:	Soil Moisture Active Passive
SMOS	:	Soil Moisture and Ocean Salinity
STR	:	Shortwave Infrared Transformed Reflectance
SVM	:	Support Vector Machine
SWAT	:	Soil Water Assessment Tool
TDR	:	Time Domain Reflectometry
TDT	:	Time-Domain Transmissivity
TOTRAM	:	Thermal-Optical Trapezoid Model
ubRMSE	:	unbiased Root Mean Square Error

1. INTRODUCTION

The availability of water resources to fulfil socio-economic and environmental needs is continuously decreasing, with this situation being intensified by global issues such as climate change and population growth. Therefore, monitoring variables that influence water resources management decisions is necessary to ensure its sustainability in the long term (Franz *et al*., 2016). The soil is one such variable that is a crucial component of the hydrological cycle (Duygu and Akyürek, 2019). It is an important variable across several sectors including agriculture, hydrology and ecology owing to its significant influence on water-related variables such as evaporation, transpiration and runoff (Lakshmi, 2012; Myeni e*t al*., 2019).

Therefore, the soil plays a key role in the distribution of water and energy fluxes between the land and atmosphere (Peng *et al*, 2017). As a result, soil moisture estimation has several applications relating to water resources management. It can aid in improving weather and climate forecasting and can also be used to develop early warning systems for flood and drought monitoring (Entekhabi e*t al*., 2010). In addition, soil moisture data can be used to provide a better indication of crop water requirements, which will result in more efficient irrigation scheduling options and increased crop yields (Myeni *et al*., 2019). These examples demonstrate the importance of continuously monitoring and measuring soil moisture content.

One of the main issues pertaining to soil moisture is the fact that it is both spatially and temporally variable as it is sensitive to several variables including topography, vegetation, temperature etc. (Vather, 2015; Myeni *et al*., 2019). In general, there are three methods that are currently used for soil moisture estimation i.e. *in-situ* methods, remote sensing and the use of hydrological models (Brocca *et al*., 2017). The aforementioned methods vary in many ways each with their respective advantages and disadvantages, which are important to acknowledge when determining the suitability of a method for a specific study or task.

In-situ methods for measuring soil moisture involve the use of ground instruments that mostly provide point-based observations (Zreda e*t al*., 2012). Soil moisture data acquired from *in-situ* methods generally have a greater degree of accuracy when compared to other methods (Peng *et al*., 2017). However, a major limitation associated with conventional *in-situ* methods is their inability to account for the spatial and temporal variability of soil moisture due to their small footprint (Zreda e*t al*., 2012).

The use of remote sensing methods for soil moisture estimation has become very common in recent years and it can be used to estimate soil moisture over large areas in a short period of time and can provide continuous data records that can be used for monitoring long-term temporal variations in soil moisture (Brocca *et al*., 2017; Kim *et al*., 2018; Chen *et al*., 2019). There are numerous satellite-based soil moisture products that are currently available and can be used to acquire information pertaining to soil moisture. Several studies have evaluated these products and demonstrated how this data can be used for hydrological applications (Peng *et al*., 2015, Brocca e*t al*., 2017, Montzka *et al*., 2017).

The use of remote sensing techniques for the estimation of soil moisture has been shown to have significant potential for hydrological applications (Brocca *et al*., 2017). However, these techniques are limited by their ability to adequately represent spatial heterogeneity due to their coarse spatial resolution (Peng *et al*., 2017). Subsequently, the accuracy of the estimates provided by these products needs to be determined prior to their use for research applications (Myeni *et al*., 2019).

Traditionally, the accuracy of these remote sensing derived soil moisture products is established through comparisons against measurements acquired from conventional *in-situ* techniques (Schrön *et al*., 2017). However, differences in the spatial scale for which these estimates are representative may contribute to further uncertainty regarding the accuracy of remote sensing soil moisture estimates (Vather e*t al*., 2018; Franz *et al*., 2020).

Modelling is also a common method used for soil moisture estimation (Brocca e*t al*., 2017). It can simulate soil moisture at different spatial scales using the required input data (Uniyal *et al*., 2017). There are many models that are available for soil moisture estimation including hydrological models and models that use satellite derived data as an input (Sagedhi *et al*., 2017; Pomeon e*t al*., 2018). Modelling of data is considered useful for monitoring soil moisture at a catchment scale, provided that the required input data is available. However, uncertainties arise due to the input data that is being used and the model itself (Brocca *et al*., 2017).

Modelling is also incorporated in the development of reanalysis soil moisture products (Robock *et al*., 2004). These products are developed though the integration of observations and modelled data, therefore, are primarily used for studying climate and present weather patterns (Bezak *et al*., 2021). Advantages associated with reanalysis products include the acquisition of data on a global scale and long data records (Robock *et al*., 2004). However, similarly with satellite-based soil moisture products, reanalysis products are also generally associated with coarse spatial resolutions and require validation (Cheng *et al*., 2019).

To reduce coarse resolution products to a finer resolution, in attempt of minimizing spatial scale differences, downscaling methods can be used (Kim *et al*., 2018). This can potentially improve the quality of data that is produced (Peng *et al*., 2017). There are several methods that can be used to downscale soil moisture products including statistical approaches and optical techniques (Peng *et al*., 2017). Despite downscaling being identified as a suitable approach, there is often insufficient field data to train and test algorithms (Paul and Singh, 2020). In addition, remote sensing data access and processing can be resource and computationally intensive and generally requires technical expertise in multiple aspects such as data science, remote sensing, coding etc. (Gorelick *et al*., 2017). These aforementioned factors will ultimately limit many researchers from applying downscaling methods.

However, with advancements in geospatial cloud computing using Google Earth Engine (GEE), coupled with the ease of application of advanced machine learning techniques facilitated by freely available software such as R, many of these previous barriers are now removed. GEE is a geospatial platform that is used for the acquisition and analysis of remotely sensed data (Mutanga and Kumar, 2019). The initiatives of this platform include addressing global issues such as droughts, floods, famine and disease, and removing access and capacity barriers for a wide range of researchers (Gorelick *et al*., 2017). The latter initiative is achieved due to the fact that large amounts of data can be easily accessed and manipulated over a short space of time without the requirement of supercomputing resources (Gorelick *et al*., 2017). As a result, this platform has advanced scientific research and is ideal for earth observation studies (Mutanga and Kumar, 2019).

R is a freely accessible tool used for statistical analysis and machine learning (Bischl *et al*., 2016). Statistical analysis is command driven having readily available packages that automatically carry out their respective functions (Tippmann, 2015). The number of packages available have increased significantly over the years. This is due to the fact that the source code is available to all users allowing researchers on a global scale to create packages, which can be accessed by scientists for future research purposes (Tippmann, 2015). Due to R being open source and its ability to automatically perform statistical analyses, it has become a well-known tool for a vast range of studies (Bischl *et al*., 2016).

While the applicability of downscaling has become easier owing to these platforms, validation is still required to assess the performance of the data that is produced. However, similarly to the issue associated with soil moisture products, conventional *in-situ* methods for estimating soil moisture generally operate over small footprints and may not be a reliable validation approach as downscaled data generally still has a coarse spatial resolution when compared to *in-situ* measurements.

In recent times, the use of the Cosmic Ray Neutron Sensor (CRNS) has gained much attention as a potential solution to address the aforementioned limitation. The CRNS is a recent *in-situ* instrument that provides area averaged soil moisture at an intermediate scale (Zreda *et al*., 2012; Vather *et al*., 2018). The CRNS measures the number of neutrons that are found above the surface, which is used to derive the soil moisture content (Vather e*t al*., 2018). The concentration of hydrogen atoms in the soil has a significant impact on the neutrons that reach the soil surface, therefore, it plays a key role in deriving the soil moisture content (Franz *et al.,* 2012). This instrument shows significant potential in bridging the gap between *in-situ* methods and remote sensing as it can measure soil moisture over a large area at a reliable resolution (Zreda e*t al*., 2008).

In-situ methods for soil moisture are highly reliable due to their high-resolution observations, however, their applicability is limited due to their small footprint diameter. Remote sensing observations and reanalysis data occur over much larger spatial extents and allow for continuous monitoring of soil moisture, however, these methods are unable to adequately account for spatial heterogeneity due to the coarse resolutions associated with available soil moisture products. Therefore, while field scale soil moisture estimates may be unsuitable for hydrological modelling applications conducted at catchment scales or larger, remote sensing observations and reanalysis data may be too coarse for field scale applications. The application of downscaling to soil moisture products could possibly mitigate this issue. While the processing associated with downscaling has made it difficult for researchers in the past to carry out such studies, platforms such as GEE and R have made it possible to overcome these complexities. Therefore, the use of these platforms makes downscaling a favourable method to produce soil moisture data at a finer resolution.

As a result, the purpose of this study is two-fold. Firstly, it focuses on downscaling coarse resolution soil moisture through the implementation of different data processing platforms. Secondly, an evaluation on satellite-based, reanalysis and downscaling approaches will be undertaken as there is currently a lack of knowledge as to which of these approaches would be more suitable for applications at larger spatial scales. This study will be undertaken over Two Network Regions, South Africa and Nebraska. The reason for conducting this study at different sites was to account for the influence of different climatic and land use conditions on soil moisture. The overall aim, objectives, research questions and hypothesis of this research study are detailed below:

1.1 Aim

The aim of this project is to develop and assess a downscaling approach for satellite-based and reanalysis soil moisture products using different data processing platforms

1.2 Objectives

- Identify and evaluate the use of currently available satellite-based and reanalysis soil moisture products
- Implement and assess a downscaling technique for satellite-based and reanalysis soil moisture products
- Evaluate the data processing platforms that were used for downscaling soil moisture
- Evaluate the performance of satellite-based soil moisture, reanalysis soil moisture and the downscaling technique using the Cosmic Ray Neutron Sensor

1.3 Research Questions

- How does the estimates obtained from each satellite-based and reanalysis soil moisture product compare with the Cosmic Ray Neutron Sensor?
- Does the application of a downscaling technique to the satellite-based and reanalysis data produce a better outcome?
- Are the data processing platforms used for downscaling soil moisture feasible?
- How does the soil moisture data obtained from all three methods (i.e. Satellite-based, reanalysis and downscaled estimates) compare across two different regions given their different climate and land use?

1.4 Research Hypothesis

Downscaling of satellite-based and reanalysis products is likely to improve their performance due to the estimates being produced at a finer resolution, therefore, resulting in more representative soil moisture data

1.5 Organization of dissertation

This dissertation consists of a total of six chapters. The first chapter provides a background and significance of the study followed by the overall aim, objectives and research questions. The second chapter is a review of all literature that was sourced during this study, which allowed for gaps in research to be identified. The third chapter entails the methodology of this project including the study site description and the methods that were used for data acquisition and analysis. The fourth chapter provides the results that were found for this study followed by a detailed discussion relating to all findings in the fifth chapter. The sixth chapter concludes the research pertaining to this project and provides some recommendations for future studies.

2. LITERATURE REVIEW

This chapter of the study details the different methods associated with soil moisture estimation including *in-situ*, remote sensing and modelling approaches. The review of literature pertaining to these methods is then followed by an investigation into downscaling techniques.

2.1 Field-Based Approaches

Conventional methods for monitoring soil moisture use ground-based instruments or techniques to provide point measurements. These techniques can monitor soil moisture at different depths and the data acquired is generally accurate and reliable (Peng e*t al*., 2017). However, soil moisture is both spatially and temporally variable (Vather, 2015). Therefore, in order to continuously monitor soil moisture over large scales using *in-situ* methods, dense networks are required (Brocca e*t al*., 2017). The costs associated with setting up and maintaining these dense soil moisture networks is relatively high and in addition, setting up these networks may not be possible given the topographic and climatic characteristics of a specific location (Myeni *et al*., 2019). The CRNS is a relatively new *in-situ* method that provides soil moisture at an intermediate scale (Zreda *et al*., 2012). Therefore, this method can provide data over larger scales when compared to conventional *in-situ* methods and can play a significant role in reducing the gap between *in-situ* methods and remote sensing (Vather *et al*., 2018).

There are several conventional *in situ* methods available for soil moisture measurement. Some common methods that are currently being used include gravimetry, Time Domain Reflectometry (TDR) and tensiometers. The gravimetric method is one of the most common and oldest techniques that are used for soil moisture measurement (Johnson, 1962; Walker, 2004). The process starts off with collecting a soil sample, which is then weighed to determine the initial mass (Johnson, 1962). Thereafter, the soil sample is oven dried at 105°C for 24 hours to remove the water present in the soil (DeAngelis, 2007). The oven dried sample is then weighed once again and the weight loss is eventually derived using both masses (i.e. the initial and oven-dried masses), which allows for the moisture content present in the soil to be determined (Walker, 2004). This method is highly suitable for the calibration of data produced by other soil moisture methods due to its level of accuracy, however, monitoring soil moisture over large spatial scales is not recommended due to the time and labor requirements involved in collecting soil samples.

A TDR probe is a portable instrument that is used to indirectly measure soil moisture content (Edaphic Scientific, 2020). It comprises of a sensor and parallel rods, which are directly placed into the soil at the required depth (Su *et al*., 2014). An electromagnetic pulse is then launched by the sensor, which travels along the parallel rods and after seconds is reflected back to the TDR sensor (Su *et al*., 2014). The travel time of the pulse is used to determine the dielectric permittivity of the medium (Su *et* al., 2014). In general, an electromagnetic pulse tends to travel slower in wet soils when compared to drier soils (Edaphic Scientific, 2020).

A tensiometer is used to directly determine the matric potential of soil moisture (Smajstrla and Harrison, 1981). The instrument consists of a porous cup, which is attached to a transparent airtight tube and a vacuum gauge that is used to record the matric potential (Johnson, 1962). The porous cup is filled with water and the tensiometer is then inserted into the soil at the required depth. The water present in the porous cup comes into contact with the surrounding soil and eventually reaches equilibrium with the medium (Johnson, 1962; Su *et al*., 2014). The porous cup will lose or gain water depending on the moisture status of the surrounding soil (Su *et al*., 2014). These changes in moisture content in the tensiometer results in changes in the pressure, which is recorded by the vacuum gauge (Johnson, 1962; Smajstrla and Harrison, 1981). The soil moisture content is then indirectly derived using a moisture-tension curve, which relates soil moisture content to the matric potential of the soil (Johnson, 1962).

As mentioned above, conventional *in-situ* methods generally provide reliable soil moisture estimates due to their fine scale observations, however, there are also certain limitations associated with these methods. The advantages and disadvantages associated with the aforementioned methods are provided in the table below:

Table 2.1 Advantages and disadvantages of conventional methods for soil moisture

***In-situ* method**	**Advantages**	**Disadvantages**
Gravimetry	Direct form of measurement Simple and relatively inexpensive Independent of soil type and salinity	Labour intensive Time consuming Destructive sampling method
Time Domain Reflectometry	Rapid data acquisition Automated Implementation costs are relatively low	Probe must have good contact with the medium to prevent the formation of air gaps The measurement depth is restricted to the length of the parallel rods
Tensiometer	Easy to install Relatively inexpensive	Frequent maintenance required Degree of accuracy is dependent on soil type

The CRNS is a more recently developed *in-situ* method that is used to measure soil moisture at an intermediate scale (Zreda *et al*., 2008; Vather *et al*., 2018; Duygu and Akyürek, 2019). Thus, this instrument plays a key role in addressing scaling issues that are associated with conventional *in-situ* techniques (gravimetry, TDR, tensiometer, etc.) being used for the validation of remote sensing methods and models (Zreda *et al*., 2008). Due to the potential of this method, a number of studies have been conducted to assess the use of the CRNS for soil moisture estimation and its ability to bridge the measurement gap between conventional *in-situ* methods and remote sensing products. CRNS installations have been established in several countries including South Africa, USA, UK, China and Australia (Franz *et al*., 2016).

The probe measures the low energy neutrons in the surrounding air and soil, which is ultimately related to the moisture content of the soil (Zreda *et al*., 2012). Neutrons present in the soil and atmosphere are produced by cosmic rays originating from outer space (Vather *et al*., 2018). Cosmic rays containing high energy can pass through the earth's magnetic field, however, when they enter the atmosphere, they collide with nuclei present (Franz *et al*., 2016). As a result, nuclei colliding with cosmic rays are disintegrated into many secondary particles (Zreda *et al*., 2012). Secondary neutrons can move through the atmosphere resulting in more collisions occurring as they approach the earth's surface, thus producing fast-moving electrons (Vather *et al*., 2018).

Once these fast-moving neutrons reach the earth's surface, some neutrons will be absorbed by the soil and the rest of the neutrons will be scattered back into the atmosphere (Franz *et al*., 2016). Fast moving neutrons that collide with hydrogen, most likely in the form of water, lose energy and are absorbed by the hydrogen atoms, therefore, the presence of hydrogen strongly influences fast-moving electrons (Franz *et al*., 2012a; Franz *et al*., 2016). As a result, by measuring the neutron intensity above the land surface, an indication of the hydrogen content is given, which is important for deriving the soil moisture content (Zreda *et al*., 2008; Vather *et al*., 2018). As shown in Figure 2.1, wetter soils will generally result in fewer neutrons being emitted in the air (lower neutron intensity) due to a higher concentration of hydrogen (Vather *et al*., 2018).

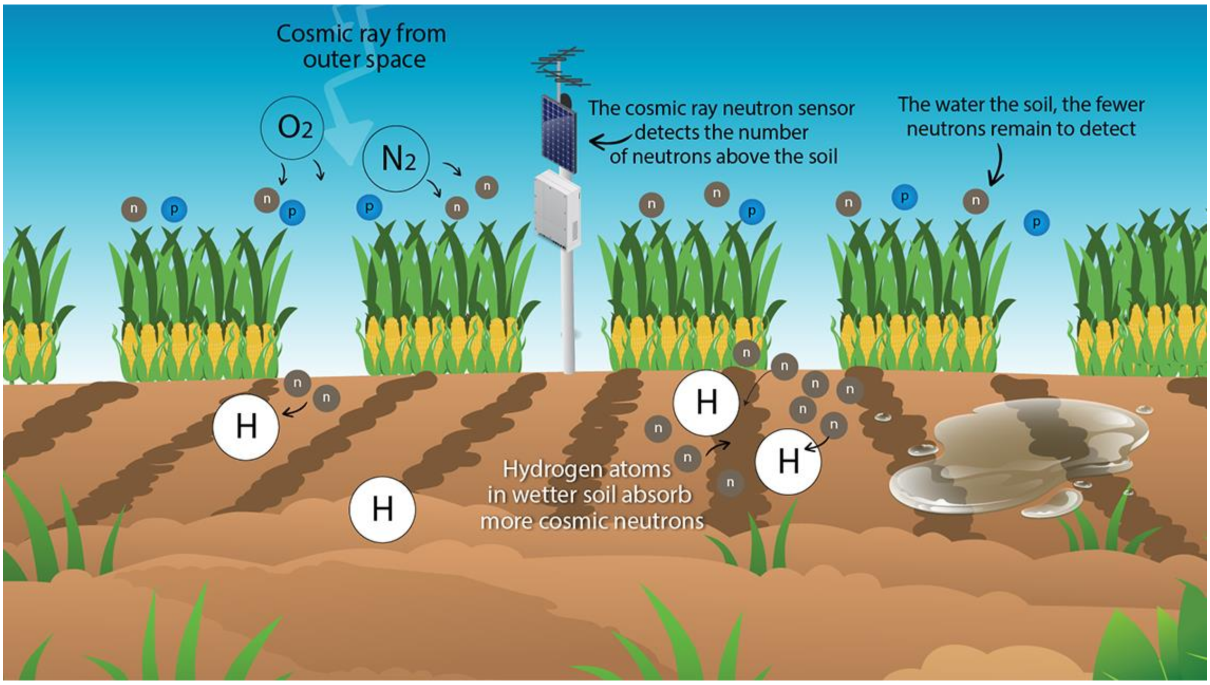

Figure 2.1 The use of Cosmic Rays for measuring soil moisture (IAEA, 2020)

A labeled diagram of the CRNS can be seen in Figure 2.2, which consists of solar panels and batteries used to power the instrument; a data logger to log the neutron counts and monitor atmospheric conditions such as pressure, temperature and humidity; neutron tubes to measure fast and slow-moving neutrons; a satellite link to enable data transfer check; and a tripod upon which the instrument is mounted for support (Vather *et al*., 2018)

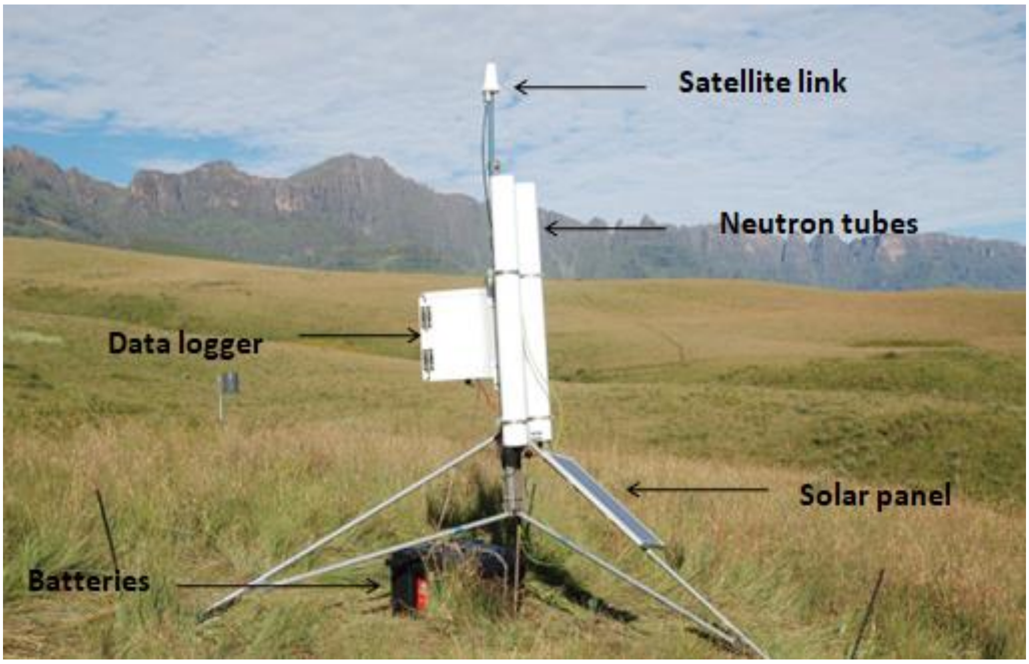

Figure 2.2 The Cosmic Ray Neutron Sensor located at the Cathedral Peak Catchment (Everson, 2016)

Conventional methods for soil moisture measurement generally produce more accurate data as compared to other available methods, however, these methods have small measurement footprints. The CRNS overcomes this limitation, as it produces area-averaged soil moisture data at an intermediate scale having a footprint diameter of ~240 m at sea level (Schron *et al*., 2017, Duygu and Akyürek, 2019). The footprint diameter varies depending on atmospheric conditions, such as air density and humidity, however, it is independent of the soil moisture content (Zreda *et al*., 2008; Franz *et al*., 2012). The soil moisture depth measured by the CRNS lies between 0.12 m to 0.70 m, however, this is strongly dependent on the soil moisture content in contrast to the footprint size (Vather *et al*., 2018).

The CRNS method is highly useful in bridging the gap between point-based methods and remote sensing owing to its ability to produce soil moisture data at an intermediate scale (Schron e*t al*., 2017). It can be used for the validation of soil moisture data acquired from satellite soil moisture products and for the calibration of land surface models that use satellite data as an input (Duygu and Akyürek, 2019). In addition to the CRNS producing data at an intermediate scale, other advantages include the fact that it is easily automated, portable and requires little energy to operate (Vather *et al.,* 2018). This method is non-invasive and independent of certain soil characteristics such as texture and bulk density (Zreda *et al*., 2012). The CRNS is not solely used for soil moisture monitoring, as it also has other applications such as monitoring changes in biomass and measuring snow-depth (Zreda *et al*., 2008; Vather *et al*., 2018).

A limitation associated with the CRNS is the fact that it requires calibration due to the method needing site specific parameters to infer soil moisture from the monitored neutron intensity (Vather *et al*., 2018). This process can be costly and requires technical expertise. There are currently three options available for calibrating the CRNS, with the most common option involving the use of area-average soil moisture that is acquired through the collection of several soil samples within the CRNS diameter (Vather *et al*., 2018). Once these samples are collected, gravimetric analysis i.e. oven-drying, is undertaken to produce area-averaged soil moisture datasets (Zreda *et al*., 2012). While this option for calibration is commonly used, it is considered to be time consuming and labor intensive due to the collection and analysis of a large number of soil samples (Vather *et al*., 2018). The second option involves the use of a calibration function that is developed through chemical analysis of soil samples (Zreda *et al*., 2012). The last option is the use of a universal function that is generally applied in cases where no soil samples are available for analysis (Zreda *et al*., 2012). Despite the need for calibration, the CRNS method shows great potential with regards to addressing scaling issues between point-based methods and remote sensing products.

2.2 Remote Sensing for Soil Moisture Estimation

Remote sensing methods can acquire soil moisture data over large spatial scales in a short duration (Brocca *et al*., 2017). In addition, they have the ability to provide data for remote areas, at which *in-situ* monitoring networks cannot be installed. Several satellite-based products are available for monitoring soil moisture at a global scale (Brocca *et al*., 2017, Kim *et al*., 2018). These include the Advanced Scatterometer (ASCAT), Soil Moisture and Ocean Salinity (SMOS), Soil Moisture Active Passive (SMAP), Advanced Microwave Scanning Radiometer (AMSR-2) and the Climate Change Initiative Soil Moisture (CCI SM).

2.2.1 Advanced Scatterometer (ASCAT)

The ASCAT soil moisture product was first launched in October 2006 by the European Space Agency (ESA) and thereafter, operated by the European Organization for the Exploitation of Meteorological satellites (EUMETSAT) (Al-Yaari *et al*., 2014). Since 2006, ASCAT has been launched onboard three Meteorological Operational satellites (MetOp): MetOp-A (2006), MetOp-B (2012) and MetOp-C (2018) (Mousa and Shu, 2019). The main purpose for launching these satellites was to acquire data for Numeral Weather Prediction (NWP).

The product consists of several flags that are provided with the soil moisture data to quantify uncertainties with regards to backscatter measurements and model parameterizations (Al-Yaari *et al*., 2014; Brocca *et al*., 2017). ASCAT monitors wind speed and direction over oceans, however, it also has other applications such as monitoring soil moisture, vegetation and sea ice extent. In addition, it is also an input into the ESA's Climate Change Initiative and Copernicus Climate Change Service (Brocca *et al*., 2017).

2.2.2 Soil Moisture and Ocean Salinity (SMOS)

The SMOS mission was launched in November 2009 by the ESA and is the second earth explorer opportunity mission following the Gravity Field and Steady-State Ocean Circulation Explorer (GOCE) satellite (Kerr *et al.*, 2012). The Microwave Imaging Radiometer using aperture synthesis (MIRAS) instrument onboard the SMOS satellite is used to acquire soil moisture data through its relationship with the emissivity of the earth's surface (Kerr *et al*., 2012).

The SMOS satellite was launched with the main purpose of monitoring soil moisture over land and sea surface salinity, which makes it ideal for improving weather forecasting systems (El Hajj *et al*., 2018). However, one of the main challenges and uncertainties associated with SMOS is the use of the MIRAS instrument as it is newly developed (El Hajj *et al*., 2018).

2.2.3 Soil Moisture Active Passive (SMAP)

The SMAP satellite was initiated by The National Aeronautics and Space Administration (NASA) and launched in January 2015 (Chan *et al.,* 2016). The main mission was to measure soil moisture and be able to differentiate between whether the areas being mapped are in a freeze or thaw state (Entekhabi *et al*., 2010). SMAP carried both a radar and radiometer that were both used for mapping soil moisture, however, the radar instrument had a hardware failure in 2015, shortly after the satellite was launched (Mousa and Shu, 2019). As a result, the radar was no longer available for use, making the radiometer the only instrument onboard the SMAP satellite to be used for monitoring the earth's surface (Chan *et al*., 2016).

The SMAP mission was launched with the aim of acquiring surface soil moisture through modelling and data assimilation (Entekhabhi *et al.,* 2010). In addition, the application of SMAP data from a broader perspective would be for better monitoring the water cycle, improving weather forecasting systems as well as agricultural productivity.

2.2.4 Advanced Microwave Scanning Radiometer (AMSR-2)

The AMSR-2 product was launched by the Japan Aerospace Exploration Agency (JAXA) in May 2012 onboard the Global Change Observation Mission 1–water (GCOM-W1) (Cui *et al*., 2017). It is the successor to the AMSR-E product that was launched in May 2002 onboard the Aqua satellite (Peng *et al*., 2015). However, due to issues with the antenna of AMSR-E, it was no longer operational from 2011, therefore, initiating the development of AMSR-2 (Griesfeller *et al*., 2016). The AMSR-2 product is relatively similar to AMSR-E, however, there were some modifications made with the aim of improving and overcoming some limitations that were associated with AMSR-E (Parinussa *et al*., 2014).

AMSR-2 provides numerous data pertaining to the hydrological cycle including soil moisture, rainfall, sea surface temperature and snow depth (Parinussa *et al*., 2014). The data acquired from this mission can be used to gain knowledge surrounding climate change and also improve weather forecasting.

2.2.5 Climate Change Initiative Soil Moisture (CCI SM)

The ESA CCI is a merged soil moisture product that was created by the ESA using a combination of data acquired from both active and passive microwave sensors (Cui *et al*., 2017). The passive microwave sensors used include the Scanning Multichannel Microwave Radiometer (SMMR), the Special Sensor Microwave Imager (SSM/I), the Advanced Microwave Scanning Radiometer (AMSR-E/AMSR-2) and the soil moisture and ocean salinity (SMOS) product (Ciabatta *et al*., 2017). The active microwave sensors that are used include the European Remote Sensing Satellites (ERS-AMI) and the Advanced Scatterometer

(ASCAT) product (Dorigo *et al*., 2014; Ciabatta *et al*., 2017). These active and passive sensors were used to develop three different products i.e. a passive microwave product, an active microwave product and a product that combines both active and passive microwave soil moisture retrievals (Peng *et al*., 2015).

The ESA CCI has a climate data record exceeding 40 years with the passive microwave product extending from 1978 – 2019, the active microwave product extending from 1991 – 2019 and the merged active/passive product operating from 1978 – 2019 (Cui *et al*., 2017). Soil moisture data is acquired from this product at a daily time-step with a spatial resolution of 25 km (Peng *et al*., 2015). The main mission of the ESA CCI product is to develop a long and complete data record for an improved understanding of the earth's climate.

Each of these aforementioned soil moisture products can be distinguished based on properties such as their frequency, orbit, resolution and swath width (Mousa and Shu, 2019). Their frequencies are characterized using X, C and L bands (Kerr *et al*., 2012; Al-Yaari *et al*., 2014). Satellite products that carry an L band radiometer have the ability to acquire soil moisture data at a greater depth as compared to other satellite products (Mousa and Shu, 2019). Common organizations responsible for launching these satellite products include the European Space Agency (ESA), National Aeronautics and Space Administration (NASA), Barcelona Expert Center (BEC) and the Japan Aerospace Exploration Agency (JAXA). The properties of different satellite soil moisture products are listed in Table 2.2.

Table 2.2 Properties associated with different satellite soil moisture products

Satellite Product	Launch	Spatial Resolution	Temporal Resolution	Band Type	Altitude	Swath Width	Purpose
ASCAT	ESA, October 2006	25-50 km	2 days	C-band	817 km	550 km	Numerical Weather Prediction
SMOS	ESA, November 2009	35-60 km	3 days	L-band	757 km	900 km	Soil moisture and sea surface salinity
SMAP	NASA, January 2015	36-40 km	2-3 days	L-band	685 km	1000 km	Surface soil moisture through modelling and data assimilation
AMSR-2	JAXA, May 2012	10-25 km	1 day	C-band; X-band	700 km	1450 Km	Soil moisture, rainfall, seas surface temperature and snow depth
CCI SM	N/A	25 km	1 day	N/A	N/A	N/A	To develop long and complete data records

One of the limitations associated with remote sensing satellite-based products is their coarse resolution, which could potentially affect the performance of the data that is obtained (Peng *et al*, 2017). Numerous studies have been carried out till date to evaluate these soil moisture products, which has mainly been done through validation procedures using *in-situ* measurements (Djamai *et al*., 2016, Kim *et al*., 2018, Chen *et al*., 2019). However, the outcomes of some of these studies have highlighted the fact that the validation of these products using *in-situ* measurements are not reliable owing to the spatial scale differences between both methods. Studies have also been undertaken to overcome the issue of coarse resolutions associated with remotely sensed products using downscaling approaches. However, from these studies, it is evident that downscaling approaches have yielded mixed reviews. (Tian *et al*., 2016, Peng *et al*., 2017).

2.3 Modelling of Soil Moisture

There are several models that are available for simulating soil moisture. This includes hydrological models and models that use satellite data as an input (Sagedhi e*t al*., 2017). Models can provide data at different scales, however, are more commonly used for studies at a catchment scale (Pomeon *et al*., 2018).

2.3.1 Models for soil moisture estimation

Several models have been studied for the estimation of soil moisture (Pomeon *et al*.,2017; Sagedhi *et al*., 2017; Babaeian *et al*., 2018). Different models vary in terms of their structure and scale at which they produce soil moisture data (Brocca *et al*., 2017). The potential of models to produce reliable outputs is entirely dependent on the quality of the input data and the model itself (Brocca *et al*., 2017). Examples of models used for soil moisture include the Soil Water Assessment Tool (SWAT) and the Optical Trapezoid Model (OPTRAM).

The SWAT model is a physically based, semi-distributed river basin model (Pomeon *et al*., 2018). The main purpose for developing SWAT, was to monitor the environmental impacts that climate change and land use have on hydrology, crop yields and sediments (Chen *et al*., 2011). The model operates at a daily time step and requires a large amount of input data, which includes climate parameters, soil parameters, land use and land cover information and discharge data (Pomeon *et al*., 2018). The river basin is delineated into several smaller sub-basins using a digital elevation model, which are then further divided to produce Hydrological Response Units (HRUs), each consisting of uniform characteristics (Chen *et al*., 2011). HRUs are developed by the overlay of soil and land use maps, which are used to identify characteristics, such as the dominant land use, soil type and slope, that will define each HRU (Uniyal *et al*., 2017; Pomeon *et al*., 2018).

Advantages of the SWAT model include the ability to integrate with Geographic Information Systems (GIS) and the fact that it is an open-source model (Vavara *et al*., 2015). The SWAT model usually uses soil, land use and climate data as inputs to output parameters, such as runoff and sediment yield (Vavara *et al*., 2015). However, a number of studies that use the SWAT model for soil moisture estimation have also been conducted. Uniyal *et al*. (2013), conducted a study that evaluated SWAT simulated soil moisture at a catchment scale. Remotely sensed soil moisture data was used for the verification of the results produced by the SWAT model.

NDVI and Brightness Temperature (BT) data were acquired from the Landsat satellite product for calculating soil moisture, which was then used for the evaluation of soil moisture from SWAT. The main issue highlighted in the study was the scale mismatch between the hydrological model and satellite product, therefore, recommending the need for further research (Uniyal *et al*., 2017).

OPTRAM is a physically based model that uses optical remote sensing data as an input to estimate soil moisture (Huang *et al*., 2019). It was developed with the intention of overcoming the limitations associated with the Thermal-Optical Trapezoid Model (TOTRAM), which used both thermal and optical remotely sensed data for soil moisture estimation (Sagedhi *et al*., 2017). The OPTRAM model provides soil moisture data using a linear relationship between shortwave infrared transformed reflectance (STR) and vegetation cover that is acquired from satellite products (Babaeian *et al*., 2019). The OPTRAM model is similar to the TOTRAM model, however, OPTRAM uses STR instead of LST data and does not require thermal data as an input (Sagedhi *et al*., 2017).

Huang *et al*. (2019), conducted a study involving the estimation of soil moisture using the OPTRAM model in a semi-aid agricultural region. Landsat 8 was used to acquire STR and NDVI data during the growing season, to be able to monitor seasonal changes in soil moisture. Results of the study showed that the use of OPTRAM has potential for estimating soil moisture. Due to the fact that OPTRAM uses STR data, atmosphere parameters have minimal impact on the soil moisture data that is obtained (Sagedhi *et al*., 2017).

2.3.2 Reanalysis products for soil moisture estimation

Reanalysis products have gained a lot of attention in the recent years. They are developed by merging both modelled and observational data, which produces a range of climatic data including soil moisture, rainfall, air temperature, sea-surface temperature and pressure (ECMWF, 2020). Similarly to remote sensing soil moisture products, reanalysis products are also able to address scarce data issues associated with limited *in-situ* observations and can produce long continuous data records for monitoring soil moisture (Cheng *et al*., 2019). Examples of reanalysis products are the Reanalysis 5 (ERA5) and the National Centers for Environmental Prediction / the National Center for Atmospheric Research (NCEP/NCAR) (Lu *et al*., 2005; Cheng *et al*., 2019).

The global ERA5 product was created by the European centre for Medium- Range Weather Forecasts (ECMWF). It is the fifth version of the ECMWF product and has recently replaced the previous ERA-Interim product (Dee *et al*., 2011). The initiation of this product was based on the characteristics of the previous products developed by ECMWF and modifications were made accordingly (ECMWF, 2020). Data is obtained at a 31 km spatial resolution at an hourly time-step, thus producing a long data record extending from 1979 to present (Cheng *et al*., 2019). The product produces volumetric (m^3/m^3) soil moisture data at different depths i.e. 0-7 cm, 7-28 cm, 28-100 cm and 100-289 cm (Peng *et al*., 2015). Due to the uncertainty relating to this product as well as other reanalysis products, data that is provided needs to be verified through observational data (Dee *et al*., 2011).

The NCEP/NCAR reanalysis product produces data on a global scale by combining data retrieved from models and observations (Lejiang *et al*., 2010). The aim for the development of this product was to create a 40-year long data record consisting of several atmospheric and surface variables and was scheduled to finish in 1997 (Kalnay *et al*., 1996). However, this product is currently still being updated and consists of a data record extending from 1948 to present with a 6-hour time-step (Lejiang *et al*., 2010). NCEP/NCAR acquires soil moisture data at two different depths i.e. 0- 10 cm and 10- 200 cm (Lu *et al*., 2005).

Due to the provision of long continuous datasets and the estimation of several climatic variables, reanalysis products are ideal for climate change studies (ECMWF, 2020). However, there are also limitations associated with these products including their coarse spatial resolutions and the lack of research, which makes it difficult to determine their accuracy (Cheng *et al*., 2019).

2.4 Downscaling Approaches for Soil Moisture Products

Satellite-based and reanalysis products show significant potential for estimating and monitoring soil moisture content. However, the coarse resolution issue associated with these products has been a concern for many researchers due to the uncertainty pertaining to the quality of data that is produced (Peng *et al*., 2017; Cheng *et al*., 2019).

Downscaling techniques can be used to disaggregate coarse resolution soil moisture products in order to produce data at higher spatial resolutions and in turn attempt to reduce the scale mismatch that exists between *in-situ* and remote sensing methods (Kim *et al*., 2018; Chen *et* *al*., 2019). Bridging the gap between *in-situ* and remote sensing methods is crucial especially for validation studies (Peng *et al*., 2017). Downscaling methods usually relate remote sensing data to fine scale variables through statistical analysis, regressions or the development of a physically based model (Peng *et al*., 2017; Kim *et al*., 2018). Commonly used fine scale variables for the application of downscaling include Normalized Difference Vegetation Index (NDVI) and Land Surface Temperature (LST) as these variables are significantly indicative of soil moisture conditions (Djamai *et al*., 2016; Peng *et al*., 2017; Fontanet *et al*., 2018). There are several downscaling techniques that are currently available for soil moisture, which differ based on the characteristics of the method and the input data used (Peng *et al*., 2017). Examples of downscaling methods include optical-based and statistical techniques (Sabaghy *et al*., 2018).

Optical-based techniques use optical observations that have a high spatial and temporal resolution (Sabaghy *et al*., 2018). These observations are used to develop a downscaling factor that can be applied to the satellite product being studied with the aim of improving the accuracy of the soil estimates acquired (Peng., *et al* 2017). An example of an optical-based technique is the Disaggregation based on Physical and Theoretical Scale change (Dispatch) algorithm (Fontanet *et al*., 2018). This method relates soil moisture data from satellite products to finer scale observations using NDVI and LST acquired from the Moderate Resolution Imaging Spectroradiometers (MODIS) sensor (Peng *et al*., 2017; Fontanet *et al*., 2018). In addition, other variables are also used, such as the soil evaporative efficiency and soil temperature, which can be derived using NDVI and LST (Fontanet *et al*., 2018). Advantages of optical-based techniques include the high spatial and temporal resolutions associated with optical observations, however, these observations are not available in cloudy conditions and are also affected by vegetation cover (Sabaghy *et al*., 2018).

Statistical downscaling techniques involve the use of a model that uses variables related to soil moisture to produce remote sensing satellite data at high resolutions (Xu *et al*., 2019). Common variables used for statistical downscaling include LST and vegetation data (Xu *et al*., 2019). A study by Loew and Mauser (2008), evaluated the use of a statistical downscaling technique that involved creating a linear regression between the fine scale soil moisture and satellite-based soil moisture (Peng *et al*., 2017). The linear relation between the fine scale and coarse scale soil moisture data is assumed to be stable in time, however, this may not necessarily be the case (Loew and Mauser, 2008). Advantages of statistical downscaling methods include their simplicity in comparison to other downscaling methods and their accountability for bias that exists between the model and satellite soil moisture data (Peng *et al*., 2017).

Recently, machine learning has become a popular method for downscaling satellite soil moisture and involves an automatic approach to analysing datasets (Adab *et al*., 2020). The primary functions of machine learning include the simulation of soil moisture data and the integration of soil moisture from different sources (Sungmin and Orth, 2021). There are several machine learning algorithms that each operate differently and vary in complexity. These algorithms analyse and use data that is given to develop relationships and patterns between datasets (Adab *et al*., 2020; Paul and Singh, 2020). Examples of such algorithms include the Generalized Linear Model (GLM), Support Vector Machine (SVM), K Nearest Neighbor (KNN), Recursive Partitioning and Regression Trees (RPART), Artificial Neural Network (ANN), Boosted Regression Trees (BRT), Multiple Linear Regression (MLR) and Ranger Random Forest (RFF) (Srivastava *et al*., 2013; Guevara and Vargas, 2019; Liu *et al*., 2020; Paul and Singh, 2020; Acharya *et al*., 2021). A study by Sarker (2021), detailed each of these algorithms including the way in which they are built to perform as well as the way in which they are used. A number of studies have been conducted with relation to estimating soil moisture using machine learning (Guevara and Vargas, 2019; Paul and Singh, 2020; Sarker, 2021). Whilst each of these algorithms vary in complexity, their performance is dependent on the given study. Therefore, it is not necessarily the case that a complex algorithm will yield a better outcome than a simple algorithm.

Several studies surrounding the development and use of downscaling methods have been conducted (Loew and Mauser, 2008; Lakshmi, 2012; Fontanet *et al*., 2018; Sebaghy *et al*., 2018). In some cases, downscaling does lead to an improvement in soil moisture data that is acquired from soil moisture products, however, the performance of these methods still remains questionable (Peng *et al*., 2017). Overall, the application of downscaling to reduce the coarse resolutions associated with soil moisture products is a suitable approach, however, the complexity of most methods as well as the intensive processing that is involved may prevent many researchers from being able to undertake such investigations (Gorelick *et al*., 2017).

2.5 Data Processing Platforms

While downscaling approaches have shown potential over the years, certain constraints, which were mentioned above, have limited their application (Gorelick *et al*., 2017; Peng *et al*., 2017).

Furthermore, the acquisition and processing of remote sensing and reanalysis data is resource and computationally intensive. With the use of data processing platforms such as GEE and R, however, these three processes can now be carried out with ease (Tippmann, 2015; Altaweel, 2016).

GEE is a cloud-based infrastructure that was developed in 2010 to provide users with access to large amounts of data (Gorelick *et al*., 2017). This platform has proven to be extremely useful for its ability to store and carry out different types of processing depending on what is required by the user (Altaweel, 2016). Several datasets are accessible for use on GEE, including data from satellite-based and reanalysis products (Gorelick *et al*., 2017). However, datasets and imagery from other data sources can be imported into the platform in cases where any required dataset or product is not yet available (Mutanga and Kumar, 2019).

GEE uses cloud computing for processing and manipulating datasets, thereby allowing a vast number of users to undertake research and analysis using this platform (Gorelick *et al*., 2017; Mutanga and Kumar, 2019). This is due to the fact that most platforms require more advanced software to be able to process large amounts of data, which may limit researchers situated in poor nations due to financial constraints, however, with GEE, fast processing software is not necessary (Mutanga and Kumar, 2019). Additionally, processing of large datasets occurs at a much faster rate on GEE as compared to processing and acquiring data from other sources (e.g. NASA and ESA), with the latter requiring the user to download several images that usually have to be individually processed for further analysis (Gorelick *et al*., 2017). GEE proves to be extremely valuable for the advancement of research due to the aforementioned reasons.

The application of machine learning for downscaling purposes can be facilitated using R, which is a freely accessible and open-source tool (Bischl *et al*., 2016). R includes a series of functions in the form of packages that can be used to carry out statistical analyses (Lantz, 2019). Due to R being open source, these packages are made available by multiple experts, however, they need to be installed prior to use (Lantz, 2019). The R tool and its functions has allowed for machine learning to be undertaken with ease and hence, has become favourable for such studies.

2.6 Review of Case Studies

Several studies have been conducted for investigating the different methods that are available for soil moisture estimation. Some of the key case studies are detailed below, while additional studies were also reviewed in Table 2.3.

2.6.1 Discussion of key studies

Peng *et al*. (2015), carried out a study investigating the performance of four satellite soil moisture products and a reanalysis product. The satellite products used were AMSR-E, ASCAT, SMOS and the recently developed CCI SM, whilst the ERA-Interim reanalysis product was evaluated. These methods were validated using *in-situ* measurements extending from 2008 until 2012. Results that were obtained showed that all the products, including the reanalysis product, were able to map seasonal changes in soil moisture. The CCI SM and ERA-Interim products performed the best amongst all the products that were evaluated having a correlation coefficient (R) value of 0.72 and 0.78 respectively, whilst AMSR-E and SMOS data had larger deviations, possibly due to radio frequency interference (RFI). As a result of the outcomes of this study, CCI SM and ERA-Interim were recommended for water management decisions.

A study conducted by Montzka *et al*. (2017), involved a validation of remotely sensed products and a model using the CRNS. The CRNS was used for the validation study, as it provides area-averaged soil moisture at an intermediate scale. Soil moisture products that were evaluated included AMSR-2, ASCAT, SMOS and SMAP. The Global Land Data Assimilation System version 2 (GLDAS2) model was validated using the CRNS. Results of the study showed that there was a good correlation observed in most cases, however, comparing results was an issue due to scale mismatch that still exists despite the probes intermediate scale. Nevertheless, the CRNS was still recommended as a long-term monitoring network due to its performance. Additional recommendations included the need for biomass correction relating to probe measurements and the need for higher resolution soil moisture records.

Liu *et al*. (2020), investigated the use of machine learning algorithms to downscale coarse resolution soil moisture across four study sites. For this study, the Essential Climate Variable (ECV) product was downscaled using six different algorithms i.e. ANN, BAYE (Bayesian), CART, KNN, RF and SVM. Each of these algorithms were then validated to assess their performance in downscaling soil moisture. From the statistical results that were produced, it was deduced that RF performed best amongst all the algorithms that were analysed having a coefficient of determination (R^2) value of 0.66 (i.e. average of all four study sites). This R^2 value also indicated a better correlation with the *in-situ* data as compared to the original ECV data, which had an R^2 value of 0.65. BAYE and KNN also performed relatively well and have shown potential for downscaling coarse resolution soil moisture data. ANN, CART and SVM did not produce good results owing to several outliers. In addition to the assessment of machine learning algorithms for downscaling, the impact of variables related to soil moisture were also evaluated. The outcome of this component of the study indicated that topography, temperature and vegetation had the most impact on soil moisture content. Further research regarding downscaling through machine learning methods is encouraged.

A study by Fang *et al*. (2018), evaluated a downscaling approach for SMAP soil moisture using NDVI and LST. The downscaling approach used involved the creation of an algorithm that relates all three variables i.e. soil moisture, NDVI and LST. LST estimates derived from MODIS was then used with the downscaling algorithm to produce soil moisture at a 1 km spatial resolution. Both the SMAP satellite soil moisture and the downscaled estimates were then validated using *in-situ* measurements that were obtained. The R^2 results for the downscaled data ranged between 0.19 to 0.70, whilst the R^2 results for the original SMAP satellite data ranged between 0.003 to 0.60. From these results and the analyses that were undertaken, it was evident that the 1 km downscaled soil moisture performed better than the original SMAP data indicating that the application of downscaling resulted in improved and more reliable estimates.

A study conducted by Franz *et al*. (2016), evaluated the use of the CRNS for monitoring soil moisture in a mixed agricultural land use system. The CRNS was calibrated and validated using the Time-Domain Transmissivity (TDT) *in-situ* method. However, an independent monitoring network was used owing to disturbance associated with agricultural practices, which make it difficult to maintain point-based continuous monitoring networks. Results of the study showed that the CRNS data corresponded well with the in-situ TDT data. Due to the fact that the CRNS method can monitor soil moisture while being situated away from routine production practices, it is seen as an ideal method for the continuous monitoring of soil moisture. Therefore, the CRNS can be used for decision-making surrounding the sustainable use and management of water resources.

Several studies were reviewed as shown in Table 2.3, which involved research pertaining to the different aspects of soil moisture estimation that will be addressed in this study. This included assessments of downscaling approaches, satellite-based and reanalysis soil moisture products, CRNS as a validation technique and the use of machine learning for soil moisture estimation.

Table 2.3 Case studies related to soil moisture estimation

Authors	**Study Description**	**Study Site**	**Main outcomes**	**Recommendations**
Peng *et al*., 2015	Evaluating the performance of satellite-based and reanalysis products for soil moisture estimation	China	All products were able to map seasonal changes in soil moisture Based on statistics such as correlation coefficients and root mean square difference, the CCI SM and ERA-interim soil moisture products produced the best results	Further evaluation of the CCI SM and ERA-interim soil moisture products for future water resources management
Montzka *et al*., 2017	Evaluating the use of the CRNS as a validation technique for satellite-based soil moisture products	Australia, USA, Germany, Kenya and India	Through the use of validation scores and triple collocation, it was found that there was a good correlation between the data sets Scale mismatch continues to be a challenge for soil moisture estimation	Biomass correction methods are needed for areas that are vegetation intensive Further research regarding penetration depth differences is required
Vather *et al*., 2018	Assessing the performance of the CRNS *in-situ* method for soil moisture estimation	South Africa	The CRNS data correlated well with the TDR *in-situ* monitoring network The CRNS has the ability to provide reliable and continuous soil moisture data	Calibration methods that are time and labour efficient are needed
Duygu and Akyurek, 2019	Evaluating the use of the CRNS for validating satellite-based products and land surface models	Turkey	The CRNS correlated well with the *in-situ* TDR network The performance of the satellite products varied according to climate and land use The land surface models compared well with the CRNS soil moisture data	Further research is needed to assess the performance of satellite products over several regions that are exposed to different climates and land use Investigate the use of the CRNS for different soil layers

Peng *et al.*, 2017	An evaluation of downscaling methods for satellite-based soil moisture	N/A	From the review of different downscaling approaches, it can be deduced that significant progress was made in terms of soil moisture estimation Remote sensing products show potential, however, are limited by their coarse spatial resolution, which makes the application of downscaling necessary to improve the representativeness of these soil moisture products A highlight of this study was the limitations and uncertainties that are still associated with current downscaling techniques	An efficient validation approach is still required for downscaling methods Uncertainties such as the quality of input data and the downscaling method itself need to be addressed to be able to accurately downscale soil moisture
Fang *et al.*, 2018	Downscaling of SMAP soil moisture using LST and vegetation data	Arizona	The downscaled data was found to be more representative of the spatial and temporal variability of soil moisture as compared to the original SMAP satellite data From the validation results, it was found that the downscaled data had a better relationship with the *in-situ* measurements in comparison to the original SMAP data	In addition to LST and NDVI, the used of other fine scale variables was encouraged The implementation of more *in-situ* monitoring networks for validation purposes is needed
Liu *et al.*, 2020	Application of downscaling to produce high resolution soil moisture using machine learning	America, Tibetan Plateau, Spain, Australia	From the six machine learning algorithms that were investigated for downscaling i.e. ANN, BAYE, Classification and Regression Trees (CART), KNN, RF, and SVM, it was deduced that RF produced the best correlation results	Certain machine learning algorithms require further improvement Further research regarding downscaled soil moisture products is encouraged to continuously improve their performance

			It was also found that topography, vegetation and temperature have a significant impact on soil moisture content	
Adab *et al.*, 2020	Estimating soil moisture using machine learning approaches	Iran	The RF machine learning algorithm produced the best statistical results when compared to the other algorithms that were also being investigated	The use of other variables that are related to soil moisture is encouraged
Franz e*t al.*, 2016	Monitoring soil moisture in a mixed agricultural land use system using the CRNS	Austria	CRNS showed a good comparison with the *in-situ* monitoring network Ideal method for long-term monitoring of soil moisture	More calibration datasets Further research is encouraged
Vather *et al.*, 2020	Monitoring soil moisture and biomass content in a forested site using the CRNS	South Africa	The biomass changes correlated well with the changes the neutron ratio Measuring both the biomass and soil moisture simultaneously can prove to be beneficial in producing more reliable soil moisture datasets Ability to gain a better understanding surrounding soil moisture and vegetation relations	Further research encouraged to monitor biomass for different types of vegetation Identify more hydrogen sources using the CRNS

2.7 Evaluation of Literature

Soil moisture is an important component of the hydrological cycle and has a strong influence on the partitioning of available water between the land and atmosphere. However, it is spatially and temporally variable as it is sensitive to several factors including rainfall, temperature and topography. Therefore, continuous monitoring networks for soil moisture are crucial for climate change studies and water resources management.

From the review of literature, it is evident that there are several methods available for soil moisture estimation. However, each of these methods is associated with certain limitations, which include their spatio-temporal representativeness. While *in-situ* point measurements are relatively accurate, they are unsuitable for hydrological applications that are conducted at catchment scales or larger, due to their poor representativeness regarding the variability in soil moisture. Satellite-based and reanalysis soil moisture products can provide soil moisture estimates over larger scales, however, may be too coarse to adequately represent its spatial heterogeneity. While studies have been conducted to validate satellite-earth observation and reanalysis data using *in-situ* measurements, this has led to further uncertainties due to the differences in the spatial scales between these methods.

The application of downscaling techniques can be used to resolve the coarse spatial resolution issue associated with satellite-based and reanalysis soil moisture products. However, from the review of literature, certain limitations with the application of downscaling were highlighted. The acquisition and processing of remote sensing data is an intricate procedure being resource and computationally intensive. Additionally, downscaling requires a reasonable amount of *in-situ* data for training and testing algorithms, which may pose as a limitation for some researchers. The complexities associated with downscaling has limited many researchers from undertaking research pertaining to this aspect.

The introduction of data processing platforms such as GEE and R, however, have made it possible to overcome previous limitations associated with the application of downscaling. GEE can be used to acquire large quantities of data and perform analyses on remote sensing satellite data through cloud computing, therefore, eliminating the need for advanced software.

Additionally, GEE is time efficient as the acquisition and processing of data can be undertaken within a short time period. The R tool has also made it possible to overcome previous barriers with regards to the application of downscaling. This software is freely available and command driven, thus, allowing for the ease of application of advanced machine learning. Overall, the advancements in geospatial cloud computing coupled with the use of machine learning algorithms facilitated by R have uplifted many of the constraints associated with downscaling.

To assess the performance of the aforementioned methods for soil moisture estimation and ultimately determine whether downscaling leads to an improvement of results, validation is necessary. The CRNS is an ideal method for bridging the measurement gap between *in-situ* methods and remote sensing, as it can provide soil moisture data at an intermediate scale. In addition, it is easily portable and requires little energy to operate. However, this method acquires area-averaged soil moisture and requires calibration. Despite the time and labour requirements associated with the calibration procedure, the CRNS probe shows significant potential in being used as a validation technique for coarse resolution soil moisture products.

Research involving each of the methods mentioned above, have been carried out, however, there still remains a lack of knowledge as to which one performs better when deriving soil moisture data. The constraints associated with the processing of remote sensing data and the application of downscaling have limited many researchers from performing such studies, however, with the advancement of data processing platforms, such studies can now be undertaken with relative ease. As a result, further research is warranted to assess and improve upon currently available methods and gain a better understanding of spatio-temporal dynamics in soil moisture.

3. METHODOLOGY

From the synthesis of literature, the gaps in research relating to the methods used for soil moisture estimation were highlighted. Studies have been undertaken to evaluate satellite-based soil moisture products, reanalysis products and downscaling techniques for the estimation of soil moisture, however, further investigation is required to determine which of these methods would be more suitable for large scale applications. Therefore, this study will evaluate the aforementioned methods for soil moisture estimation, which will be facilitated through the use of data processing platforms. This section entails a description of each of the study sites, the process involved in acquiring data from GEE, the downscaling procedure using model R and the calculations that were done for analysis purposes.

3.1 General Methodology

To achieve the aim of this study the following objectives were proposed:

- Identify and evaluate the use of currently available satellite-based and reanalysis soil moisture products
- Implement and assess a downscaling technique for satellite-based and reanalysis soil moisture products
- Evaluate the data processing platforms that were used for downscaling soil moisture
- Evaluate the performance of satellite-based soil moisture, reanalysis soil moisture and the downscaling technique using the Cosmic Ray Neutron Sensor

The first component of the methodology involved the acquisition and validation of satellite-based soil moisture products. These products included the NASA-USDA SMOS Global Soil Moisture Data (SMOS), NASA-USDA SMAP Global Soil Moisture Data (SMAP_25 km) and NASA-USDA Enhanced SMAP Global Soil Moisture Data (SMAP_10 km), which were accessed on GEE. The same process was carried out for the acquisition of data from the reanalysis products that were used for this study, namely, NCEP Climate Forecast System (CFSV2) and ERA5. The CRNS was then used to validate each product for further analysis.

The downscaling section of the study was then undertaken using the best performing soil moisture product at each of the three study sites based on the validation results. NDVI and LST data from the MODIS satellite product were also acquired on GEE as these variables are related to soil moisture, which was an important requirement for the downscaling procedure. The datasets were then structured and formatted accordingly for application to the R model. The R model was run to produce the downscaled soil moisture data, which was then validated against the CRNS.

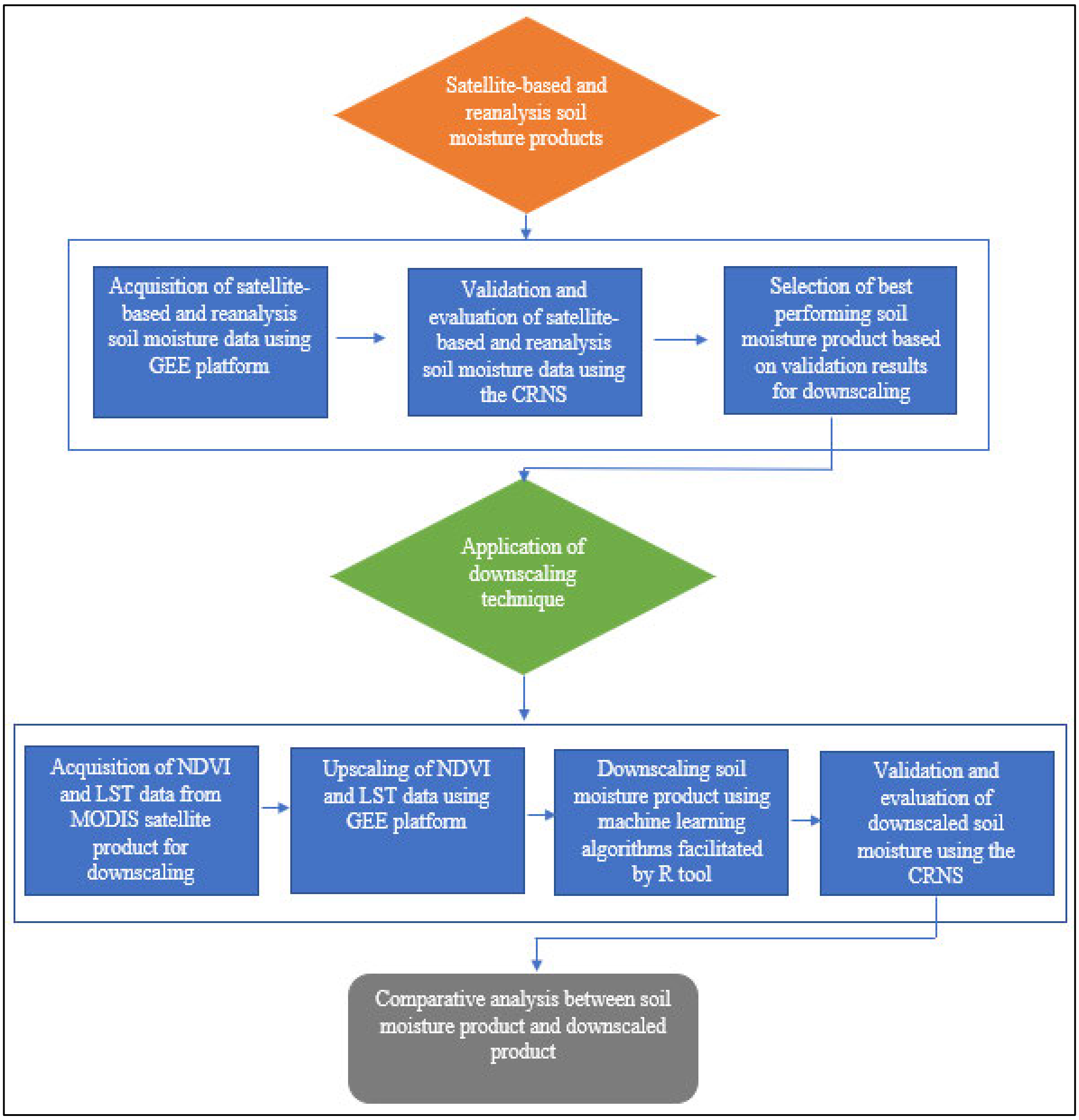

Figure 3.1 General methodology for soil moisture estimation

3.2 Study Site Description and In-situ Data Collection

For this study, research was undertaken at three sites at which extensive work has been done with regards to hydrological observations including soil moisture monitoring. The first site was the Two Streams Catchment located in KwaZulu-Natal, South Africa. The second and third site i.e., Mead CSP3 and York Benny, are both located in Nebraska, United States. Since soil moisture is highly variable, these sites were primarily selected due to data availability and the differences in climatic regimes.

3.2.1 Two Streams Catchment

The Two Streams Catchment site is located -29.21° S and 30.65° E and is characterized by a humid climate with an average annual temperature of 17°C and a mean annual precipitation ranging from 650 mm to 1300 mm (Bulcock and Jewitt, 2012; Scott-Shaw e*t al*., 2020). The rainfall distribution is greater during the summer months, which also experiences heavy mists that increase the moisture content of the soil (Bulcock and Jewitt, 2012). During the winter months, this area is associated with relatively cold and dry weather conditions (South African Government, 2020). The elevation of the area is approximately 1100 m above sea level (Vather *et al*., 2020).

Commercial plantations of black wattle (*Acacia mearnsii*) are the dominant vegetation within the study area (Clulow *et al*., 2011). Common soil forms found at the catchment include apedal and plinthic soils (Clulow *et al*., 2011).

While CRNS data for the Two Streams Catchment can be accessed on the Cosmic Ray Soil Moisture Observing System (COSMOS) site, this data was unreliable and not used since it had not been calibrated. The calibrated CRNS data for the Two Streams Catchment was sourced from Vather *et al*. (2020). The CRNS was installed at the Two Streams Catchment in 2014, having a measurement radius of 210 m and covering an area of 0.14 km^2 (Vather *et al*., 2020). A total of four calibrations were done for the CRNS, three of which were undertaken between 2016 to 2017 and the fourth calibration being carried out in 2018 due to clear-felling (Vather *et al*., 2020). The calibration procedure involved the acquisition and use of area-averaged soil moisture estimates from gravimetric analysis as well as neutron counts (Vather *et al*., 2020).

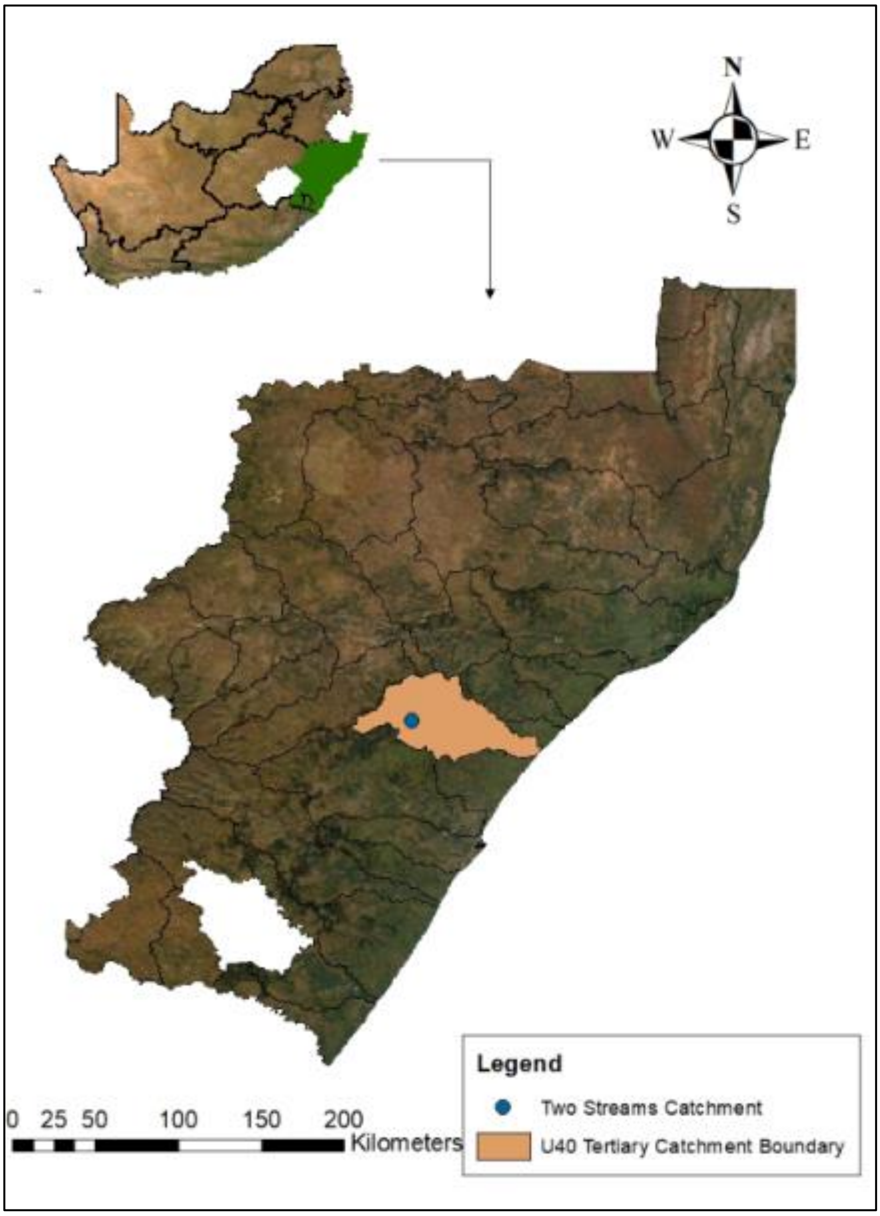

Figure 3.2 Two Streams catchment located in KwaZulu-Natal

3.2.2 Mead CSP3 Catchment

The Mead CSP3 site is located within the eastern side of Nebraska (41.18° N and -96.44° W), situated in the United States of America (Foolad *et al*., 2017; Franz *et al*., 2020). The general climate of the area is of a humid type with warm summers and cold winters being experienced (Yang *et al*., 2018). The Mead site has an average annual precipitation of 784 mm and an average annual temperature of 10.5°C (Yang *et al*.,2018; Foolad *et al*., 2020). The elevation is approximately 366 m above sea level and is relatively flat (US Climate Data, 2021).

The dominant vegetation in the area are corn and soybean, which are grown rotationally (Franz *et al*., 2020). The site is rainfed and the growing period for the vegetation generally starts in May and ends towards October (Foolad *et al*., 2017; Franz *et al*., 2020). The soil types present are mainly silt loam and silty clay loam (Foolad *et al*., 2017). The soil moisture content determines the measurement depth of the CRNS, which typically ranges between 15 – 40 cm (Foolad *et al*., 2017). The neutron counts are taken at an hourly time step and is then used to derive the soil moisture content (Foolad *et al*., 2017).

3.2.3 York Benny Catchment

The York Benny study site, which is also situated in Nebraska, is located 40.93° N and -97.46° W (COSMOS, 2021). The general climate experienced is of a hot and humid type with warm summers and cold and dry winters (Franz *et al*., 2020). The average annual temperature is approximately 19 °C and the mean annual precipitation is 725 mm (COSMOS, 2021) The elevation of the area is 494 m above sea level (COSMOS, 2021).

The dominant vegetation type at the York Benny site is corn and soybean, which are grown rotationally with the latter being grown during the even years (Franz *et al*., 2020). This site is irrigated using a center-pivot system (COSMOS, 2021). The most common soil type found in the area is silty loam (COSMOS, 2021). The soil moisture content is variable due to the climatic characteristics of the study site. The rainfall distribution is greater towards the summer months as compared to the cold and dry winter period (Franz *et al*., 2020). The most common form of precipitation during the summer months is thunderstorms (University of Nebraska, 2016). Snowfall is possible in some cases, however, this is not common due to the low amount of rainfall and recurring thaws (University of Nebraska, 2016).

With regards to the *in-situ* data collection for both the Mead CSP3 and York Benny study sites, the calibrated CRNS datasets were sourced from Franz *et al*. 2020. The measurement radius of the CRNS situated at both these sites have a range of 150 – 250 m, which is dependent on climatic and surface conditions, and cover an area of 0.28 km^2 (Franz *et al*., 2015; Franz *et al*., 2020). Calibration of the CRNS datasets was done using a portable moisture probe (Franz *et al*., 2020). This procedure was carried out over a sufficient amount of land within each study site to be able to produce representative and reliable calibration results (Franz *et al*., 2020).

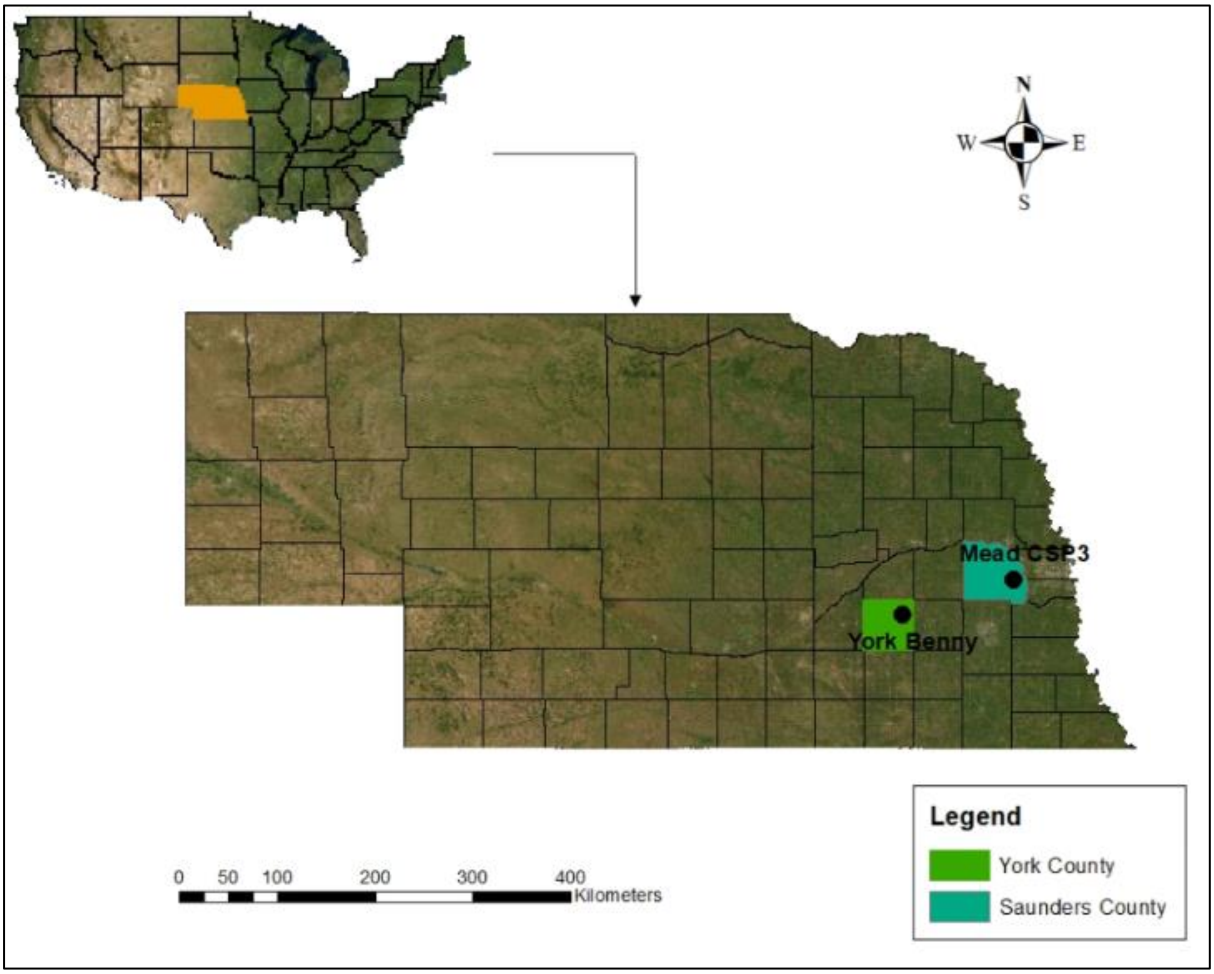

Figure 3.3 Location of the Mead CSP3 and York Benny study sites in Nebraska

3.3 Acquisition of Satellite-based and Reanalysis Soil Moisture Data

Satellite-based and reanalysis soil moisture data are available on multiple platforms including NASA, ESA, BEC, JAXA etc. However, these data sources require an extended amount of time and computational resources. GEE is a recently developed platform that overcomes these shortcomings and was solely used for acquiring the satellite-based and reanalysis soil moisture data in this study as it was the most efficient method.

3.3.1 Data collection and processing in GEE

For this section, three satellite-based soil moisture products were used, namely, SMOS, SMAP_25 km and SMAP_10 km. These products were selected as they are currently available on GEE for use and are also amongst the commonly used satellite products for soil moisture studies (Peng *et al*., 2015, Montzka *et al*., 2017, Fang *et al*., 2018). The SMOS product available on GEE is one of the processed versions derived from the original SMOS product resulting in a spatial resolution of 25 km and 3-day revisit period. The bands relating to different measurements included surface soil moisture, subsurface soil moisture and anomaly data, however, for the purpose of this study, surface soil moisture was solely used. The length of record available on GEE for SMOS extends from 2010 to 2020. The SMAP_25 km product was also a processed version and had the same characteristics as SMOS including its spatial resolution, revisit period and band data. The data record for SMAP_25 km on GEE is for a period extending from 2015 to 2020. SMAP_10 km is a modified version of the SMAP_25 km through data assimilation. This product has a finer resolution of 10 km, however, other characteristics such as the revisit period and band measurements are the same as the SMAP_25 km product. The record length for the SMAP_10 km product is from 2015 to present.

The first step for acquiring all three datasets on GEE involved defining the region of interest (ROI), therefore, buffer zones were created for each study site (i.e. Two Streams Catchment, Mead CSP3 and York Benny). This involved the use of Google Earth Pro to locate and construct the buffer zones for each area using their respective coordinates, which were ultimately produced as keyhole markup language (KML) files. However, these KML files were not compatible with GEE and had to be converted to a shapefile format. This conversion process was done using ArcGIS. Using ArcMap version 10.6, the KML files for each of the study sites were changed to a layer format using the 'KML to layer' conversion tool. Once in layer format, these files were then exported as shapefiles, which were compatible with GEE. The three shapefiles were then uploaded on GEE and imported into the code script. The area of the buffer zones at the Two Streams Catchment, Mead CSP3 and York Benny were 0.65 km^2, 0.63 km^2 and 0.62 km^2 respectively.

After defining the region of interest, the soil moisture datasets were individually acquired using the necessary filter methods available on GEE. The image collection for the soil moisture product was imported and the ROI and time period were then specified. The time period used was dependent on the availability of data on GEE as the entire image collection for each product was acquired. For example, the SMAP_25 km dataset on GEE extended from 2015 to 2020, hence, the complete five-year long record was used.

The surface soil moisture band was then selected together with its maximum and minimum values. A function to derive the mean value for each polygon (buffer zone) was also included in the code script to account for every pixel contained within the study area. However, it was generally the case that the three study sites fell within a single pixel due to the coarse resolutions of the satellite-based soil moisture products.

Additional specifications were made to the code such as assigning a time factor to every feature, sorting of the data for the desired outcome and the creation of a time series plot. The code was then run to produce the soil moisture dataset for analysis. However, when acquiring the SMOS and SMAP_25 km data, the model did not run for the three study sites. This was due to the fact that the buffer zones created for the three study sites were too small to be represented by the coarse resolution products, therefore, they were slightly enlarged to be able to acquire the soil moisture datasets. This did not have an impact on the data that was produced as the enlarged buffer zones still fell within the same pixels as the original buffer zones. The SMAP_10 km data did not require the use of the enlarged buffer zones as the product was able to produce data for the original sizes of the study sites due to its lower spatial resolution of 10 km. Therefore, it is important to ensure that the chosen ROIs are sufficiently large to acquire data from the images.

The same code was used for the acquisition of data from each satellite-based soil moisture product. The only modifications made each time the model was run included changing the image collection, ROI and date range where necessary.

ERA5 and CFSV2 products were also accessed, processed and evaluated using GEE. ERA5 is a global reanalysis product combining observations and modelled data to produce data for a total of 50 variables. Some of these variables include volumetric soil moisture, temperature, precipitation, runoff and evaporation. Volumetric soil moisture data was available at different depths, however, the variable pertaining to the surface soil moisture (0 – 7 cm depth) was only used, as the penetration depth of the SMOS and SMAP products also fell within this range (Cui *et al*., 2017). Data availability for the ERA5 product extends from 1981 to present.

The CFSV2 reanalysis product is a version 2 of the CFS product, which commenced in March 2011. It is a 6 hourly product that has a long record of data extending from 1979 to present. Similarly, to the ERA5 reanalysis product, CFSV2 also produces data for several variables including volumetric soil moisture content, temperature, surface pressure, latent heat flux etc. The variable measuring the surface volumetric soil moisture content (0 – 5 cm depth) was used.

The procedure involved in acquiring both these datasets from GEE was the same as the satellite-based soil moisture products. The only changes made to the code script was the image collection to specify which dataset was required, the ROI as the model was run separately for each study site and the time of interest had to also be specified as the record lengths available varied among the different products.

3.4 Analysis of Satellite-based and Reanalysis Soil Moisture Data

Upon the acquisition and processing of the soil moisture products, a validation was undertaken using CRNS measurements. This procedure was done for each of the products across all three study sites to evaluate their performance. It involved graphical and statistical analyses after which, the best performing product at each of the three sites was selected for the downscaling component of this study. The graphical analyses involved the creation of time series plots to visualize the trends observed by each of the products against the CRNS data. However, some of the datasets acquired from GEE had to be adjusted prior to plotting the time series. SMOS, SMAP_25 km and SMAP_10 km were the only datasets acquired in mm units, which were then converted to a volumetric unit to ensure consistency amongst all datasets. This was computed using the following equation (Voroney, 2019):

$$Volumetric\ soil\ water\ content = \frac{Depth\ of\ water\ (mm)}{Depth\ of\ soil\ \ (mm)} \tag{3.1}$$

In addition to the graphical analyses, several performance metrics were used to further evaluate the performance of each soil moisture product against the CRNS data. The correlation coefficient (r) and coefficient of determination (R^2) statistics were derived using scatter plot graphs. The r value ranges between 0 to 1 and is used to represent the linear relationship between the product and the CRNS data. The closer the value is to 1, the better the correlation between both variables. The R^2 value gives an indication of the strength of the linear relationship between the product and the CRNS. The closer the R^2 value is to 1, the better the relationship is between the datasets. In addition to these statistics, the Root Mean Square Error (RMSE), unbiased Root Mean Square Error (ubRMSE), Mean Absolute Error (MAE) and relative volume were also calculated using the following equations (Liu *et al*., 2018; Gruber *et al*., 2020; Acharya *et al*., 2021):

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \hat{x}_i)^2}{N}} \tag{3.2}$$

$$\text{ubRMSE} = \sqrt{\frac{\sum_{i=1}^{N} [(x_i - y_i) - (\hat{x}_i - \hat{y}_i)]^2}{N}} \tag{3.3}$$

$$\text{MAE} = \frac{\sum_{i=1}^{N} |x_i - \hat{x}_i|}{N} \tag{3.4}$$

$$\text{Relative Volume} = \frac{x_i - \hat{x}_i}{x_i} \times 100 \tag{3.5}$$

where:

N = The total number of data values

x_i = Observed data values

$\hat{x}_i$ = Predicted data values

y_i = Averaged observed data value

$\hat{y}_i$ = Averaged predicted data value

i = Variable i

The RMSE and MAE values are indicative of the differences observed between the soil moisture product and CRNS data. Both these values (RMSE and MAE) range between 0 to ∞. ubRMSE is used to address the bias and error in the comparisons, which are caused by the depth difference between the remote sensing and CRNS data. This metric has a target value of 0.04 $m^3\ m^{-3}$, therefore, a value lower than 0.04 $m^3\ m^{-3}$ implies that the product meets accuracy requirements with statistical significance. The relative volume is a percentage derived to examine the relationship between the observed (CRNS) and predicted (soil moisture product) datasets. This value could be negative or positive, indicating an oversimulation and undersimulation respectively.

3.5 Downscaling of Soil Moisture Data using Machine Learning

The downscaling of the soil moisture datasets was performed using the R statistical software. The method adopted herein is based on the method applied in a study that was conducted by Gokool *et al*. (2022). The best performing product at each of the three study sites, based on the validation results, was selected for this application.

3.5.1 Acquisition of NDVI and LST data

In general, the application of several downscaling approaches involves estimating soil moisture from finer resolution variables that are related to soil moisture (Peng *et al*., 2017; Dandridge *et al*., 2019). This is usually undertaken by developing relationships in the form of regressions, statistics etc. between the variables (Peng *et al*., 2017; Kim *et al*., 2018). For the downscaling of the selected soil moisture product, NDVI and LST data from the MODIS satellite, was acquired as both these fine scale variables have a significant influence on soil moisture (Peng *et al*., 2017; Fang *et al*., 2018; Fontanet *et al*., 2018). These datasets were also acquired using GEE.

The MODIS Terra Surface Reflectance Daily Global 1 km and 500 m satellite product was used for the acquisition of NDVI data having a record length of 21 years (2000 – present). This product provides seven bands each representing different wavelengths of surface reflectance. The NDVI data was acquired at a 500 m spatial resolution and were unitless estimates ranging between -1 to 1. For the acquisition of LST, the MODIS Terra Land Surface Temperature and Emissivity product was used, which has also been providing data for a 21-year period (2000 – present). The LST estimates were produced in Kelvin units on a daily time step at a 1 km spatial resolution.

For the acquisition of NDVI data, the same buffer zones created for each of the study sites for the validation component, were used. The image collection for the MODIS Terra Surface Reflectance product was imported to the code script. The ROI was then specified with the code script being run individually for each study site. The entire dataset available on GEE was acquired, therefore, the time period defined was 2000 – 2021. The bands used to derive NDVI estimates were then selected, which were surface reflectance bands 1 and 2 representing the red and Near Infrared (NIR) wavelengths respectively. After selecting the bands, a function was used to calculate the NDVI estimates. This function derives NDVI using the following equation:

$$NDVI = \frac{NIR-Red}{NIR+Red} \tag{3.6}$$

where:

NDVI	= Normalized Difference Vegetation Index
NIR	= Near Infrared band
Red	= Red band

The function used to calculate a mean value for each polygon was applied to the data. This step was necessary as more than one pixel fell within each of the study sites.

The code script created for obtaining NDVI data was also used for the LST data, however, some adjustments had to be made to the script. The image collection for the Terra Land Surface Temperature and Emissivity product was imported. The ROI and time of interest was then specified. The entire LST dataset was acquired, therefore, the time period used was from 2000-2021. The LST band was then selected, which represented the daytime land surface temperature. There were no functions needed to calculate LST, as a band for the variable was already available. However, the dataset had to be multiplied by the LST scale factor of 0.02 to produce the required data.

3.5.2 Upscaling NDVI and LST data

The downscaling component was undertaken for the selected satellite-based product to estimate finer scale soil moisture from NDVI and LST data. The NDVI and LST datasets obtained from GEE had to be processed before their input in R. Firstly, the data had to be upscaled to derive the relationships between all the datasets at the coarsest spatial resolution. Thereafter, the finer resolution estimates of NDVI and LST were used to estimate soil moisture at a finer resolution. This is based on the assumption that the relationship derived at the coarsest spatial resolution remains consistent at a finer spatial resolution as well (Blöschl *et al*., 2009; Alemohammad *et al*., 2018; Dandridge *et al*., 2019). However, since the original LST and NDVI data were at different spatial resolutions, which was 0.5 km and 1 km respectively, the NDVI data was upscaled to be consistent with the LST data. This step was undertaken prior to the upscaling procedure between NDVI, LST and the selected soil moisture products.

3.5.3 Downscaling of the soil moisture product

The packages required to perform the necessary analyses to downscale soil moisture were initially loaded into R. One of the main packages that was used included the Classification And REgression Training (caret) package, which uses multiple packages on R to carry out the training procedure of the models (Kuhn, 2008). These packages do not have to initially be loaded on R as they will be loaded automatically when required during the application. Due to the different temporal resolutions associated with the products, the datasets were converted to a monthly time series. The length of record varied depending on the availability of data pertaining to the selected soil moisture product, NDVI and LST for each of the three study sites. It was necessary to select a period where data was available for all three variables. The monthly estimates for the three variables were then all added to one excel spreadsheet, which was imported into R. Once imported to R, the excel spreadsheet was analysed to identify any missing data and if any missing data was detected, it was infilled using the average value for that specific variable.

The next step involved splitting the selected soil moisture dataset into two separate datasets, one for training the model and the other for the validation process. The reason for splitting the data is to determine the best performing machine learning algorithm as there is a possibility that an algorithm performing well with the training data, may not perform the same when exposed to unseen data during the validation process. Prior to splitting the data, the NDVI and LST data (predictor variables) were separated from the soil moisture data (target variable). A 'createDataPartition' function was then used to split each of these datasets into training and validation data, which comprised of 80 % and 20 % of the data, respectively.

The process for building and evaluating the machine learning algorithms was then undertaken. To evaluate the performance of the machine learning model, a repeated cross validation method was used, which involved repeating the cross validation three times to produce a mean value across a total of 10 folds. A total of five machine learning algorithms were chosen for evaluation:

i. General Linear Model (glm)
ii. Recursive Partitioning and Regression Trees (rpart)
iii. K-Nearest Neighbours (knn)
iv. Suport Vector Machines (svmRadial)
v. Random Forest (ranger)

A mixture of simple and complex algorithms was used. After creating the model list, the training data was used to evaluate the performance of each algorithm. The RMSE results were produced to assess the performance of the algorithms. In addition, the correlation between the individual models were tested to assist in identifying which models to include in the ensemble. Based on the results that were obtained across each study, algorithms that had a correlation exceeding 0.75 with the other algorithms, were not used for further analysis as they would most likely produce similar predictions.

In addition to the four algorithms selected for analysis, two model ensembles were also created using the 'caretStack' function with the aim of identifying the best model to use. The first model ensemble was created by using the simple glm, whilst the second ensemble was created using the more complex ranger algorithm. The RMSE statistics were also produced for both ensembles, which were then compared to the results of the original algorithms used.

The validation dataset was then used to evaluate the performance of the models. However, based on the results in the previous procedure, only the three best performing models were used. The reason for selecting three models instead of solely using the best performing model is due to the possibility that a model that performs well for the training dataset may not necessarily perform the same when exposed to the validation data. The RMSE, MAE and correlation statistics were derived for each of the three models. Based on the results obtained, the best performing model was selected. The CRNS validation data and original NDVI and LST data (1 km spatial resolution) were then imported to the R model to perform the final step of the downscaling component. The best performing model used the NDVI and LST data to produce the downscaled soil moisture estimates, which were then compared to and evaluated using the CRNS data. The predicted (downscaled soil moisture) and observed data (CRNS) were then exported from the R model as an excel file for further analysis.

This entire downscaling procedure was undertaken individually for each of the three study sites. The length of record produced for the downscaling soil moisture estimates was dependent on the CRNS data availability. For example, the Two Streams Catchment had a CRNS data record of 3 years, therefore, the downscaled soil moisture estimates produced for this catchment was for the same 3-year period.

3.6 Data Availability

The number of years of soil moisture data that was available for each product can be seen in Figure 3.4. The reanalysis and satellite-based soil moisture products each had the same length of record available for all three study sites. The ERA5 and CFSV2 reanalysis products had the longest records of data exceeding 40 years. The satellite-based soil moisture products had record lengths ranging between 5 to 11 years. The availability of CRNS validation data varied amongst the study sites. The Mead CSP3 site had the longest CRNS data record of 11 years, the York Benny site had a record length of 7 years, and the Two Streams Catchment had the shortest record of 3 years.

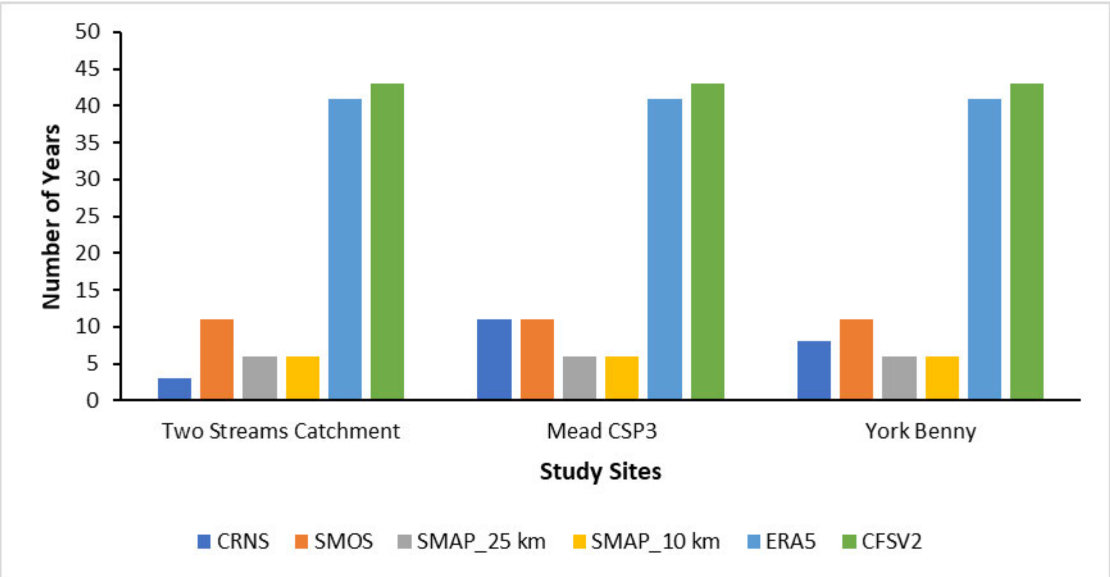

Figure 3.4 Bar graph illustrating the number of years of data available for each soil moisture product

4. RESULTS

This section entails the results acquired pertaining to the different aspects undertaken in the methodology with the aim of accomplishing the objectives of the study. The chapter is structured as follows:

i. Validation of the three satellite-based soil moisture products (SMOS, SMAP_25 km and SMAP_10 km) using the CRNS *in-situ* data
ii. Validation of the two reanalysis soil moisture products (ERA5 and CFSV2) using the CRNS *in-situ* data
iii. An assessment of the use of satellite-based and reanalysis products to estimate soil moisture
iv. An assessment of the performance of the machine learning algorithms that were used to downscale the CFSV2 reanalysis soil moisture product
v. Validation of the downscaled soil moisture estimates that were produced using the R model

The aforementioned aspects were carried out for the soil moisture estimates produced across all three study sites (Two Streams Catchment, Mead CSP3 and York Benny). To produce the necessary results, various analyses and statistics were used including:

i. Time series analysis
ii. Scatter plot graphs, which were used to derive the coefficient of determination (R^2)
iii. Correlation coefficient (r)
iv. RMSE
v. ubRMSE
vi. MAE
vii. Relative volume

The analyses and statistics used were considered the most suitable for the different aspects of research in this study (Liu *et al*., 2018; Acharya *et al*., 2021). However, these analyses were not used for every section and were applied only where it was necessary.

4.1 Validation of the Satellite-based and Reanalysis Soil Moisture Products

The validation of the soil moisture estimates was undertaken using the CRNS data across the three study sites. The satellite-based soil moisture products that were investigated were SMOS, SMAP_25 km and SMAP_10 km. The selected reanalysis products were ERA5 and CFSV2. For the time series analyses, the period used extended from 2015 – 2018, as data was available for all the soil moisture products as well as the CRNS. The datasets were also converted to average monthly soil moisture estimates due to the different temporal resolutions of each product.

4.1.1 Two Streams study site validation

The time series analysis for the CRNS and satellite derived soil moisture is illustrated in Figure 4.1. The three satellite products and the CRNS exhibit a variation in soil moisture during the winter and summer months. During the summer months, the satellite-based and CRNS soil moisture estimates are relatively higher when compared to the rest of the year. This is likely due to the fact that the Two Streams Catchment has a greater distribution of rainfall during summer as compared to winter.

From Figure 4.1, the three satellite-based soil moisture products follow a similar trend to the CRNS, however, majority of the time, these estimates are either overestimated or underestimated. SMOS and SMAP_25 km mainly underestimates soil moisture during the winter months, whilst the products during summer months exhibit much higher peaks when compared to the CRNS data. The SMAP_10 km product generally underestimates soil moisture during both summer and winter months when compared to the CRNS. In terms of the variability in soil moisture, while SMOS and SMAP_25 km follows the same trend, the SMAP_25 km product appears to have the better relationship with the CRNS. Although the product underestimates soil moisture in most cases, the trend followed by both SMAP_25 km and the CRNS are relatively similar.

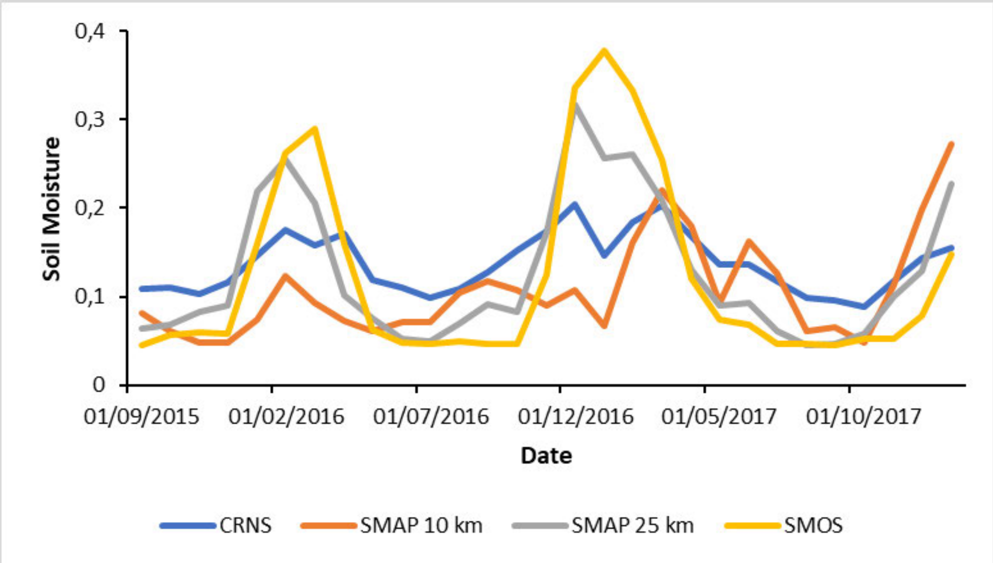

Figure 4.1 **Time series analysis of the CRNS and satellite-based soil moisture at the Two Streams study site**

The ERA5 and CFSV2 reanalysis soil moisture estimates for the Two Streams site were also evaluated against the CRNS data using a time series analysis. From Figure 4.2, it can be seen that the data produced by the reanalysis products and the CRNS indicate seasonal variation of soil moisture. The soil moisture content increases towards the summer months and decreases when approaching winter. This is likely due to the greater rainfall distribution occurring in summer and dry conditions associated with winter. The ERA5 and CFSV2 reanalysis products, in Figure 4.2, overestimate soil moisture during the entire time-period that was selected. Despite the overestimation of soil moisture, the trend observed by the CFSV2 product is relatively similar to the CRNS trend

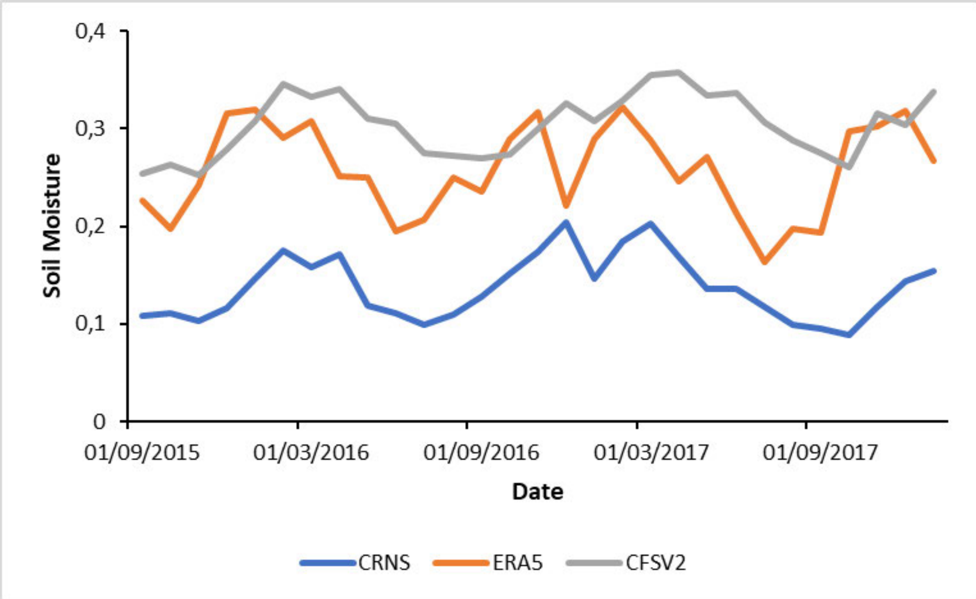

Figure 4.2 **Time series analysis of CRNS and reanalysis soil moisture at the Two Streams study site**

Several performance metrics were derived to further evaluate the performance of each soil moisture product against the CRNS data (Table 4.1). The statistical results acquired in Table 4.1 collectively indicate SMAP_25 km and SMOS performed best amongst the soil moisture products. The CFSV2 reanalysis product also performed relatively well, however, the RMSE and MAE values were the highest observed when compared to the other products. The relative volume percentage for CFSV2 was the largest, -121.96 %, indicating there is an oversimulation and the predicted value (CFVS2) is more than double the actual value (CRNS). However, despite the large percentage value, the difference between the actual and predicted value is relatively small. The ERA5 product also has a large relative volume error percentage indicating an oversimulation. The relative volume error of the CFSV2 and ERA5 products correspond with their trends observed in Figure 4.2, where it can be seen that both products oversimulate soil moisture when compared to the CRNS *in-situ* data. From the time series analysis and the statistical results, it is evident that SMAP_25 km performed best.

Table 4.1 Statistics comparing each of the soil moisture products against the CRNS at the Two Streams site

	SMOS	SMAP 25 km	SMAP 10 km	ERA5	CFSV2
r	0.76	0.83	0.54	0.43	0.75
R^2	0.58	0.68	0.29	0.18	0.56
RMSE	0.08	0.06	0.06	0.13	0.17
MAE	0.07	0.05	0.05	0.12	0.17
Relative Volume (%)	-10.75	-6.85	-22.08	-88.48	-121.96

4.1.2 Mead CSP3 study site validation

The satellite-based soil moisture products at the Mead CSP3 study site were plotted against the CRNS data using a time series analysis. At the Mead CSP3 site, summer months generally experience more rainfall than the winter months, therefore, it is expected that soil moisture content would be greater during summer and lower in winter. However, from Figure 4.3, it can be seen that soil moisture is highly variable at this catchment with highs and lows experienced throughout the year.

The SMAP_25 km product performed relatively well when compared to the CRNS. However, the soil moisture data derived from this product is overestimated throughout the period. The same trend was observed for the SMOS product. The SMAP_10 km product performed differently with the soil moisture being both overestimated and underestimated throughout the year as shown in Figure 4.3. However, the highs and lows observed by the SMAP_10 km product follow a similar pattern to the CRNS trend.

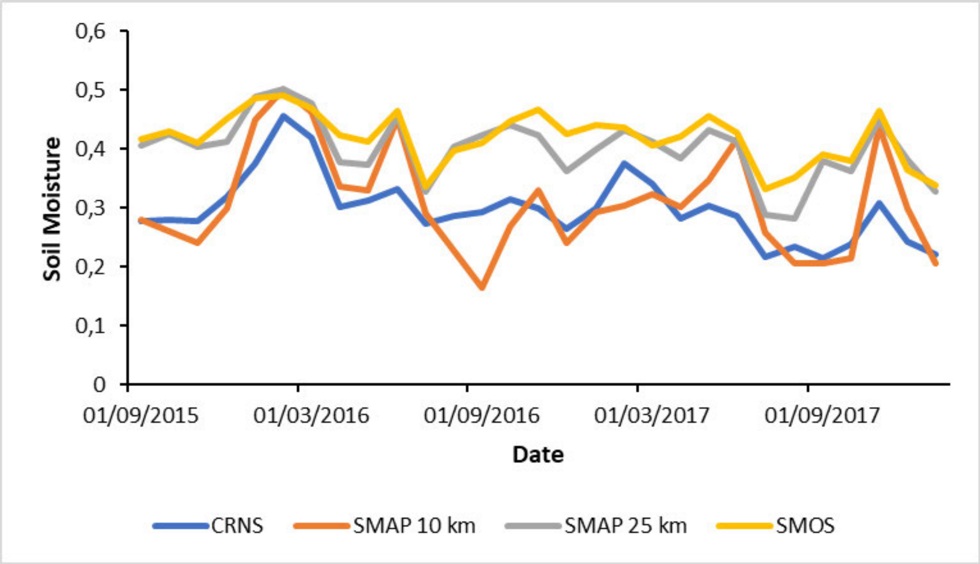

Figure 4.3 Time series analysis of the CRNS and satellite-based soil moisture at the Mead CSP3 study site

The time series analysis for the reanalysis products and CRNS data can be seen in Figure 4.4. All three datasets show variation in soil moisture throughout the time-period allocated. However, there is no distinctive trend in terms of seasonal variation as the soil moisture estimates vary during both summer and winter months. The CFSV2 compares well with the CRNS data, which can be seen by the trend followed in Figure 4.4. The trend observed for the ERA5 product does not correspond as well as the CFSV2 when being compared to the CRNS data. However, despite the CFSV2 product having a similar trend, both the reanalysis products overestimate and underestimate soil moisture during the time-period used.

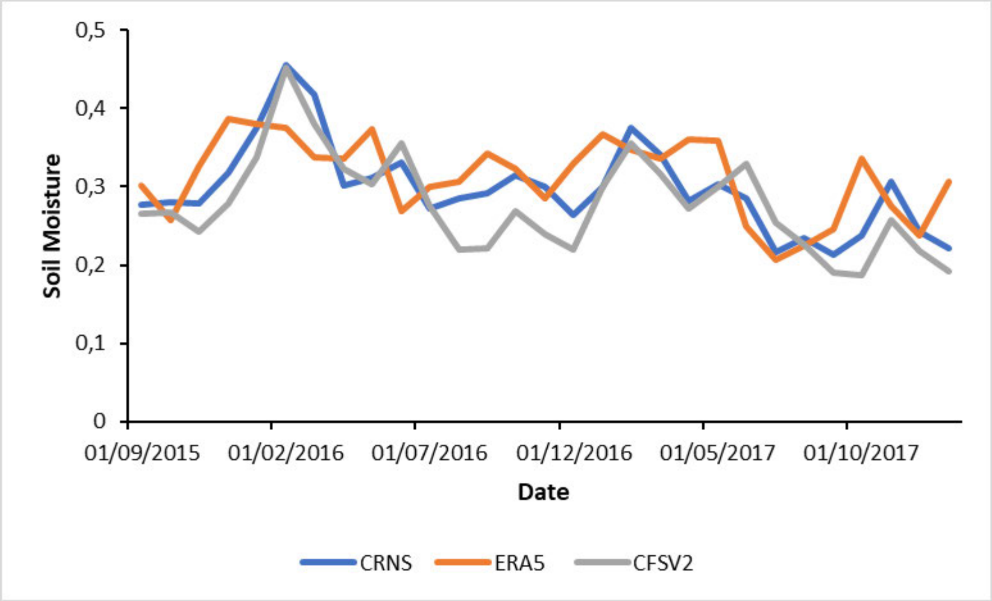

Figure 4.4 **Time series analysis of the CRNS and reanalysis soil moisture at the Mead CSP3 study site**

From the trend series analyses (Figure 4.3 and Figure 4.4), the SMAP_25 km and CFSV2 performed relatively well against the CRNS. From the statistical results acquired in Table 4.2, it is evident that the CFSV2 reanalysis product produced the best results, which corresponds with its trend observed in Figure 4.4. The SMAP_25 km product showed satisfactory performance, however, the RMSE and MAE values were amongst the highest when compared to the other soil moisture products. Both the SMAP_25 km and SMOS product had the highest relative volume error percentages, with the negative values indicating an oversimulation against the CRNS. This corresponds with the trend in Figure 4.3, where it is evident that both products overestimate soil moisture for the entire period that was used.

Table 4.2 **Statistics comparing each of the soil moisture products against the CRNS at the Mead CSP3 site**

	SMOS	**SMAP 25 km**	**SMAP 10 km**	**ERA5**	**CFSV2**
r	0.37	0.66	0.56	0.47	0.73
R^2	0.14	0.44	0.32	0.22	0.53
RMSE	0.11	0.11	0.09	0.06	0.06
MAE	0.10	0.10	0.07	0.05	0.05
Relative Volume (%)	-28.14	-30.10	-12.02	-1.25	-9.44

4.1.3 York Benny study site validation

The satellite-based soil moisture was plotted as a time series (Figure 4.5) to evaluate their performance using the CRNS data. The soil moisture data from each of the satellite products as well as the CRNS show variation throughout the data period used, with no distinctive pattern in terms of seasonal changes. Highs and lows experienced by the soil moisture data occur during both the summer and winter months. The rainfall distribution at the York Benny site is generally higher during the summer months as compared to winter. Therefore, it was expected that soil moisture content would have followed the trend in rainfall, however, this was not the case.

From Figure 4.5, it can be seen that the SMOS and SMAP_25 km products overestimate soil moisture when compared against the CRNS data. The trends observed for both products, however, are similar to the trend of the CRNS data. The SMAP_10 km product does not perform as well as the other two satellite products with soil moisture being overestimated in some instances and underestimated in other cases. The highs and lows of soil moisture exhibited by all three products correspond relatively well with CRNS data. However, it is evident that soil moisture at this site is highly variable throughout the year despite rainfall being distributed more towards the summer months.

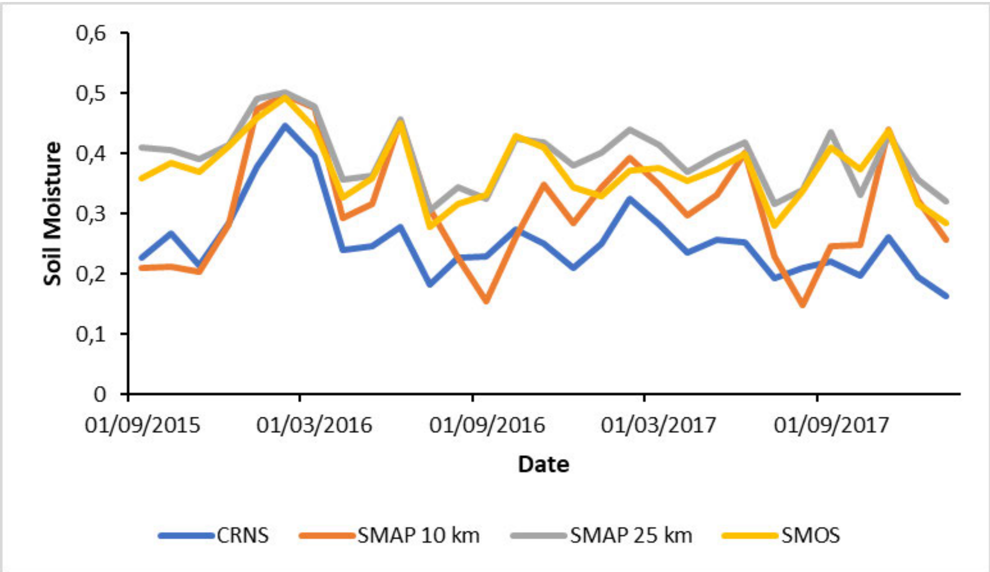

Figure 4.5 Time series analysis of the CRNS and satellite-based soil moisture at the York Benny study site

The time series analysis for the reanalysis products and CRNS at the York Benny study site can be seen in Figure 4.6. The patterns observed at this catchment are similar to the Mead CSP3 site, with variations in soil moisture throughout the time period. The highs and lows of soil moisture were experienced during both the summer and winter periods even though more rainfall is generally experienced during the summer months, which would have ideally resulted in the soil moisture content being higher.

Amongst both the reanalysis soil moisture products, the CFSV2 product corresponds better with the CRNS data. However, this mainly refers to the trend observed by the product, as it overestimates and underestimates soil moisture in some cases. The ERA5 product has a different trend to that of the CRNS *in-situ* data, which can be seen in Figure 4.6. The highs and lows observed by the CRNS data is contradicted by the ERA5 trend in certain periods. Additionally, this product overestimates soil moisture in most instances during this time-period.

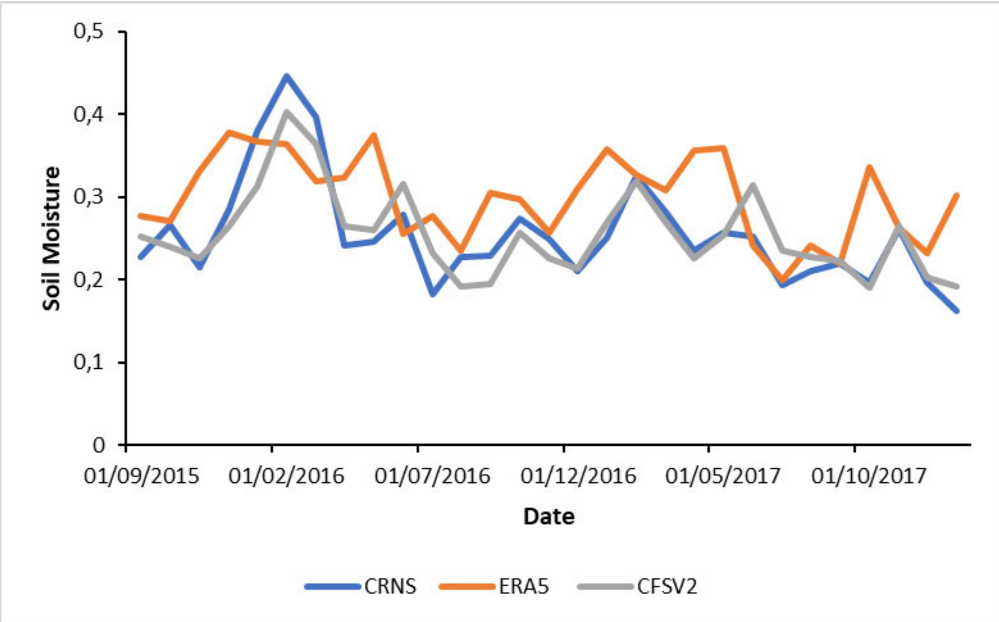

Figure 4.6 Time series analysis of the CRNS and reanalysis soil moisture at the York Benny study site

The statistical results acquired for each of the soil moisture products can be seen in Table 4.3. The SMAP_25 km product produced the highest r and R^2 values amongst all the soil moisture products. However, the RMSE and MAE values obtained were the highest when compared to the other datasets. The relative volume error was also high with the negative value indicating an overestimation of soil moisture. The SMOS product produced similar statistical results as shown in Table 4.3. The CFSV2 produced the best statistical results in this instance having an r value of 0.66 and R^2 of 0.44. Additionally, the RMSE, MAE and relative volume values were relatively low, which indicated that the actual and predicted datasets did not exhibit large differences.

Table 4.3 Statistics comparing each of the soil moisture products against the CRNS at the York Benny site

	SMOS	SMAP 25 km	SMAP 10 km	ERA5	CFSV2
R	0.62	0.70	0.59	0.37	0.66
R^2	0.38	0.49	0.35	0.13	0.44
RMSE	0.12	0.15	0.10	0.08	0.05
MAE	0.12	0.14	0.09	0.06	0.04
Relative Volume (%)	-41.76	-48.71	-26.09	-16.56	-4.92

For the downscaling component of the study, the best performing product for each study site had to be selected for further analysis. The SMAP_10 km and ERA5 soil moisture products were not selected for the downscaling study as both these datasets performed poorly across all three study sites. For the Two Streams site, SMOS and SMAP_25 km performed well against the CRNS. However, based on the validation results, SMAP produced better results and was chosen for the downscaling application. With regards to the Mead CSP3 site, the CFSV2 product outperformed the other soil moisture products and was chosen for the downscaling application. For the York Benny site, both SMAP_25 km and CFSV2 products performed well. The SMAP_25 km product, however, was not selected for the downscaling component due to having the highest RMSE and MAE values. Therefore, the CFSV2 product was selected to be downscaled at the York Benny site due to its performance.

4.2 Validation of the Downscaled Soil Moisture

The downscaled soil moisture estimates were produced across all three study sites. This was necessary in order to thoroughly assess the performance of the downscaling technique. The downscaling section of the study involved analysing a list of machine learning algorithms and selecting the best performing algorithm, which would be used to downscale the selected soil moisture product for each study area. A total of seven algorithms were initially selected, including the ensembles that were created, which were applied to the training dataset. From the results obtained, the three best performing algorithms were chosen for further analysis, which was done using the validation dataset. The best performing machine learning algorithm was then selected for downscaling of the soil moisture product. As mentioned above, the SMAP_25 km product was chosen for downscaling at the Two Streams site, whilst the CFSV2 product was downscaled for both the Mead CSP3 and York Benny sites.

4.2.1 Two Streams study site validation

The statistical results for the seven machine learning algorithms that were used during the downscaling component can be seen in Table 4.4. These results were produced for the training datasets. The statistical analyses used included the R^2 value, MAE and RMSE. The glm, knn and ranger algorithms have the smallest R^2 values when compared to the other algorithms. The MAE and RMSE statistics for these three algorithms were amongst the highest values for the Two Streams study site. Therefore, these algorithms were not selected for downscaling of the SMAP_25 km product. Based on the statistical results, the three best performing algorithms were rpart, ensemble 1 and ensemble 2.

Table 4.4 Statistics for the machine learning algorithms for the training data at the Two Streams site

	glm	**rpart**	**knn**	**svmRadial**	**ranger**	**ensemble 1**	**ensemble 2**
R^2	0.4140	0.5100	0.4200	0.4910	0.4310	0.5080	0.7700
MAE	0.0462	0.0431	0.0465	0.0442	0.0473	0.0405	0.0274
RMSE	0.0568	0.0525	0.0567	0.0557	0.0581	0.0501	0.0355

The three selected algorithms were then evaluated using the validation datasets and the statistics produced can be seen in Table 4.5. The correlation results produced by all three algorithms indicate a poor relationship and are significantly lower than the correlation results that were produced for the training data. Ensemble 2 was the best performing algorithm. It has the highest correlation amongst the three algorithms with a value of 0.183 and the lowest values for the MAE (0.0616) and RMSE (0.0723). Therefore, the ensemble 2 model was chosen for downscaling the SMAP_25 km soil moisture product.

Table 4.5 Statistics for the three best machine learning algorithms for the validation data at the Two Streams site

	ensemble 1	ensemble 2	rpart
Correlation	0.0710	0.1830	0.0473
MAE	0.0724	0.0616	0.0676
RMSE	0.0842	0.0723	0.0807

The downscaled soil moisture produced using the ensemble 2 model was validated against the CRNS data (Figure 4.7). This process was undertaken within the R model. Both datasets (Downscaled soil moisture and CRNS data) have a positive relationship with an R^2 value of 0.34. A Kruskal-wallis test was also done for the data using the R model to indicate whether there was a significant difference between the datasets. The p-value was 0.50, which meant there was no significant difference.

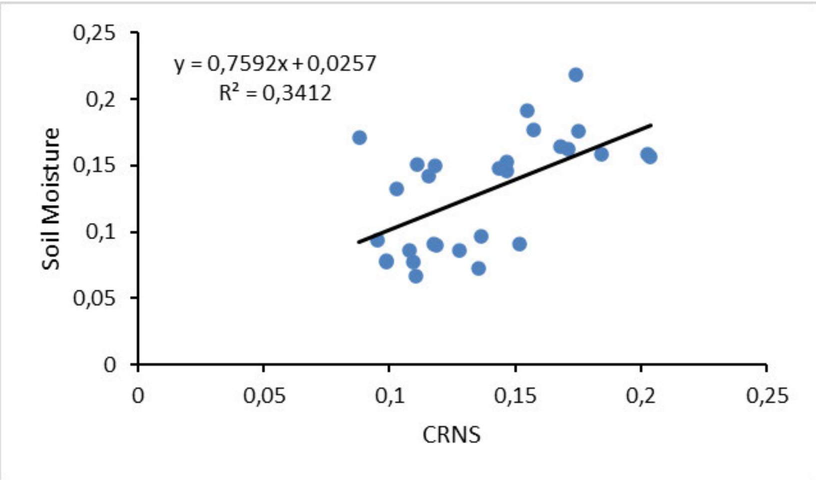

Figure 4.7 Scatter plot illustrating the relationship between the downscaled soil moisture and CRNS for the Two Streams study site

4.2.2 Mead CSP3 study site validation

The statistical analyses for the seven machine learning algorithms can be seen in Table 4.6. The glm, rpart and svmRadial algorithms were not selected for further analysis due to exhibiting the lowest R^2 values and highest MAE and RMSE statistics. The knn, ranger, ensemble 1 and ensemble 2 models produced the best statistical results amongst the seven algorithms. However, the knn algorithm was excluded due to its correlation with some of the other algorithms exceeding 0.75. Therefore, the ranger, ensemble 1 and ensemble 2 models were chosen for further analysis.

Table 4.6 Statistics for the machine learning algorithms for the training data at the Mead CSP3 site

	glm	rpart	knn	svmRadial	ranger	ensemble 1	ensemble 2
R^2	0.4080	0.3810	0.4480	0.4190	0.4410	0.4580	0.5380
MAE	0.0435	0.0447	0.0422	0.0432	0.0423	0.0418	0.0368
RMSE	0.0521	0.0532	0.0502	0.0515	0.0506	0.0498	0.0460

The three algorithms were then evaluated using the validation datasets. The statistical results produced can be seen in Table 4.7. It can be deduced that the ensemble 1 model performed the best amongst the three algorithms with a correlation value of 0.717. The MAE (0.0386) and RMSE (0.0471) values were higher than the other two algorithms. Therefore, the ensemble 1 model was used for downscaling the CFSV2 product at the Mead CSP3 study site.

Table 4.7 Statistics for the three best performing machine learning algorithms for the validation data at the Mead CSP3 site

	ensemble 1	ensemble 2	ranger
Correlation	0.7170	0.6790	0.7020
MAE	0.0386	0.0432	0.0401
RMSE	0.0471	0.0511	0.0488

The results for the downscaled soil moisture derived using the ensemble 1 model were then analysed. The estimates produced were plotted against the CRNS data (Figure 4.8). Both the downscaled soil moisture and CRNS estimates have a positive relationship with an R^2 value of

0.59. This value shows that the downscaled soil moisture corresponds satisfactorily when compared against the CRNS data. The kruskal-wallis test produced a p-value of 0.50 for the data at the Mead CSP3 site, therefore, there was no significant difference between the datasets.

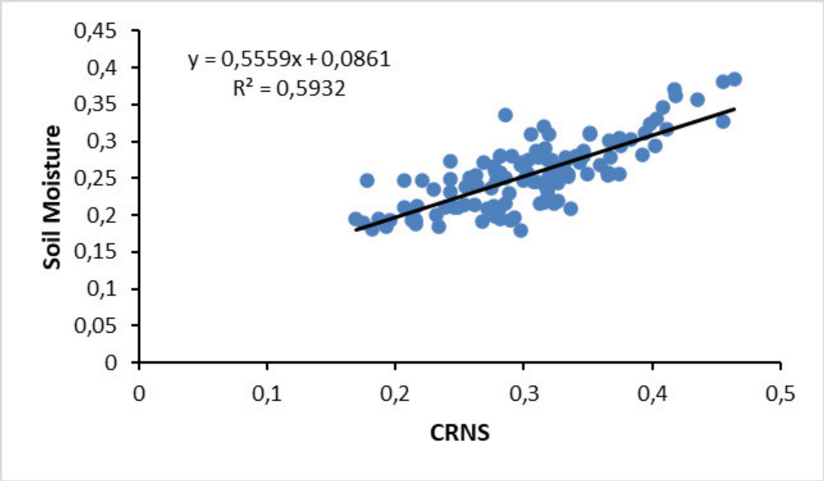

Figure 4.8 Scatter plot illustrating the relationship between the downscaled soil moisture and CRNS for the Mead CSP3 study site

4.2.3 York Benny study site validation

The statistical analyses produced for each of the machine learning algorithms can be seen in Table 4.8. The rpart, knn and svmRadial algorithms produced the lowest R^2 results. The MAE and RMSE for these aforementioned algorithms were amongst the highest, therefore, based on the statistical analyses that were carried out, these products were not used further. The glm, ranger, ensemble 1 and ensemble 2 models performed the best based on the results acquired. The glm and ranger algorithms produced similar results. However, based on the fact that the ranger algorithm performed better than the glm algorithm at the other two study sites, the ranger was selected for further analysis together with the ensemble 1 and ensemble 2 models.

Table 4.8 Statistics for the machine learning algorithms for the training data at the York Benny site

	glm	rpart	knn	svmRadial	ranger	ensemble 1	ensemble 2
R^2	0.3910	0.3280	0.3810	0.3900	0.3910	0.4000	0.5240
MAE	0.0387	0.0415	0.0390	0.0385	0.0385	0.0379	0.0325
RMSE	0.0471	0.0501	0.0476	0.0475	0.0473	0.0464	0.0414

The three selected models were then evaluated using the validation data. The statistical results acquired can be seen in Table 4.9. From the results produced, it can be deduced that the ranger algorithm performs the best with a correlation value of 0.587. The MAE (0.0433) and RMSE (0.0547) were the lowest values amongst the three algorithms. Therefore, the ranger algorithm was used to downscale the CFSV2 product at the York Benny study site.

Table 4.9 Statistics for the three best performing machine learning algorithms for the validation data at the York Benny site

	ensemble 1	**ensemble 2**	**ranger**
Correlation	0.5400	0.4470	0.5870
MAE	0.0452	0.0498	0.0433
RMSE	0.0570	0.0608	0.0547

The downscaled soil moisture results were then validated using the CRNS data. Both datasets had a positive relationship (Figure 4.9). The R^2 value is 0.41, which indicates a relatively poor relationship between the downscaled soil moisture and CRNS estimates. From the kruskal-wallis test, a p-value of 0.50 was derived, which meant that there was no significant difference observed between the datasets.

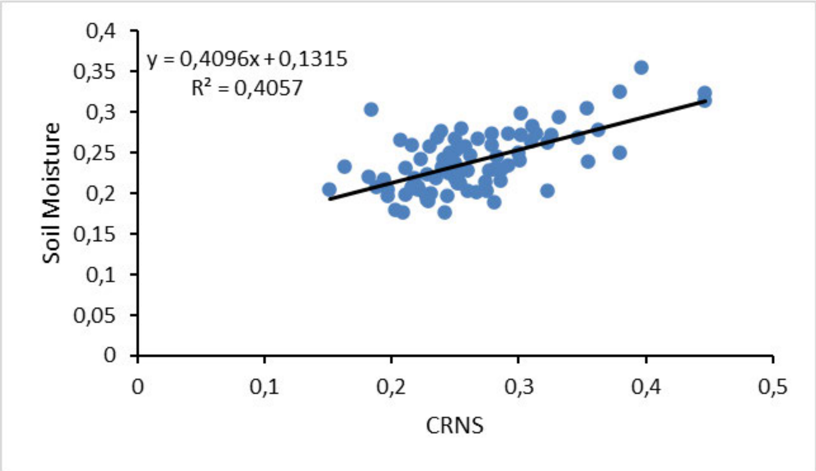

Figure 4.9 Scatter plot illustrating the relationship between the downscaled soil moisture and CRNS for the York Benny study site

4.3 Evaluation of the Downscaled Soil Moisture and Satellite Product

The downscaled soil moisture estimates produced for each study site exhibited an average performance when validated against the CRNS data. For this section, the downscaled soil moisture estimates and original satellite-based product were compared to the CRNS data across all three study sites. This step was undertaken to determine whether the downscaling procedure resulted in improved soil moisture estimates when compared to the original satellite-based data.

Multiple statistics were calculated for the Two Streams site (Table 4.10), comparing both the downscaled soil moisture and the SMAP_25 km product against the CRNS estimates. The SMAP_25 km product produced an r value of 0.83 and R^2 value of 0.68, which were significantly higher than the downscaled results as shown in Table 4.10. The RMSE and MAE, however, indicated improvements with the downscaled product. This is further proven by the ubRMSE of the downscaled product, which lies within the benchmark target of 0.04 m^3 m^{-3}. While the r and R^2 values are relatively higher for the SMAP_25 km product, it is evident that in most of the key performance indicators there was an improvement in results when compared to the original SMAP_25 km product.

Table 4.10 Statistics comparing the downscaled and original SMAP_25 km product against the CRNS at the Two Streams catchment

	Downscaled Product	**SMAP_25 km**
r	0.58	0.83
R^2	0.34	0.68
RMSE	0.04	0.06
ubRMSE	0.03	0.05
MAE	0.03	0.05
Relative Volume (%)	5.31	-6.85

The same analysis was done for the Mead CSP3 study site, which can be seen in Table 4.11. The downscaled product had a higher r value when compared to the CFSV2 reanalysis product. The RMSE and MAE values were lower for the CFSV2 soil moisture product, however, the differences in these statistics between both the downscaled and original product are relatively small. Additionally, the ubRMSE results revealed that the downscaled data lies closer to the benchmark target of 0.04 m^3m^{-3}. A collective analysis of the statistical results indicated that the downscaled soil moisture product did perform better than the CFSV2 reanalysis product.

Table 4.11 Statistics comparing the downscaled and original CFSV2 product against the CRNS at the Mead CSP3 catchment

	Downscaled Product	CFSV2
r	0.77	0.73
R^2	0.59	0.53
RMSE	0.063	0.056
ubRMSE	0.04	0.05
MAE	0.054	0.046
Relative Volume (%)	16.08	-9.44

The downscaled soil moisture data and CFSV2 product were analysed against the CRNS at the York Benny study site (Table 4.12). From an overall analysis of the statistical results that were acquired, it can be deduced that there were similar levels of performance between the downscaled product and CFSV2 soil moisture data.

Table 4.12 Statistics comparing the downscaled and original CFSV2 product against the CRNS at the York Benny catchment

	Downscaled Product	CFSV2
r	0.64	0.66
R^2	0.41	0.44
RMSE	0.051	0.050
ubRMSE	0.04	0.05
MAE	0.040	0.042
Relative Volume (%)	9.19	-4.92

5. DISCUSSION

The aim of this research project was to apply and evaluate a downscaling technique for satellite-based and reanalysis soil moisture products. The acquisition of all datasets and the application of downscaling were facilitated by data processing platforms, GEE and R.

The research questions that needed to be addressed included:

- How does the estimates obtained from each satellite-based and reanalysis soil moisture product compare with the CRNS?
- Does the application of a downscaling technique to the satellite-based and reanalysis data produce a better outcome?
- Are the data processing platforms used for downscaling soil moisture feasible?
- How does the soil moisture data obtained from all three methods (i.e. Satellite-based, reanalysis and downscaled estimates) compare across two different regions given their different climate and land use?

Multiple soil moisture products were chosen for this study to be validated. This included both satellite-based and reanalysis products. The main purpose for choosing and evaluating many products was to be able to select a soil moisture product that performs best at a particular site, which could then be downscaled. This step was crucial as a product that performs poorly will unlikely result in improved estimates when downscaled due to the possibility of the issue being related to the conceptualisation of the product itself. Additionally, comparisons between satellite-based and reanalysis products were made to determine which of the two methods is more suitable for soil moisture estimation.

The selected satellite-based and reanalysis soil moisture products for this study were validated against the CRNS data at each study site. The results acquired for the Two Streams catchment showed that all the soil moisture products exhibited seasonal variation in soil moisture, which increased towards the summer months and were relatively low during winter. The SMAP_25 km product produced the best results and corresponded well against the CRNS *in-situ* data. However, it was evident that all the soil moisture products either overestimated or underestimated soil moisture. The Two Streams site is dominated by commercial plantations of black wattle, which may have potentially affected the performance of the satellite-based soil moisture products. A study by Everson *et al*. 2014, investigated the impact of wattle trees on hydrological processes at the Two Streams site. This study highlighted the fact that interception due to canopy and litter cover results in only approximately 65 % of the total precipitation being available water that drains into the soil. It was also found that the tree water use is highly variable to the aspect and slope across the catchment. Therefore, it is possible that the soil moisture retrieval algorithms are not sensitive to such factors.

The Mead CSP3 and York Benny sites both had similar trends for all the soil moisture products that were investigated. This can be expected as both sites are exposed to similar climatic conditions by being situated in eastern side of Nebraska and are also dominated by the same vegetation cover (corn and soybean). Both these sites depicted variation in soil moisture content throughout the period of study with no distinctive pattern. Rainfall distribution is relatively higher during the summer months at these catchments and low in winter, however, the soil moisture content derived from all the products did not correspond with this trend. This was expected for the York Benny site due to the area being irrigated, however, the same situation occurred for the Mead CSP3 site despite this catchment being rainfed.

From the results acquired across all three catchments, it is evident that CFSV2 and SMAP_25 km produced the best results when compared to the other soil moisture products. This deduction is based on the collective interpretation of both the graphical and statistical analyses. A study by Yang *et al.* (2021) evaluating satellite-based and reanalysis soil moisture products found that the latter performed better than the satellite-based products, with the ERA5 reanalysis product producing the best results. The creation of reanalysis soil moisture products involves an assimilation of data, which also includes ground-based observations that are generally associated with high accuracy. With satellite-based soil moisture products, however, the acquisition of data is undertaken using a retrieval algorithm, which could easily result in estimation errors for many reasons such as issues with the algorithm itself as well as the characteristics of the study area (eg. Climate and vegetation conditions). Therefore, it is expected that reanalysis products will outperform satellite-based products, which was the case in the aforementioned study by Yang *et al*. (2021) and a previous study by Peng *et al*. (2015).

Due to these findings, it was expected that the reanalysis products will produce the best results. More specifically, it was expected that the ERA5 reanalysis product would perform well due to its development via data assimilation as well as its finer spatial resolution of 10 km. However, based on the results obtained in this study, while the CFSV2 reanalysis product did show satisfactory performance, the ERA5 product performed poorly when validated against the CRNS. These findings could possibly be related to the fact that the average monthly estimates generated for each product are based on different population sizes, which may have influenced the data comparisons and impacted the representativeness of the products for estimating soil moisture. Due to the different population sizes, some products may have better captured the variability in soil moisture as compared to other products. Additionally, factors such as climate and surface conditions including topography and soil properties may have also affected the performance of the product. Most of the studies investigating the ERA5 product were undertaken in China (Peng *et al*., 2015; Ling *et al*., 2021; Wu *et al*., 2021; Yang *et al*., 2021), therefore, it is possible that the product is sensitive to climatic and surface factors, which could explain it performing well across areas in China but poorly for the Nebraska and South Africa sites chosen for this study.

Based on the validation results obtained for the satellite-based and reanalysis soil moisture products, the best performing product was selected for each of the study sites for the downscaling application of this study. The SMAP_25 km product was selected for the Two Streams site, whilst the CFSV2 product was chosen for the Mead CSP3 and York Benny sites. Downscaling of soil moisture has the potential of addressing the issue of scale mismatch between *in-situ* and coarse resolution products, which could potentially result in improved soil moisture estimates from satellite-based and reanalysis products. Five machine learning algorithms, excluding the ensemble models created, were initially selected and were each evaluated using the training datasets. Additionally, the machine learning algorithms were also used to create ensembles models to downscale soil moisture, which did produce good results across all three study sites. Based on the performance of the selected algorithms for the training and validation datasets, the most suitable model was then used to downscale the selected soil moisture product. This procedure was undertaken individually for each study site to identify which model is best suited to each of the catchments. During the selection process, it was found that a different model was selected for each of the three study sites based on their performance statistics, indicating that a catchment's characteristics does influence the performance of these algorithms.

The downscaling datasets were then validated against the CRNS *in-situ* data at each study site. Based on the statistical results acquired at both the Two Streams and Mead CSP3 study sites, it was revealed that the downscaled products exhibited better performance than the original soil moisture products. With regards to the York Benny study site, both the downscaled product and the CFSV2 data had similar levels of performance. From these results, it is evident that the application of downscaling does show potential, as improvements in soil moisture estimation were evident at both the Two Streams and Mead CSP3 sites. While the results for the York Benny site indicated that both the downscaled and original products had similar levels of performance, some key indicators such as the ubRMSE and MAE had shown slight improvements for the downscaled soil moisture. Despite the potential shown by the application of downscaling, it is evident from the statistical analyses that the extent of improvement was relatively small, which could be the case due to a number of factors.

One of the reasons that may have influenced the performance of the downscaling technique is the initial relationships between the soil moisture product, NDVI and LST. These results were computed using the R tool prior to the downscaling application (See Appendix). It was revealed that these relationships had an average performance across all three study sites. This could have attributed to the moderate performance of the downscaled product. A stronger relationship between the soil moisture product and fine scale variables may have resulted in better downscaled estimates. A study by Fang *et al*. (2018), which involved downscaling SMAP data using LST and NDVI estimates, recommended the use of other fine scale variables to downscale soil moisture. This is due to the fact that other factors in addition to vegetation and temperature may also influence soil moisture variability and certain factors at a particular study site may have a greater impact on soil moisture as compared to other variables.

The performance of the downscaled products may have also been influenced by the data availability for the soil moisture products and CRNS validation data at each of the three study sites. The Mead CSP3 and York Benny sites both had long data records of CFSV2 for training and testing the algorithms. In addition, the CRNS validation datasets were also sufficient with the Mead CSP3 site having CRNS data extending from 2011 and the York Benny site having CRNS data from 2014. However, with regards to the Two Streams site, the data records were much shorter. Due to the fact that the SMAP_25 km product was selected at this catchment, only a five-year record of data was available for training and testing the algorithms as data was only produced from 2015 till 2020. Furthermore, the CRNS validation dataset was also a short record of two years (2015 – 2017). The Mead CSP3 site having the longest data records, may have allowed for a more thorough evaluation in terms of the variability of soil moisture as compared to the Two Streams site, which was only validated for a two-year period. Therefore, the short data records used for the Two Streams site may have influenced the results obtained, which could possibly explain the low r and R^2 results that were obtained for the downscaled product. However, despite this limitation, the inclusion of other key statistics did reveal improvements with the application of downscaling at this site.

Looking specifically at the Mead CSP3 site where the downscaled product performed better than the original CFSV2 product, even though the downscaled product did perform better than the CFSV2 product, it was only a small improvement, which could have been attributed to the conceptualisation of the reanalysis product. Very few studies on the evaluation of reanalysis products with regards to soil moisture have been carried out. This particularly applies to the CFSV2 product, however, a study by Tian *et al*. (2016) did perform an evaluation on downscaling of the CFSV2 product for the forecasting of precipitation and temperature. The study deduced that downscaling did not result in an improvement of the CFSV2 data. Therefore, this product does require further investigation to fully understand it and be able to determine whether downscaling is influenced more by the conceptualisation of the product or the catchment conditions. This also applies to the SMAP_25 km soil moisture product. While the application of downscaling to the Two Streams data did result in an overall improvement, the differences in the statistics between the downscaled and original product were relatively small.

While downscaling approaches are often recommended in studies, certain factors generally limit many researchers from applying such techniques, which include the complexity of some techniques, the requirement of sufficient *in-situ* data to be able to train and test algorithms and the acquisition of satellite-based data being resource and computationally intensive. Therefore, this study also aimed at addressing this issue using GEE and R. GEE was used to obtain all the datasets used in this study. Additionally, it was used to upscale the NDVI and LST data from MODIS. The R tool was then used for the application of machine learning techniques to downscale soil moisture. Both these data processing platforms show significant potential in removing many of the previous aforementioned barriers.

The acquisition of remote sensing data usually takes an extensive amount of time to obtain and process, however, with the use of GEE, this process is much more time efficient as data that normally takes several hours to acquire can now be retrieved within minutes. Additionally, GEE is a cloud computing platform, which eliminates the need for advanced modern software allowing many researchers with financial constraints to conduct desired studies. The downscaling application was undertaken with ease due to the facilitation of machine learning using R. The R tool is freely available and open source making it possible for many researchers to conduct analyses and allowing developers to share packages they have created via this tool.

While both these platforms have proven to carry out such processes with ease, GEE and R are different language platforms that each require user training prior to application. Therefore, it would have been beneficial to have a single platform that could conduct all the analyses that were undertaken using GEE and R in this study.

6. CONCLUSIONS AND RECOMMENDATIONS

This chapter entails the conclusions and recommendations of the research that was undertaken for this study.

6.1 Conclusions

Soil moisture is essential to monitor mainly for the purpose of sustaining water resources in the long term. It has several applications across different sectors, which makes it crucial to produce reliable soil moisture data and be monitored on a regular basis. There are three methods for deriving soil moisture content, namely, *in-situ*, remote sensing and modelling. Each of these methods have their respective advantages and limitations. Remote sensing satellite-based and modelled reanalysis soil moisture products can produce data over large spatial scales over a short period of time, however, they are limited by their coarse resolutions. Therefore, to overcome this limitation, remote sensing and reanalysis data are validated using *in-situ* methods of soil moisture. The scale mismatch between the coarse resolution products and *in-situ* data, however, makes it unsuitable to use *in-situ* estimates for the validation of these soil moisture products.

Therefore, the objective of this study was firstly to implement a downscaling technique to produce finer resolution soil moisture and address the scale mismatch issue that currently exists. This procedure was facilitated by data processing platforms, GEE and R. Additionally, the performance of satellite-based and reanalysis products in deriving soil moisture data were also investigated. While this step was necessary for the downscaling application, it was also of importance to determine whether satellite-based or reanalysis products are more suitable for soil moisture estimation. Moreover, with regards to reanalysis products, there has not been much research undertaken to assess their performance. This research was undertaken at three study sites, with one being in South Africa and the other two sites being situated in the USA. The location of these sites allowed for comparisons to be made in terms of climate and vegetation.

The first component of the study involved the evaluation of the satellite-based and reanalysis soil moisture products using the CRNS instrument. Based on the results that were obtained for this study, it is evident that some soil moisture products perform better than others. When comparing the satellite-based products to the reanalysis products, there is no definitive answer as to which of these methods produce more reliable soil moisture estimates. This is due to the fact that each product performs differently. For example, the ERA5 performed poorly across the three study sites, however, the CFSV2 product was amongst the best performing products. The same was deduced for the satellite-based soil moisture products. Additionally, it is also evident that the performance of these soil moisture products is dependent on the given circumstances as each product performed differently across all three study sites. This could have been due to differences in catchment conditions such as climate, vegetation, topography and soil properties. From the validation procedure, the SMAP_25 km and CFSV2 products had the best performance against the CRNS data.

With regards to the downscaling application of the study, the downscaled datasets for each study site were validated against the CRNS data. It was deduced that the downscaling application was successful for the Two Streams and Mead CSP3 study sites. With regards to the York Benny site, there was no definitive outcome as the downscaled product performed similarly to the original CFSV2 product. While the downscaled products for the Two Streams and Mead CSP3 sites did perform better than the original products, the statistical improvements were relatively small. The reasoning owing to the performance of the downscaling technique includes potential issues that could be related to the conceptualisation of the original soil moisture products, data availability and the relationship between the soil moisture product and fine scale variables.

Through the conduction of this study, it can be deduced that downscaling can be undertaken with ease using data processing platforms such as GEE and R. With regards to the results obtained for the downscaling of soil moisture data, it is evident that downscaling does produce better results in some cases, however, this is not always the outcome. For this study, the performance of the downscaled products may have been influenced by certain factors that were mentioned above. As a result, further studies are recommended to investigate whether the application of downscaling procedures truly offers any improvements in the accuracy of soil moisture estimates at localized levels. Such research is particularly relevant as improved estimation of soil moisture can be used to improve planning and management operations for various purposes.

6.2 Recommendations

The following recommendations listed below can be used in future research to address the limitations that were encountered during this study:

i. The scale mismatch between the remote sensing and *in-situ* methods still exists due to the coarse resolution associated with satellite-based and reanalysis soil moisture products. Downscaling of soil moisture may reduce the scale mismatch by disaggregating coarse resolution products and produce more reliable outcomes. The application of downscaling in this study did result in improvements across two study sites, and hence, holds significant potential for future studies and is highly recommended.

ii. NDVI and LST were the selected variables for the downscaling application of this study, however, the use of other fine scale variables related to soil moisture should be investigated.

iii. While the use of GEE and R for acquiring data in future studies is highly encouraged, the use of a single platform to conduct all the analyses rather than having to use two separate platforms and languages may also be beneficial. The models that were used were subjective in this study, however, are adaptable for other products.

iv. For this study, average monthly estimates were generated for each product that were based on different population sizes, which may have influenced the data comparisons. Further investigations into the temporal resolutions of the soil moisture products could potentially address whether the averaging of such datasets affects their performance.

v. The CRNS *in-situ* data was used for the validation of the soil moisture products and downscaled datasets. However, longer validation data lengths are recommended. This applies particularly for the training and testing of downscaling algorithms to produce more representative relationships and ultimately strengthen the performance of these algorithms.

vi. An assessment of the CRNS data that was used for validation purposes in this study, was not undertaken due to not having access to gravimetric soil moisture or TDR data. Therefore, this investigation is recommended for future studies to determine whether the scale of the CRNS did offer any improvements for validation.

7. REFERENCES

Acharya, U, Daigh, AL and Oduor, PG. 2021. Machine learning for predicting field soil moisture using soil, crop, and nearby weather station data in the Red River Valley of the north. *Soil Systems*, *5*(4), p.57.

Adab, H, Morbidelli, R, Saltalippi, C, Moradian, M and Ghalhari, GAF. 2020. Machine learning to estimate surface soil moisture from remote sensing data. *Water*, *12*(11), p.3223.

Alemohammad, SH, Kolassa, J, Prigent, C, Aires, F and Gentine, P. 2018. Global downscaling of remotely sensed soil moisture using neural networks. *Hydrology and Earth System Sciences*, *22*(10), pp.5341-5356.

Al-Yaari, A, Wigneron, JP, Ducharne, A, Kerr, YH, Wagner, W, De Lannoy, G, Reichle, R, Al Bitar, A, Dorigo, W, Richaume, P and Mialon, A. 2014. Global-scale comparison of passive (SMOS) and active (ASCAT) satellite based microwave soil moisture retrievals with soil moisture simulations (MERRA-Land). *Remote sensing of environment* 152**,** 614-626.

Babaeian, E, Sadeghi, M, Franz, TE, Jones, S and Tuller, M. 2018. Mapping soil moisture with the OPtical TRApezoid Model (OPTRAM) based on long-term MODIS observations. *Remote sensing of environment* 211**,** 425-440.

Berk Duygu, M and Akyurek, Z. 2019. Using Cosmic-ray Neutron Probes in validating the satellite soil moisture products and land surface models. *EGUGA***,** 1060.

Bezak, N, Jemec Auflič, M and Mikoš, M. 2021. Reanalysis of soil moisture used for rainfall thresholds for rainfall-induced landslides: The Italian case study. *Water*, *13*(14), p.1977.

Bischl, B, Lang, M, Kotthoff, L, Schiffner, J, Richter, J, Studerus, E, Casalicchio, G and Jones, ZM. 2016. mlr: Machine learning in R. *The Journal of Machine Learning Research*, *17*(1), pp.5938-5942.

Blöschl, G, Komma, J and Hasenauer, S. 2009. Hydrological downscaling of soil moisture. *Final Report to the H-SAF (Hydrology Satellite Application Facility) via the Austrian Central Institute for Meteorology and Geodynamics (ZAMG), Vienna University of Technology, 1040,* p.15.

Brocca, L, Crow, WT, Ciabatta, L, Massari, C, De Rosnay, P, Enenkel, M, Hahn, S, Amarnath, G, Camici, S, Tarpanelli, A and Wagner, W. 2017. A review of the applications of ASCAT soil moisture products. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10(5)**,** 2285-2306.

Brocca, L, Filippucci, P, Hahn, S, Ciabatta, L, Massari, C, Camici, S, Schuller, L, Bojkov, B and Wagner, W. 2019. SM2RAIN–ASCAT (2007–2018): global daily satellite rainfall data from ASCAT soil moisture observations. *Earth System Science Data* 11(4).

Bulcock, HH and Jewitt, GPW. 2012. Field data collection and analysis of canopy and litter interception in commercial forest plantations in the KwaZulu-Natal Midlands, South Africa. *Hydrology & Earth System Sciences*, *16*(10).

Chan, SK, Bindlish, R, O'Neill, PE, Njoku, E, Jackson, T, Colliander, A, Chen, F, Burgin, M, Dunbar, S, Piepmeier, J and Yueh, S. 2016. Assessment of the SMAP passive soil moisture product. *IEEE Transactions on Geoscience and Remote Sensing* 54(8)**,** 4994-5007.

Chen, F, Crow, WT, Bindlish, R, Colliander, A, Burgin, MS, Asanuma, J and Aida, K. 2018. Global-scale evaluation of SMAP, SMOS and ASCAT soil moisture products using triple collocation. *Remote sensing of environment* 214**,** 1-13.

Chen, S, She, D, Zhang, L, Guo, M and Liu, X. 2019. Spatial downscaling methods of soil moisture based on multisource remote sensing data and its application. *Water* 11(7)**,** 1401.

Cheng, M, Zhong, L, Ma, Y, Zou, M, Ge, N, Wang, X and Hu, Y. 2019. A study on the assessment of multi-source satellite soil moisture products and reanalysis data for the Tibetan Plateau. *Remote sensing* 11(10)**,** 1196.

Ciabatta, L, Massari, C, Brocca, L, Gruber, A, Reimer, C, Hahn, S, Paulik, C, Dorigo, W, Kidd, R and Wagner, W. 2018. SM2RAIN-CCI: a new global long-term rainfall data set derived from ESA CCI soil moisture. *Earth System Science Data* 10(1)**,** 267.

Clark, S. 2015. Radar on NASA's SMAP Earth observing satellite declared lost. [Internet]. Spaceflight Now. Available from: https://spaceflightnow.com/2015/09/03/radar-on-nasas-smap-earth-observing-satellite-declared-lost/. [Accessed: 26 June 2020].

Clulow, AD, Everson, CS and Gush, MB. 2011. The long-term impact of Acacia mearnsii trees on evaporation, streamflow and ground water resources. *Water Research Commission Report No. TT505/11, ISBN*, pp.978-1.

COSMOS. 2021. York Irrigated Maize. [Internet]. COSMOS. Available from: http://cosmos.hwr.arizona.edu/Probes/StationDat/090/index.php. [Accessed: 17 March 2021].

Cui, C, Xu, J, Zeng, J, Chen, KS, Bai, X, Lu, H, Chen, Q and Zhao, T. 2018. Soil moisture mapping from satellites: An intercomparison of SMAP, SMOS, FY3B, AMSR2, and ESA CCI over two dense network regions at different spatial scales. *Remote sensing* 10(1)**,** 33.

Cui, H, Jiang, L, Du, J, Zhao, S, Wang, G, Lu, Z and Wang, J. 2017. Evaluation and analysis of AMSR-2, SMOS, and SMAP soil moisture products in the Genhe area of China. *Journal of Geophysical Research: Atmospheres*, *122*(16), pp.8650-8666.

Dandridge, C, Fang, B and Lakshmi, V. 2019. Downscaling of SMAP soil moisture in the lower mekong river basin. *Water*, *12*(1), p.56.

DeAngelis, K. 2007. Measurement of soil moisture content by gravimetric method. *American Society of Agronomy***,** 1-2.

Dee, DP, Uppala, SM, Simmons, A, Berrisford, P, Poli, P, Kobayashi, S, Andrae, U, Balmaseda, MA, Balsamo, G, Bauer, DP and Bechtold, P. 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society* 137(656)**,** 553-597.

Djamai, N, Magagi, R, Goïta, K, Merlin, O, Kerr, Y and Roy, A. 2016. A combination of DISPATCH downscaling algorithm with CLASS land surface scheme for soil moisture estimation at fine scale during cloudy days. *Remote Sensing of Environment*, *184*, pp.1-14.

Dorigo, W., Gruber, A., De Jeu, R., Wagner, W., Stacke, T., Loew, A, Albergel, C, Brocca, L, Chung, D, Parinussa, RM and Kidd, R. 2015. Evaluation of the ESA CCI soil moisture product using ground-based observations. *Remote sensing of environment* 162**,** 380-395.

Duygu, MB and Akyürek, Z. 2019. Using Cosmic-Ray Neutron Probes in Validating Satellite Soil Moisture Products and Land Surface Models. *Water* 11(7), 1362.

Edaphic Scientific. 2020. Time Domain Reflectometry (TDR). [Internet]. Edaphic Scientific. Available from: https://www.edaphic.com.au/soil-water-compendium /time -domain-reflectometry-tdr/. [Accessed: 07 May 2020].

El Hajj, M, Baghdadi, N, Zribi, M, Rodríguez-Fernández, N, Wigneron, JP, Al-Yaari, A, Al Bitar, A, Albergel, C and Calvet, JC. 2018. Evaluation of SMOS, SMAP, ASCAT and Sentinel-1 soil moisture products at sites in Southwestern France. *Remote sensing* 10(4)**,** 569.

Entekhabi, D, Njoku, EG, O'Neill, PE, Kellogg, KH, Crow, WT, Edelstein, WN, Entin, JK, Goodman, SD, Jackson, TJ, Johnson, J and Kimball, J. 2010. The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE* 98(5)**,** 704-716.

European Centre for Medium-Range Weather Forecasts (ECMWF). 2020. Climate Reanalysis. [Internet]. ECMWF. Available from: https://www.ecmwf.int/en/research/ climate-reanalysis. [Accessed: 02 June 2020].

European Space Agency (ESA). 2012. Soil Moisture CCI. [Internet]. ESA. Available from: https://www.esa-soilmoisture-cci.org/node/139. [Accessed: 26 June 2020].

Everson, C. 2016. New technology provides validation data for flood forecasting models. [Internet]. SAEON. Available from: https://www.ee.co.za/article/new-technology-provides-validation-data-flood-forecasting-model-2.html. [Accessed: 26 June 2020].

Everson, CS, Clulow, AD, Becker, M, Watson, A, Ngubo, C, Bulcock, H, Mengistu, M, Lorentz, S and Demlie, M. 2014. The long-term impact of Acacia mearnsii trees on evaporation, streamflow, low flows and ground water resources. Phase II: Understanding the controlling environmental variables and soil water processes over a full crop rotation. Pretoria, South Africa.

Fang, B, Lakshmi, V, Bindlish, R and Jackson, TJ. 2018. Downscaling of SMAP soil moisture using land surface temperature and vegetation data. *Vadose Zone Journal*, *17*(1), pp.1-15.

Fontanet Ambrós, M, Fernández García, D and Ferrer, F. 2018. The value of satellite remote sensing soil moisture data and the DISPATCH algorithm in irrigation fields. *Hydrology and Earth system sciences* 22(11)**,** 5889-5900.

Foolad, F, Franz, TE, Wang, T, Gibson, J, Kilic, A, Allen, RG and Suyker, A. 2017. Feasibility analysis of using inverse modeling for estimating field-scale evapotranspiration in maize and soybean fields from soil water content monitoring networks. *Hydrology and Earth System Sciences*, *21*(2), pp.1263-1277.

Franz, TE, Pokal, S, Gibson, JP, Zhou, Y, Gholizadeh, H, Tenorio, FA, Rudnick, D, Heeren, D, McCabe, M, Ziliani, M and Jin, Z. 2020. The role of topography, soil, and remotely sensed vegetation condition towards predicting crop yield. *Field Crops Research*, *252*, p.107788.

Franz, TE, Wahbi, A, Vreugdenhil, M, Weltin, G, Heng, L, Oismueller, M, Strauss, P, Dercon, G and Desilets, D. 2016. Using cosmic-ray neutron probes to monitor landscape scale soil water content in mixed land use agricultural systems. *Applied and Environmental Soil Science* 2016.

Franz, TE, Wang, T, Avery, W, Finkenbiner, C and Brocca, L. 2015. Combined analysis of soil moisture measurements from roving and fixed cosmic ray neutron probes for multiscale real-time monitoring. *Geophysical Research Letters*, *42*(9), pp.3389-3396.

Franz, TE, Zreda, M, Ferre, T, Rosolem, R, Zweck, C, Stillman, S, Zeng, X and Shuttleworth, WJ. 2012a. Measurement depth of the cosmic ray soil moisture probe affected by hydrogen from various sources. *Water Resources Research* 48(8).

Franz, TE, Zreda, M, Rosolem, R and Ferre, T. 2012b. Field validation of a cosmic-ray neutron sensor using a distributed sensor network. *Vadose Zone Journal* 11(4)**,** vzj2012.0046.

Gokool, S, Kunz, RP and Toucher, M. 2022. Deriving moderate spatial resolution leaf area index estimates from coarser spatial resolution satellite products. *Remote Sensing Applications: Society and Environment*, *26*, p.100743.

González-Zamora, Á, Sánchez, N, Pablos, M and Martínez-Fernández, J. 2019. CCI soil moisture assessment with SMOS soil moisture and in situ data under different environmental conditions and spatial scales in Spain. *Remote sensing of environment* 225**,** 469-482.

Gorelick, N, Hancher, M, Dixon, M, Ilyushchenko, S, Thau, D and Moore, R. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, *202*, pp.18-27.

Griesfeller, A, Lahoz, WA, de Jeu, RA, Dorigo, W, Haugen, L, Svendby, TM and Wagner, W. 2016. Evaluation of satellite soil moisture products over Norway using ground-based observations. *International journal of applied earth observation and geoinformation* 45**,** 155-164.

Gruber, A, De Lannoy, G, Albergel, C, Al-Yaari, A, Brocca, L, Calvet, JC, Colliander, A, Cosh, M, Crow, W, Dorigo, W and Draper, C. 2020. Validation practices for satellite soil moisture retrievals: What are (the) errors?. *Remote sensing of environment*, *244*, p.111806.

Guevara, M and Vargas, R. 2019. Downscaling satellite soil moisture using geomorphometry and machine learning. *PloS one*, *14*(9), p.e0219639.

Huang, F, Wang, P, Ren, Y and Liu, R. 2019. Estimating soil moisture using the Optical Trapezoid Model (OPTRAM) in a semi-arid area of SONGNEN plain, China based On Landsat-8 data. In: *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*: IEEE), 7010-7013.

International Atomic Energy Agency (IAEA). 2020. Using cosmic rays to measure moisture levels in soil [Internet]. IAEA. Available from: https://www.iaea.org/newscenter/news/using-cosmic-rays-to-measure-moisture-levels-in-soil. [Accessed: 26 June 2020].

Japan Aerospace Exploration Agency (JAXA). 2013. SHIZUKU (GCOM-W1) to Provide Geophysical Quantity Products. [Internet]. JAXA. Available from: https://global.jaxa.jp/press/2013/05/20130517_shizuku_e.html. [Accessed: 26 June 2020].

Jet Propulsion Laboratory. 2020. SMAP Overview. [Internet]. NASA. Available from: https://smap.jpl.nasa.gov/observatory/overview/. [Accessed: 16 May 2020].

Johnson, AE. 1962. Methods of measuring soil moisture in the field: US Geology. *Survey Water-Supply Paper*, *1619*.

Kerr, Y.H., Waldteufel, P., Richaume, P., Wigneron, J.P., Ferrazzoli, P., Mahmoodi, A, Al Bitar, A, Cabot, F, Gruhier, C, Juglea, SE and Leroux, D. 2012. The SMOS soil moisture retrieval algorithm. *IEEE Transactions on Geoscience and Remote Sensing* 50(5)**,** 1384-1403.

Kim, D, Moon, H, Kim, H, Im, J. and Choi, M. 2018. Intercomparison of downscaling techniques for satellite soil moisture products. *Advances in Meteorology* 2018.

Kovačević, J, Cvijetinović, Ž, Stančić, N, Brodić, N and Mihajlović, D. 2020. New downscaling approach using ESA CCI SM products for obtaining high resolution surface soil moisture. *Remote Sensing*, *12*(7), p.1119.

Kuhn, M. 2008. Building predictive models in R using the caret package. *Journal of statistical software*, *28*, pp.1-26.

Lakshmi, V. 2013. Remote sensing of soil moisture. *ISRN Soil Science* 2013.

Lantz, B. 2019. Machine learning with R: expert techniques for predictive modeling. Packt publishing ltd, 2019.

Lejiang, Y, Zhanhai, Z, Mingyu, Z, Zhong, S, Lenschow, D, Hsu, H, Huiding, W and Bo, S. 2010. Validation of ECMWF and NCEP-NCAR reanalysis data in Antarctica. *Advances in Atmospheric Science*, *27*(5), pp.1151-1168.

Li, H, Robock, A, Liu, S, Mo, X and Viterbo, P. 2005. Evaluation of reanalysis soil moisture simulations using updated Chinese soil moisture observations. *Journal of Hydrometeorology*, *6*(2), pp.180-193.

Ling, X, Huang, Y, Guo, W, Wang, Y, Chen, C, Qiu, B, Ge, J, Qin, K, Xue, Y and Peng, J. 2021. Comprehensive evaluation of satellite-based and reanalysis soil moisture products using in situ observations over China. *Hydrology and Earth System Sciences*, *25*(7), pp.4209-4229.

Liu, Y, Jing, W, Wang, Q and Xia, X. 2020. Generating high-resolution daily soil moisture by using spatial downscaling techniques: A comparison of six machine learning algorithms. *Advances in Water Resources*, *141*, p.103601.

Liu, Y, Yang, Y and Yue, X. 2018. Evaluation of satellite-based soil moisture products over four different continental in-situ measurements. *Remote Sensing*, *10*(7), p.1161.

Loew, A and Mauser, W. 2008. On the disaggregation of passive microwave soil moisture data using a priori knowledge of temporally persistent soil moisture fields. *IEEE Transactions on Geoscience and Remote Sensing* 46(3)**,** 819-834.

Lu, CH, Kanamitsu, M, Roads, JO, Ebisuzaki, W, Mitchell, KE and Lohmann, D. 2005. Evaluation of soil moisture in the NCEP–NCAR and NCEP–DOE global reanalyses. *Journal of Hydrometeorology* 6(4)**,** 391-408.

Mark Altaweel. 2019. How Google Earth Engine Has Changed Access to Remote Sensing Data. [Internet]. GIS Lounge. Available from: https://www.gislounge.com/how-google-earth-engine-has-changed-access-to-remote-sensing-data/. [Accessed: 09 May 2021].

Mohanty, BP, Cosh, MH, Lakshmi, V and Montzka, C. 2017. Soil moisture remote sensing: State-of-the-science. *Vadose Zone Journal* 16(1)**,** 1-9.

Montzka, C, Bogena, HR, Zreda, M, Monerris, A, Morrison, R, Muddu, S and Vereecken, H. 2017. Validation of spaceborne and modelled surface soil moisture products with cosmic-ray neutron probes. *Remote sensing* 9(2)**,** 103.

Mousa, B and Shu, H. 2020. Spatial evaluation and assimilation of SMAP, SMOS, and ASCAT satellite soil moisture products over Africa using statistical techniques. *Earth and Space Science* 7(1)**,** e2019EA000841.

Mutanga, O and Kumar, L. 2019. Google earth engine applications. *Remote Sensing*, *11*(5), p.591.

Myeni, L, Moeletsi, M and Clulow, A. 2019. Present status of soil moisture estimation over the African continent. *Journal of Hydrology: Regional Studies* 21**,** 14-24.

National Aeronautics and Space Administration (NASA). 2020. Meteorological Operational Satellite A. [Internet]. NASA. Available from: https://www.nasa.gov/directorates/heo/scan/services/missions/earth/MetOpA.html. [Accessed: 26 June 2020].

Parinussa, RM, Holmes, TR, Wanders, N, Dorigo, WA and de Jeu, RA. 2015. A preliminary study toward consistent soil moisture from AMSR2. *Journal of Hydrometeorology* 16(2)**,** 932-947.

Paul, S and Singh, S. 2020. Soil Moisture Prediction Using Machine Learning Techniques. In: *2020 The 3rd International Conference on Computational Intelligence and Intelligent Systems* (pp. 1-7).

Peng, J, Loew, A, Merlin, O and Verhoest, NE. 2017. A review of spatial downscaling of satellite remotely sensed soil moisture. *Reviews of Geophysics* 55(2)**,** 341-366.

Peng, J, Niesel, J, Loew, A, Zhang, S and Wang, J. 2015. Evaluation of satellite and reanalysis soil moisture products over southwest China using ground-based measurements. *Remote sensing* 7(11)**,** 15729-15747.

Philip Briggs. 2020. Weather and Climate: South Africa. [Internet]. Safari Bookings. Available from: https://www.safaribookings.com/south-africa/climate. [Accessed: 08 July 2020].

Poméon, T, Diekkrüger, B, Springer, A, Kusche, J and Eicker, A. 2018. Multi-Objective Validation of SWAT for Sparsely-Gauged West African River Basins—A Remote Sensing Approach. *Water* 10(4)**,** 451.

Sabaghy, S, Walker, JP, Renzullo, LJ and Jackson, TJ. 2018. Spatially enhanced passive microwave derived soil moisture: Capabilities and opportunities. *Remote sensing of environment* 209**,** 551-580.

Sadeghi, M, Babaeian, E, Tuller, M and Jones, SB. 2017. The optical trapezoid model: A novel approach to remote sensing of soil moisture applied to Sentinel-2 and Landsat-8 observations. *Remote sensing of environment* 198**,** 52-68.

Sarker, IH. 2021. Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, *2*(3), pp.1-21.

Schrön, M, Köhli, M, Scheiffele, L, Iwema, J, Bogena, HR, Lv, L, Martini, E, Baroni, G, Rosolem, R, Weimar, J and Mai, J. 2017. Improving calibration and validation of cosmic-ray neutron sensors in the light of spatial sensitivity. *Hydrology and Earth System Sciences*, *21*(10), pp.5009-5030.

Scott-Shaw, BC, Hill, TR and Gillham, JS. 2020. Calibration of a modelling approach for sediment yield in a wattle plantation, KwaZulu-Natal, South Africa. *Water SA*, *46*(2), pp.171-181.

Smajstrla, AG, Harrison, DS and Duran, FX. 1981. Tensiometers for soil moisture measurement and irrigation scheduling. *Circular-Florida Cooperative Extension Service (USA)*.

South African Government. 2020. Geography and Climate. [Internet]. South African Government. Available from: https://www.gov.za/about-sa/geography-and-climate. [Accessed: 08 July 2020].

Srivastava, PK, Han, D, Ramirez, MR and Islam, T. 2013. Machine learning techniques for downscaling SMOS satellite soil moisture using MODIS land surface temperature for hydrological application. *Water resources management*, *27*(8), pp.3127-3144.

Sungmin, O and Orth, R. 2021. Global soil moisture data derived through machine learning trained with in-situ measurements. *Scientific Data*, *8*(1), pp.1-14.

SU, SL, Singh, D and Baghini, MS. 2014. A critical review of soil moisture measurement. *Measurement* 54, 92-105.

Tian, D, Wood, EF and Yuan, X. 2017. CFSv2-based sub-seasonal precipitation and temperature forecast skill over the contiguous United States. *Hydrology and Earth System Sciences*, *21*(3), pp.1477-1490.

Tippmann, S. 2015. Programming tools: Adventures with R. Nature News, 517 (7532), p.109.

University of Nebraska. 2016. City of York, Nebraska Hazard Mitigation Plan December 2016. City of York, Nebraska.

Uniyal, B, Dietrich, J, Vasilakos, C and Tzoraki, O. 2017. Evaluation of SWAT simulated soil moisture at catchment scale by field measurements and Landsat derived indices. *Agricultural Water Management* 193**,** 55-70.

US Climate Data. 2021. Climate Mead Nebraska. [Internet]. US Climate Data. Available from: https://www.usclimatedata.com/climate/mead/nebraska/united-states/usne0316. [Accessed: 15 March 2021].

Varvara, M, Christos, V, Ourania, T and Kostas, K. 2015. Using Remote Sensing technology with SWAT hydrological modeling to estimate soil moisture of an insular basin. In *Proceedings of the IWA Balkan Young Water Professionals Conference*.

Vather, T. 2015. *Comparison between satellite-based and cosmic ray probe soil moisture estimates: a case study in the Cathedral Peak catchment.*

Vather, T, Everson, C, Mengistu, M and Franz, T. 2018. Cosmic ray neutrons provide an innovative technique for estimating intermediate scale soil moisture. *South African Journal of Science* 114(7-8)**,** 79-87.

Vather, T, Everson, CS and Franz, TE. 2020. The applicability of the cosmic ray neutron sensor to simultaneously monitor soil water content and biomass in an Acacia mearnsii Forest. *Hydrology*, *7*(3), p.48.

Voroney, P. 2019. Soils for horse pasture management. *Horse Pasture Manag*, pp.65-79.

Wahbi, A, Weltin, G, Heng, L, Dercon, G, Vreugdenhil, M, Oismueller, M and Strauss, P. 2015. Cosmic Ray Neutron Probe: Uses, Calibration and Validation in Austria.

Walker, JP, Willgoose, GR and Kalma, JD. 2004. In situ measurement of soil moisture: a comparison of techniques. *Journal of Hydrology* 293(1-4)**,** 85-99.

Wu, Z, Feng, H, He, H, Zhou, J and Zhang, Y. 2021. Evaluation of soil moisture climatology and anomaly components derived from ERA5-land and GLDAS-2.1 in China. *Water Resources Management*, *35*(2), pp.629-643.

Xu, Y, Wang, L, Ma, Z, Li, B, Bartels, R, Liu, C, Zhang, X and Dong, J. 2019. Spatially Explicit Model for Statistical Downscaling of Satellite Passive Microwave Soil Moisture. *IEEE Transactions on Geoscience and Remote Sensing* 58(2)**,** 1182-1191.

Yang, Y, Anderson, MC, Gao, F, Wardlow, B, Hain, CR, Otkin, JA, Alfieri, J, Yang, Y, Sun, L and Dulaney, W. 2018. Field-scale mapping of evaporative stress indicators of crop yield: An application over Mead, NE, USA. *Remote Sensing of Environment*, *210*, pp.387-402.

Yang, Y, Zhang, J, Bao, Z, Ao, T, Wang, G, Wu, H and Wang, J. 2021. Evaluation of multi-source soil moisture datasets over central and eastern agricultural area of China using in situ monitoring network. *Remote Sensing, 13*(6), p.1175.

Zhang, X, Zhang, T, Zhou, P, Shao, Y and Gao, S. 2017. Validation analysis of SMAP and AMSR2 soil moisture products over the United States using ground-based measurements. *Remote sensing* 9(2)**,** 104.

Zreda, M, Desilets, D, Ferré, T and Scott, RL. 2008. Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic-ray neutrons. *Geophysical research letters* 35(21).

Zreda, M, Shuttleworth, WJ, Zeng, X, Zweck, C, Desilets, D, Franz, T and Rosolem, R. 2012. COSMOS: the cosmic-ray soil moisture observing system. *Hydrology & Earth System Sciences* 16(11).

8. APPENDIX

Scatter plot graphs (Figure 8.1) were created to derive the R^2 results, which gives an indication of the strength of the linear relationship between each product and the CRNS. The closer the R^2 value is to 1, the better the relationship is between the datasets. It can be seen that each of the three products have a positive relationship when compared to the CRNS data. The SMOS product had an R^2 value of 0.5825 against the CRNS, which is relatively good. The SMAP_10 km product did not perform as well as the other two products, having an R^2 value of 0.2949. The SMAP_25 km product had the highest R^2 value of 0.6807 against the CRNS data, indicating a good relationship between the two datasets.

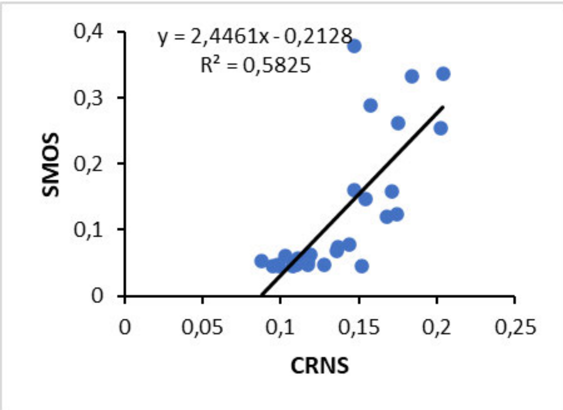

Figure 8.1a SMOS vs CRNS

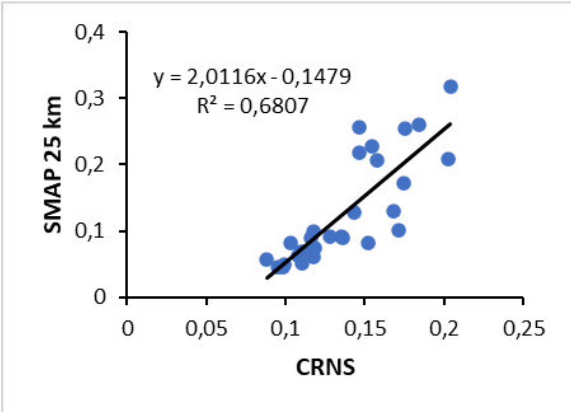

Figure 8.1b SMAP_25 km vs CRNS

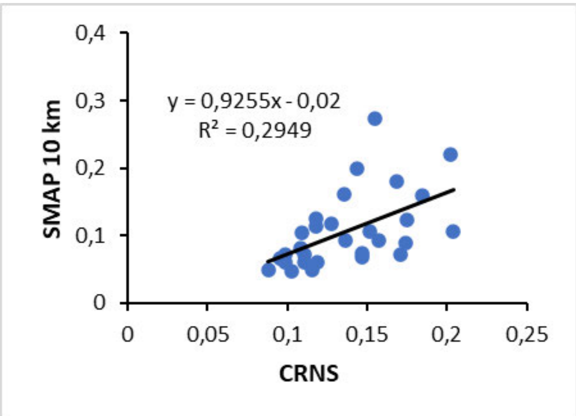

Figure 8.1c SMAP 10 km vs CRNS

Figure 8.1 Scatter plot graphs illustrating the relationship between the satellite products and CRNS at the Two Streams study site

The scatter plots in Figure 8.2 were used to investigate the performance of the reanalysis products against the CRNS data. Both the products had a positive relationship with the CRNS. The ERA5 product had an R^2 value of 0.1848, indicating that it did not correspond well with the CRNS soil moisture estimates. The CFSV2 product had an R^2 value of 0.5554 against the CRNS, which was relatively good.

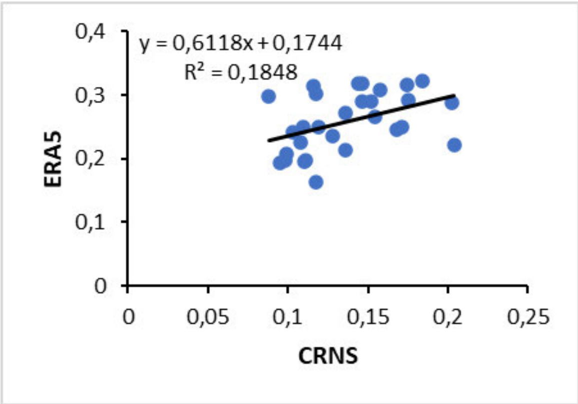

Figure 8.2a ERA5 vs CRNS

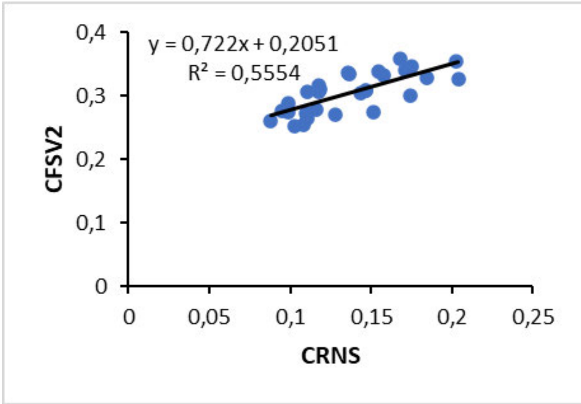

Figure 8.2b CFSV2 vs CRNS

Figure 8.2 Scatter plot graphs illustrating the relationship between reanalysis soil moisture and the CRNS at the Two Streams site

The scatter plots (Figure 8.3) were used to derive the R^2 value. All three satellite products had a positive relationship with the CRNS data. However, the R^2 results were relatively low, which implied that the relationship strength between each of the products and the CRNS were not as strong. The SMOS product had an R^2 value 0.1380 against the CRNS, which implied that the SMOS data did not correspond very well with the CRNS estimates. The SMAP_25 km and SMAP_10 km products had an R^2 value of 0.4383 and 0.3163 respectively, which were also relatively low.

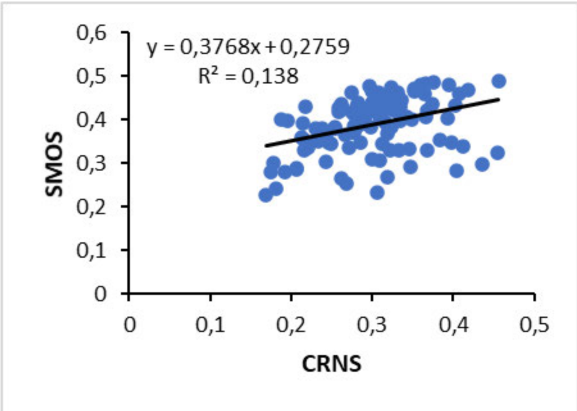

Figure 8.3a SMOS vs CRNS

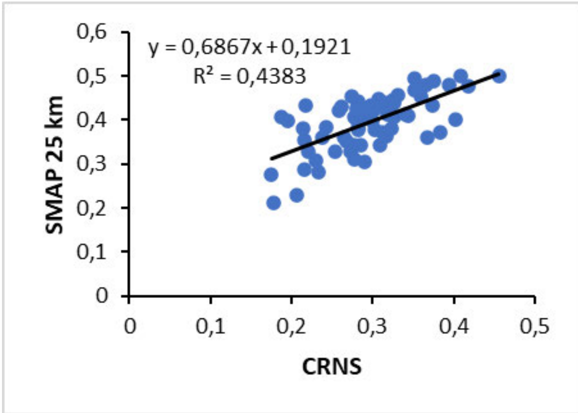

Figure 8.3b SMAP 25 km vs CRNS

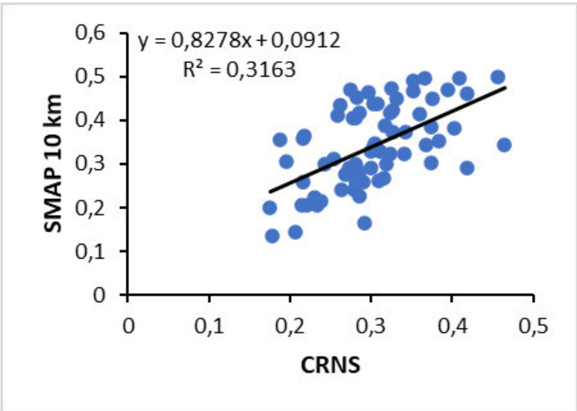

Figure 8.3c SMAP 10 km vs CRNS

Figure 8.3 Scatter plot graphs illustrating the relationship between the satellite products and CRNS at the Mead CSP3 study site

As shown in Figure 8.4, scatter plot graphs were also used to evaluate the reanalysis products at the Mead CSP3 study site. The ERA5 product did not have a strong relationship with the CRNS data, which was indicated by the R^2 value of 0.2202. The CFSV2 product performed relatively well against the CRNS having an R^2 value of 0.5262. Therefore, the data produced by the scatter plots correspond with the trends observed in the time series analysis.

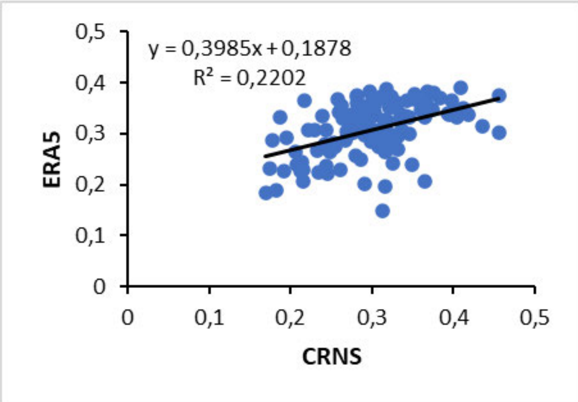

Figure 8.4a ERA5 vs CRNS

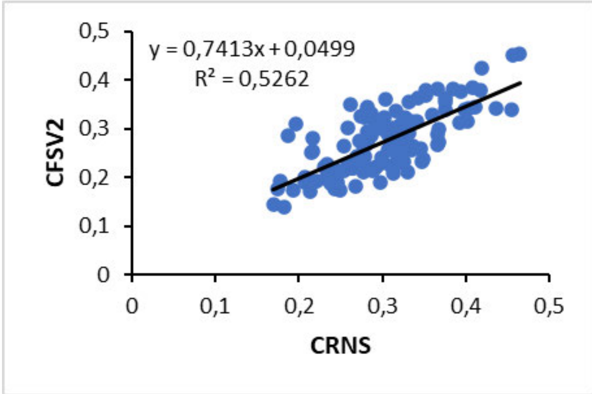

Figure 8.4b CFSV2 vs CRNS

Figure 8.4 Scatter plot graphs illustrating the relationship between reanalysis soil moisture and the CRNS at the Mead CSP3 site

The scatter plots for each of the satellite-based soil moisture products against the CRNS can be seen in Figure 8.5. All three satellite products have a positive relationship when plotted against the CRNS, however, from the R^2 results, it can be deduced that the products do not have a strong relationship with the CRNS. The SMOS and SMAP_10 km products have an R^2 value of 0.3821 and 0.346 respectively, which are relatively low. The SMAP_25 km has an R^2 value of 0.4855, which was the strongest relationship amongst the satellite products for the York Benny study site.

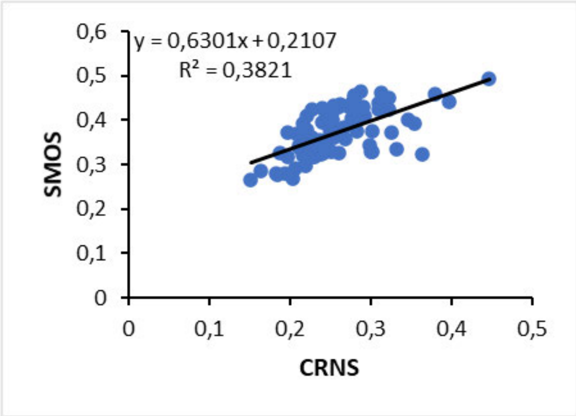

Figure 8.5a SMOS vs CRNS

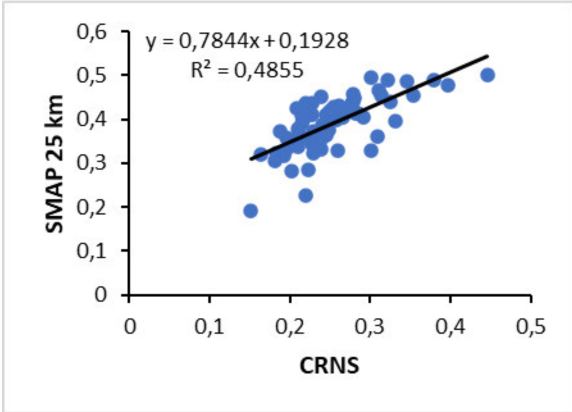

Figure 8.5b SMAP_25 km vs CRNS

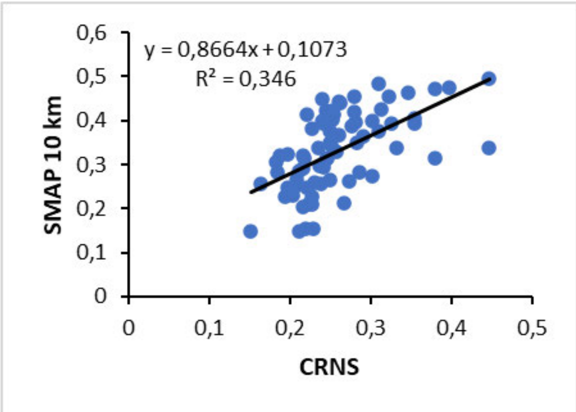

Figure 8.5c SMAP 10 km vs CRNS

Figure 8.5 Scatter plot graphs illustrating the relationship between the satellite products and CRNS at the York Benny study site

The scatter plot data for the reanalysis products against the CRNS data can be seen in Figure 8.6. The ERA5 product had an R^2 value of 0.134, which meant that it did not correspond well with the CRNS data. The CFSV2 product, however, performed better than the ERA5 product with an R^2 value of 0.4376.

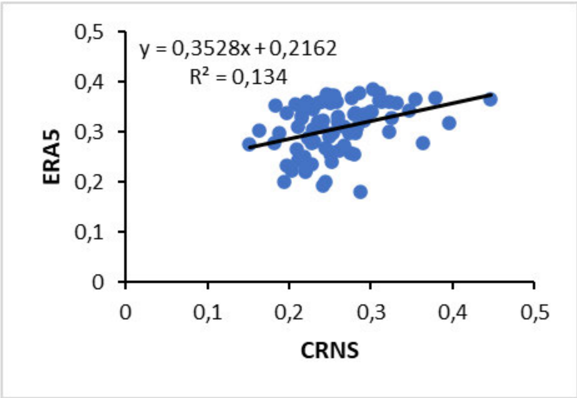

Figure 8.6a ERA5 vs CRNS

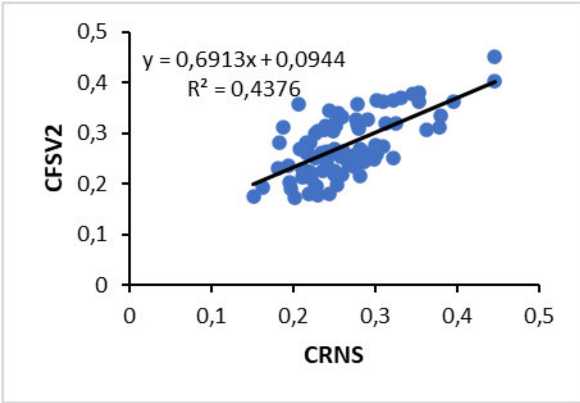

Figure 8.6b CFSV2 vs CRNS

Figure 8.6 Scatter plot graphs illustrating the relationship between reanalysis soil moisture and the CRNS at the York Benny site

The relationship between the reanalysis product, NDVI and LST is illustrated in Figure 8.7. From the correlation results, it is evident that the relationship between the CFSV2 product and the NDVI was moderate.

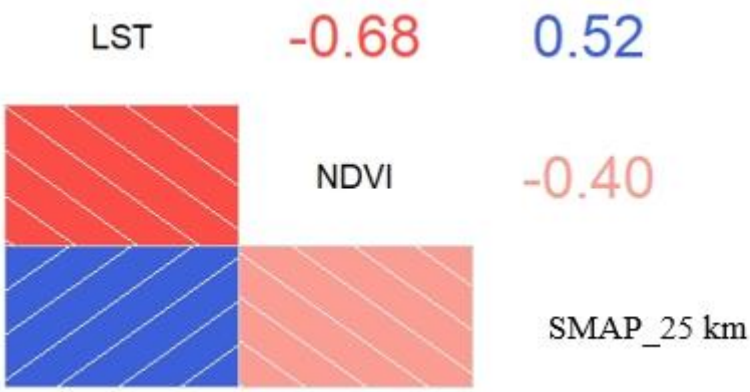

Figure 8.7 Relationship between SMAP_25 km product, NDVI and LST at the Two Streams catchment

The relationship between the reanalysis product, NDVI and LST is illustrated in Figure 8.8. Based on the correlation results that were obtained, it can be deduced that the CFSV2 product was moderately correlated with the fine scale variables.

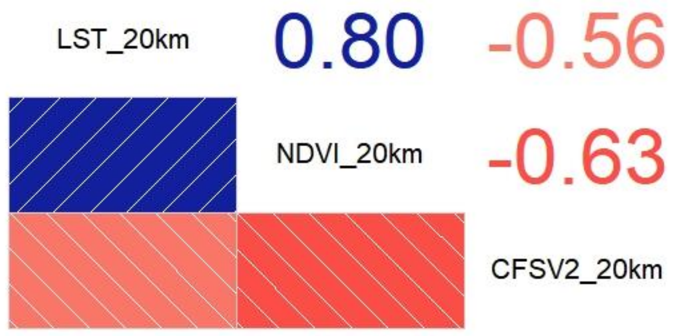

Figure 8.8 Relationship between CFSV2 product, NDVI and LST at the Mead CSP3 catchment

The relationship between the reanalysis product, NDVI and LST at the York Benny site is illustrated in Figure 8.9. Based on the correlation results that were obtained, it can be deduced that the CFSV2 product had satisfactory performance when correlated with NDVI and LST.

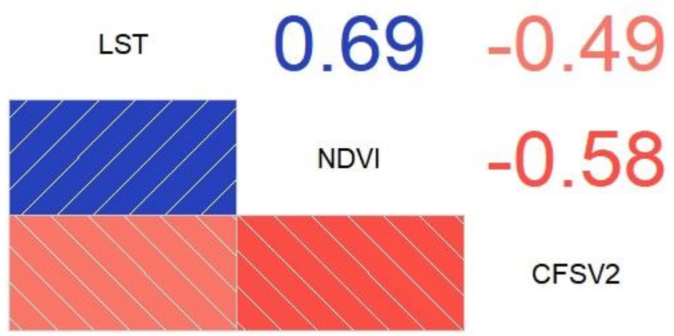

Figure 8.9 Relationship between CFSV2 product, NDVI and LST at the York Benny catchment