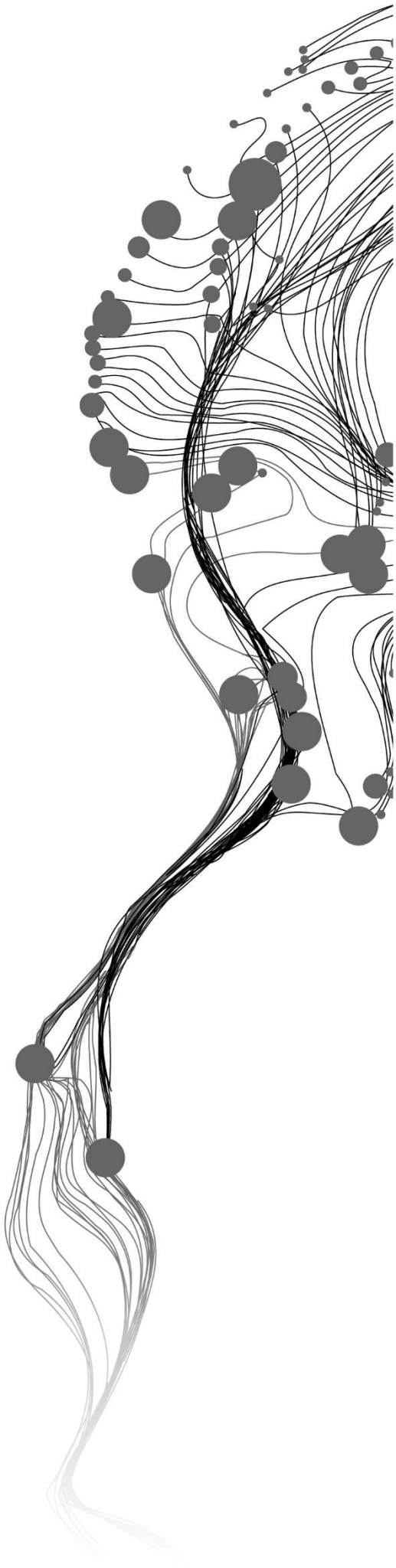
Spatial Downscaling of Satellite Soil Moisture Utilising High-Resolution UAS Data over Alento Catchment in Italy

BYUKUSENGE ELIE
JULY,2020

SUPERVISORS:
Dr. Y. Zeng
Ir. G.N. Parodi



Spatial Downscaling of Satellite Soil Moisture Utilising High-Resolution UAS Data over Alento Catchment in Italy

BYUKUSENGE ELIE
Enschede, The Netherlands, JULY,2020

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Water Resources and Environmental Management

SUPERVISORS:
Dr Y. Zeng
Ir. G.N. Parodi

THESIS ASSESSMENT BOARD:
Chair: Dr. Ir. M.W. Lubczynski
External examiner: Dr Tóth Brigitta (Szabó Brigitta) Institute for Soil Sciences and Agricultural Chemistry, Centre for Agricultural Research, Hungarian Academy of Sciences
Supervisor(s): Dr. Y. Zeng / Ir. G.N. Parodi
Others: Ir. A.M. van Lieshout

ABSTRACT

Soil moisture plays a vital role in water resources management related applications. Nevertheless, the coarse resolution of satellite-based surface soil moisture products has limited applications at field-scale, for example, precision agriculture. The current Sentinel-1 satellite mission provides soil moisture products at 1km resolution, which is still not matching the need at field scales. Therefore, the spatial downscaling approach was applied to downscale coarse resolution (1km) satellite surface soil moisture (SSM) products to high resolution (15 cm) utilising UAS measurements using the random forest (RF) machine learning-based model.

In this study, the RF model was trained using various configurations of input data prepared with remotely sensed SSM and ancillary land surface parameters of LST, DEM and NDVI. The performance of different trained RF models was evaluated to find out which RF model could represent the best relationship of SSM and surface parameters with the best capability for the prediction of SSM. The results indicated that all trained RF models have good performances. However, the trained RF model using 2018 - 2019 dataset on 78 km by 78km spatial extent outperformed the others with the highest correlation coefficient (R) of 0.83 and RMSE of 12.13 %. Therefore, this trained RF model was considered for further process and was applied with the land surface features derived from UAS imageries to predict the SSM at 15cm resolution at noon and sunrise time.

The trained RF model can also identify the relative importance of land surface parameters /features in predicting SSM. It was found that the LST has a higher impact than other features, while DEM being the least influential. The downscaled SSM can capture the spatial pattern of SSM at noon and sunrise time, when compared with the in situ measurements from the study area in Monte Cilento Sub catchment in Alento Catchment, Italy. The averaged ubRMSE, RMSE and R are reported 0.07 cm^3/cm^3, 0.21 cm^3/cm^3, and 0.60 respectively. Notably, all statistical metrics showed acceptable results even though the average of ubRMSE does not reach the SMAP and Global Climate Observing System (GCOS) mission accuracy target of 0.04 cm^3/cm^3 for soil moisture due to the downscaled SSM products were generated at 5 cm while the in situ measurements were taken at 15 cm.

In summary, this study successfully generates high spatial resolution SSM data from coarse-scale satellite products by integrating UAS measurements and RF model as a downscaling approach. The generated soil moisture products could provide useful information for better agricultural management in Monteforte Cilento sub-catchment in Alento River catchment.

Keywords: Surface soil moisture; land surface parameters; spatial downscaling; Random forest; UAS measurements

ACKNOWLEDGEMENTS

I want to thank the Dutch Government for giving me the fellowship under the OKP program to study at the University of Twente, ITC Faculty. I would want to thank all staffs of Faculty of Geo-information Science and Earth Observation for providing practical and theoretical knowledge. It was an excellent opportunity to be in a multicultural environment, which allowed me to improve my professional skills.

I want to express sincere appreciation to the supervisors, Dr Y. Zeng and Ir. G.N. Parodi for their professional guidance, suggestions and valuable support. I greatly appreciate Dr Zeng for believing in me and providing the needed support. I want to thank also my mentor, Ir. A.M. van Lieshout for his advice and motivation during my studies.

My sincere thanks also go to the researchers Ruodan Zhuang, Salvatore Manfreda and Antonino Maltese for their contributions on UAS images collection for this research. I would also like to acknowledge Nunzio Romano and Paolo Nasta who involved for in-situ data collection in the study area. Also, thanks to my colleague Lijie Zhang for his help on the RF python code, Furthermore, Ruodan, who also guided me in data processing during my MSc research.

Special thanks to my friends Frank Mugabe, Jean de la Croix Ishimwe, Olivier Nzamuye, Habimana Claude, for supporting and made me feel more confident in everyday life. It was amazing two years together, and I will never forget the memories we had together in Enschede.
Lastly, I would like to give my special gratitude to my family for their encouragement and moral support throughout my studies and continuous encouragement till now.

Elie BYUKUSENGE

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ASCAT	Advanced SCATterometer
CGLS	Copernicus Global Land Services
DEM	Digital Elevation Model
DN	Digital numbers
ESA	European Space Agency
GCOS	Global Climate Observing System
LAI	Leaf Area Index
LST	Land Surface Temperature
MODIS	MODerate resolution Imaging Spectro-radiometer
MSE	Mean Square Error
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NIR	Near InfraRed
R	Correlation Coefficient
RF	Random Forest
RMSE	Root Mean Square Error
S-1	Sentinel 1 A and B
SAR	Synthetic Aperture Radar
SMAP	Soil Moisture Active Passive
SMOS	Soil Moisture and Ocean Salinity satellite
SSM	Surface Soil moisture
TDR	Time Domain Reflectometry
TIR	Thermal InfraRed TRMM
TU-Wien	Technische Universität Wien
UAS	Unmanned Aerial system
UAV	Unmanned Aerial vehicle
ubRMSE	unbiased RMSE

1. INTRODUCTION

1.1. Background

Soil moisture is an important variable which plays a crucial role in land-atmosphere interactions (Ochsner et al., 2013). It is defined as the amount of water in the unsaturated zone (Hillel et al., 1998) and can also be considered as surface soil moisture on the top few centimetres soil (Bauer-Marschallinger & Schaufler, 2018). The surface soil moisture (SSM) has potential value for water management studies and applications (Zhao et al., 2018a). It is necessary to get accurate information on surface soil moisture (SSM) for understanding the land-atmosphere processes.

The soil moisture information can be obtained from different sources such as in-situ measurements and remote sensing observations. The most commonly applied in-situ measurement techniques are the time domain reflectometry (TDR) and gravimetric method (Luca Brocca et al., 2017). However, they are not suitable to represent the spatial variability of soil moisture due to the heterogeneity nature of soil properties(Sabaghy et al., 2018a). On the other hand, the remote sensing observations can provide the spatiotemporal information of soil moisture on a global scale.

Nowadays, the soil moisture products can be derived from multiple satellites which use different sensors such as the microwaves and optical or thermal infrared sensors. For example, Sentinel-1 C SAR products have 1km spatial resolution SSM products (Bauer-Marschallinger & Schaufler, 2018) which is much higher resolution than that of SMOS, SMAP products and other existing satellite-based soil moisture products. Nevertheless, 1km spatial resolution is still not matching the needs at field scales (e.g. submeters for precision agriculture) (Sabaghy et al., 2018a). There are still challenges to achieving high-resolution soil moisture data at field scale (Rötzer et al., 2015) and the downscaling methods have been developed to obtain needed spatial resolution (Zappa et al., 2019).

The spatial downscaling has been used to get high-resolution soil moisture data by combining different land surface parameters and coarse satellite soil moisture (Kim et al., 2018). Sabaghy et al. (2018) found that it is also efficient to combine satellite soil moisture products with high-resolution land surface parameters derived from optical observations and radiometer observations. Peng et al. (2017) stated three main types of spatial downscaling methods such as satellite-based method, geoinformation data-based method and model-based method. The satellite-based techniques mostly use the high spatial resolution optical observations and radiometer observations at higher frequencies combined with coarse resolution soil moisture products to generate downscaled soil moisture data (Sabaghy et al., 2018a).

On the other hand, model predictions have also been used in the model and data-based downscaling techniques to provide high spatial resolution soil moisture data (Peng et al., 2017a). Common methods are data assimilation and machine learning. The data assimilation uses a model by considering the prior information about the data and predict the outputs based on the initial values. At the same time, the machine

learning approach takes into consideration the relationship between the soil moisture as the response variable and land surface features/parameters as predictor variables (Sabaghy et al., 2018a). Specifically, the machine learning model can be applied to the high-resolution land surface features/parameters by assuming that considered relationship remains the same among different scales (Zappa et al., 2019).

Among the abovementioned existing spatial downscaling methods, Sabaghy et al. (2018) reported that the existing techniques could downscale the coarse resolution soil moisture at required accuracy of RMSE which is 0.04 cm^3/cm^3 in the top 5 cm suggested by the SMAP science team (Entekhabi et al. .2010) and Global Climate Observing System (GCOS) (Bauer-Marschallinger & Schaufler, 2018). The machine learning methods are better than the current available downscaling methods, but radar-based techniques also provide good accuracy (Sabaghy et al., 2018a). Previous studies showed that the Random Forest (RF) algorithm approach is more suitable for downscaling the satellite products than other Machine learning algorithms(Abbaszadeh et al., 2019).

This study aims to use (UAS) measurements to downscale the coarse surface soil moisture from one kilometre to centimetres spatial resolution by using the Random Forest regression method. The main idea is to establish a relationship between SSM and other land surface parameters. Then, apply the trained RF-model with the high-resolution land surface parameters to obtain high-resolution SSM data. The input datasets of Sentinel -1 1km SSM products and ancillary land surface parameters (LST, NDVI and DEM) were collected from 2015 to 2019 to train the RF model. The unmanned aerial system (UAS) measurements had been used for estimating the land surface parameters that are linked to soil moisture (Manfreda et al., 2018). The UAS uses different types of sensors like thermal and hyperspectral sensors and gives high-resolution images at low operational cost (Hsu & Chang, 2019). The downscaled results using UAS and RF model were validated via comparing with the in situ soil moisture measurements collected in the study area.

1.2. Research problem

Water resource management uses soil moisture information of different spatial scales for solving the related problems. In 2018 there was a drought event in Europe, and soil moisture deficiency had affected different sectors. Among others, the agricultural sector was the most affected over Alento catchment in Italy (EU Science HUB, 2019). With enough information about soil moisture products, it is expected that one can predict or propose the adaptation or mitigation measures to avoid the significant loss for farmers.

Many satellite missions have been developed to monitor and estimate the soil moisture products on a global scale. The current relatively high spatial resolution (1km) surface soil moisture (SSM) was still not matching the need for field-scale applications (e.g., submeters for precision agriculture). The submeter resolution soil moisture products could provide needed information for the agricultural management in Monteforte Cilento sub-catchment in Alento River catchment. Nevertheless, it is still challenging to obtain the desired high spatial resolution soil moisture data due to the spatial and temporal variability of soil moisture patterns.

1.3. Research objectives

The main objective of this study was to downscale the coarse resolution (1 km) surface soil moisture to high resolution (15 cm) utilising Sentinel-1 C SAR SSM products combined with high-resolution land surface parameters derived from UAS' measurements.

Specific objectives are :

- To analyse the performance of the proposed Random forest downscaling approach.
- To downscale surface soil moisture products from coarse resolution to high resolution using Random forest approach
- To validate the downscaled soil moisture products with ground soil moisture measurements.

1.4. Research questions

- What is the performance of the proposed downscaling approach?
- How do the land surface parameters and used method affect the accuracy of downscaled surface soil moisture products?
- What is the accuracy of downscaled SSM products when validating with ground measurements

2. LITERATURE REVIEW

The spatial downscaling method is a procedure to get high-resolution images from coarse resolution images (Sabaghy et al., 2018a). It is mainly based on the relationship between soil moisture and various surface parameters. It is essential to select an appropriate downscaling method and surface parameters to downscale coarse resolution soil moisture data. The next sections discussed more on soil moisture product, land surface parameters which have a more significant impact on soil moisture variability, on how to use the UAS measurement to get the land surface parameters, and lastly on the downscaling approaches and the detail of the proposed Random forest approach for this study.

2.1. Sentinel -1 surface soil moisture product

The Sentinel -1 surface soil moisture product is version 1 of the products provided by the European Space Agency mission, and it is obtained from the observations from a constellation of two satellites as Sentinel-1A and Sentinel-1B (Bauer-Marschallinger & Schaufler, 2018). The Synthetic Aperture Radar (SAR) instrument is the main instrument carried by the Sentinel-1 spacecraft, and specific information can be obtained from SAR images (Poullaouec et al., 2016).

The Sentinel-1 products have various applications among them water management and soil protection (Poullaouec et al., 2016). TU Wien change detection method (Wagner et al., 2013) was used to retrieve surface soil moisture products with 1km of spatial resolution from backscatter measurements.

Bauer-Marschallinger and Schaufler (2018) defined the surface soil moisture (SSM) as the relative water content of the top few centimetres soil. The information on the dynamics of soil moisture is essential to understand the processes in many environmental fields (for example, on the impact on agricultural productivity) (Bauer-Marschallinger & Schaufler, 2018).

2.2. Land surface parameters

It is crucial to establish the relationship between surface soil moisture (SSM) and the surface parameters in the downscaling scheme (Song et al., 2014). Previous downscaling studies have demonstrated common land surface parameters such as land surface temperature (LST), normalized difference vegetation index (NDVI) and digital elevation model (DEM) in expressing their relationship with surface soil moisture (SSM) (Song et al., 2014; Peng et al., 2016; Peng et al., 2017; Sabaghy et al., 2018a; Sabaghy et al., 2020).

In this study, the next section summarizes the role of those surface parameters for influencing the soil moisture variability.

The normalized difference vegetation index (NDVI) is considered as a ratio between the visible (red) and near-infrared bands, and it is a standard index used to determine the vegetation characteristics (Bhandari et al., 2012).

It is included to account the influence of vegetation dynamics on soil moisture variation at different space and time scales (Abbaszadeh et al., 2019). It also shows the structure of the vegetation and is related to other properties such as the leaf area index (LAI) which was defined by Borzuchowski and Schulz (2010) as the total area of leaves per unit ground area. Hawley (1983) demonstrated that the presence of vegetation tends to diminish the magnitude of soil moisture variations.

The land surface temperature (LST) represents the temperature measured at the Earth's surface, and it is a crucial parameter which is used to detect the dynamics of land-surface processes (Zhengming, 1999; Dash et al., 2002).) The soil moisture and land surface temperature are closely related as soil emission are influenced by soil moisture and soil physical temperature (Pablos et al., 2016). The soil moisture affects the magnitude of surface temperature via its influence on emissivity(Sun & Pinker, 2004). The land surface temperature is used to maintain the temporal dynamics of the soil moisture (Abbaszadeh et al., 2019).

Balasubramanian (2017) defined the Digital Elevation Model (DEM) as a digital representation of topography. It gives topographic parameters such as elevation and slope. It affects the distribution of soil moisture, particularly in the topsoil layer and the elevation was known as one of the most useful topographic features in many studies to downscale coarse-scale soil moisture (Crow et al., 2012).

In general, it is difficult to conclude the impact of land surface parameters on surface soil moisture variability. Sometimes, one or more of these contributing parameters can be neglected due to the research objectives over a given study area. This research considered land surface parameters mentioned above as most factors of surface soil moisture variability and also based on the availability of UAV data in the study area.

2.3. Unmanned aerial system measurement

Manfreda et al. (2018) reported on the use of Unmanned Aerial Systems (UAS) technology for environmental monitoring and highlighted the potential of UAS imagery to provide information that is useful in decision making and proper monitoring of hydrological processes. The next sections discuss the description of UAS as well as data collection and processing of UAS measurements.

2.3.1. System description

The unmanned aerial system (UAS) is defined as a system which considers various components such as the unmanned aerial vehicle (UAV) known as a drone, a ground-based controller, and a method of communications between those components (Jeziorska, 2019). There are three main categories of the unmanned aerial vehicle (UAV) such as fixed-wing, rotor and kite. The UAVs have an autopilot that controls the flight, Global Navigation Satellite System (GNSS) receiver, camera and an inertial platform (IMU), those instruments are communicating with ground-based controller to provide the position of the images during the acquisition. Figure 2.1 shows the basic components of an Unmanned Aircraft System.

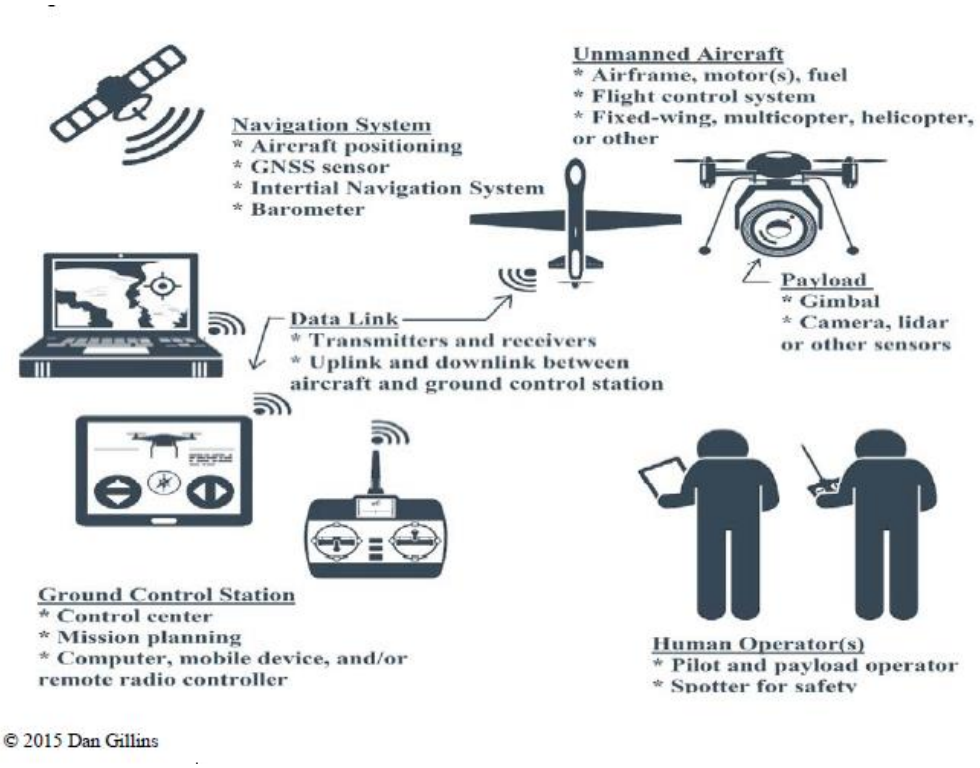


Figure 2.1: Basic Components of an Unmanned Aerial System (Gillins et al., 2018)

2.3.2. Sensing Payloads

There are many types of sensors which can be installed on the unmanned aerial vehicle (UAV). In this study, the focus was on hyperspectral and thermal sensors. Those sensors can be used to provide field measurements which can be used to derive the land surfaces parameters related to soil moisture sensitivity and dynamics (Jeziorska, 2019).

The thermal data was collected using the FLIR Tau thermal camera, which provided radiometric values that later converted to land surface temperature values in degrees Celsius (Sagan et al., 2019). The land surface temperature can be used to compute other surface parameters which have a relationship between them, such as surface soil moisture or vegetation condition (Manfreda et al., 2018).

On the other hand, the Cubert hyperspectral camera was used to provide the biophysical measurements of vegetation (e.g., NDVI) and topographic information (DEM) in the study area (Aasen et al., 2015).

2.3.3. Data collection and processing

Acquisition starts with the flight planning as it is essential to specify the drone characteristics, including the camera/sensor type, and setting the extent of the study area. The weather conditions and ground sampling distance may also be taken into consideration to get good images (Manfreda et al., 2018).

The UAV acquires the images at high overlap by using the grid flight plan. The ground control points are needed for accurate georeferencing and can be collected using a GNSS receiver in the study area. All those aspects affect data quality characteristics and further images processing. It is also essential to consider local flight regulations.

The data processing can be done using the dedicated software (eg.Pix4D, Agisoft Photoscan) to generate the orthomosaic map depending on types of the camera. It is to note that the UAV images have been acquired for this study, the data processing as explained in the above has been implemented to generate orthomosaic maps, which was further used to retrieve land surface parameters (i.e., LST, NDVI and DEM) at 15 cm of spatial resolution.

2.4. Spatial downscaling methods

There are various existing downscaling techniques, and most of them differ based on the inputs data and downscaling approach (Sabaghy et al., 2018a). In general, the spatial downscaling method uses coarse resolution soil moisture observations and additional information about the land surface parameters. The existing methods were discussed in recent studies by Peng et al. (2017) and Sabaghy et al. (2018). The next section highlighted the difference between the existing methods and the detail of the proposed downscaling method for this study.

2.4.1. Satellite-based method

The satellite-based downscaling method is mainly based on the observations derived from radar and radiometer. The main concept was first introduced by Njoku et al. (2000). For example, O'Neill, Chauhan and Jackson(1996) downscaled soil moisture using a radiative transfer model with radar-derived data on the vegetation attenuation, transmissivity and scattering parameters. Another example is the change detection method which used the linear correlation between backscatter and soil moisture content for formulating downscaling models which derived relative changes of soil moisture with higher resolution (Piles et al., 2009).

The other satellite-based method is based on optical observations. It mainly combines the strengths of high spatial resolution optical observations and high accuracy passive microwave to derive soil moisture. Therefore, the high spatial resolution of optical observations provides information on surface parameters to improve the resolution of soil moisture products (Sabaghy et al., 2018a). For example, Zhao et al. (2018) downscaled SMAP passive soil moisture from 9km to 1km of spatial resolution using the high resolutions observations from MODIS Terra/Aqua satellites.

2.4.2. Model-based method

The model predictions have been used to downscale the coarse resolution soil moisture mainly based on data assimilation and machine learning algorithm (Sabaghy et al., 2018a). Notably, the machine learning approach establishes the relationship between soil moisture and other the environmental parameters then use the constructed relationship to predict the high-resolution soil moisture products using high-resolution predictors (Chen et al., 2019). Existing machine learning methods are Artificial Neural Network (ANN), Support Vector Machine (SVM), random forest and Relevance Vector Machine (RVM) (Sabaghy et al.,

2018a). One of them, the random forest model, was developed by Breiman and was proposed as satellite soil moisture downscaling approach in this study.

2.4.2.1. The main concept of the random forest approach

Breiman (2001) described the random forest as a supervised machine learning algorithm which can be applied for regression and classification tasks. The main focus is on the random forest regression task which builds several decision trees during the training time, and the mean prediction of the individual trees is generated as the output of this method (Zhao et al., 2018a).

The variables required for random forest regression consists of predictors and the dependent variable (Bartkowiak et al., 2019). In this study, the principle of the proposed downscaling approach is to build a function (F) that understand the relationship between soil moisture and land surface parameters at original spatial resolution. Then, the constructed function can be used to predict soil moisture at high spatial resolution (Abbaszadeh et al., 2019) by using the following equations:

- $\mathbf{SSM_o = F(P_o)}$ **(2.1)**
- $\mathbf{SSM_d = F(P_d)}$ **(2.2)**
- **P = (p1, p2, p3 ,…)(2.3)**
    - $\mathbf{SSM_o}$ is the original soil moisture data
    - $\mathbf{SSM_d}$ is downscaled soil moisture data,
    - **P** is the land surface parameters (i.e., land surface temperature (LST), DEM and NDVI).

Therefore, there are predictors and dependent parameters for RF algorithm. Sentinel -1 1km SSM soil moisture product was considered as the dependent parameter and NDVI, land surface temperature (LST) and DEM as the predictor's parameters.

2.4.2.2. Main steps for training Random forest

The following steps show the workflow of random forest regression algorithm as described by Bartkowiak et al. (2019).

Step 1: Data preparation

The downloaded data cannot be used into the model; therefore, data preparation is needed to put the data into machine learning understandable terms. Data preparation requires steps such as :

- The missing data and invalid values can be identified using the statistical summary. It is needed as both can impact the analysis. The missing data values can be changed to non-values while the invalid values can be removed from the data.

- After removing the invalid values and missing data, the other process is the separation of the data into the features and targets. The targets known as the labels and are the values to predict, and the features also are values which used to make a prediction.

Step 2: Training the RF model

- After data preparation and getting the features and labels values, the next step is splitting data into training and testing sets. The training number of features and labels have to match the testing number of features and labels. Usually, the ratio between the training number and the testing number is 3:1
- The next step is to train the RF-based regression algorithm to learn the relationship between features and labels during the training phase. The tree sample sets are randomly derived from the original training data set with replacement using the bootstrap method (Bartkowiak et al., 2019). Each sample set is a bootstrap sample, and the elements that are not included in the bootstrap are the out-of-bag data (OOB) for that bootstrap sample (Jing et al., 2016).
- In this study, about 75 % of the observations (referred to as in-bag samples) are used to train RF algorithm while 25 % (referred to as out-of-the bag samples) are used in an internal cross-validation technique for checking the performance of trained RF model (Breiman, 2001).

Step 3: Make predictions with the test data & model performance

The RF algorithm has been trained to learn the relationships between the features and the labels. There is a trained RF-based regression model which is representing the relationship between labels as Sentinel -1 1km SSM soil moisture products and surface features as NDVI, LST and DEM.

The trained RF model can be applied to the test features, and the outputs are the predicted values of the surface soil moisture. The prediction results are obtained by averaging the predictions from each regression trees (Jing et al., 2016).

After that, the model performance can be evaluated using the statistical metrics obtained from comparing the predicted soil moisture values with observed satellite soil moisture values. Consequently, if performance is not acceptable, the adjustment of the model inputs datasets may be made by acquiring more data or changing the model settings.

Step 4: Variable importance

The ranking of variable importance is essential in the random forest algorithm. Random forest algorithm could indicate the relative importance of a variable as an increased mean square error (MSE) (Jing et al., 2016). An increased MSE can be calculated by randomly assigning a variable to compute the extent of the reduction in the accuracy of the random forest prediction, thus during the fitting process, the prediction error for each out-of-bag (OOB) sample was recorded and averaged over the forest (Cutler et al., 2012).

The variable which has a larger value than others indicates that it has high importance in terms of its contributions to the RF regression model (Qu et al., 2019).

Step 5: Model prediction on the UAS measurements

The final step of the workflow in this study is the prediction of the high-resolution soil moisture products. On this stage, the model has been well-trained, and it has excellent performance representing the relationship model between SSM and the other surface parameters. Next step is to apply the trained RF model to the surface test features in order to obtain the surface soil moisture at high-resolution.

2.4.2.3. Main advantages of Random forest model over other existing models

- The previous studies have shown that RF regression algorithm in comparison with other machine learning approaches is more suitable for downscaling the satellite products (Sabaghy et al., 2018a). For example, Tyralis et al. (2019) presented that the random forest method as the most flexible algorithm because a model can select, combine and fit different functional relationship between predictors and dependent parameters.
- It gives useful outputs results, for example, statistical metrics to evaluate the impacts of inputs parameters on the model performance and the relative variable importance scores to rank impacts of the used land surface parameters/features on the trained model.
- Random forest algorithm is considered as a highly accurate algorithm because to get the results; it builds multiple decision trees, and the results are the averaging of the results of the predictions from each regression trees (Jing et al., 2016).

2.5. Importance of high-resolution soil moisture measurement

The development of the field and small watershed scale (0.1–1 km^2) soil moisture measurements help into agricultural production and a better understanding of hydrological processes responses into the catchments (Robinson et al., 2008). The high-resolution soil moisture product can be used in supporting the agricultural management for preparing and adapting to future extreme events under the backdrop of climate change.
It could also help considerably in the different scientific research, for example, in minimising the uncertainty in the developed model.
Figure 2.2 summarises the temporal and spatial resolution requirements in a given water-related application. Downscaled soil moisture with sub-meters of spatial resolutions is required in precision agriculture to monitor the agricultural field variations and to deal with them using alternative strategies (Zhang and Kovacs 2012). Precision agriculture mainly uses high-resolution satellite images to understand those variations in the field, like soil moisture conditions (Sabaghy et al., 2018a). It was also found out by Torbett et al. (2007) that when the farmers know the spatial variation in soil properties, they adopt precision agriculture.

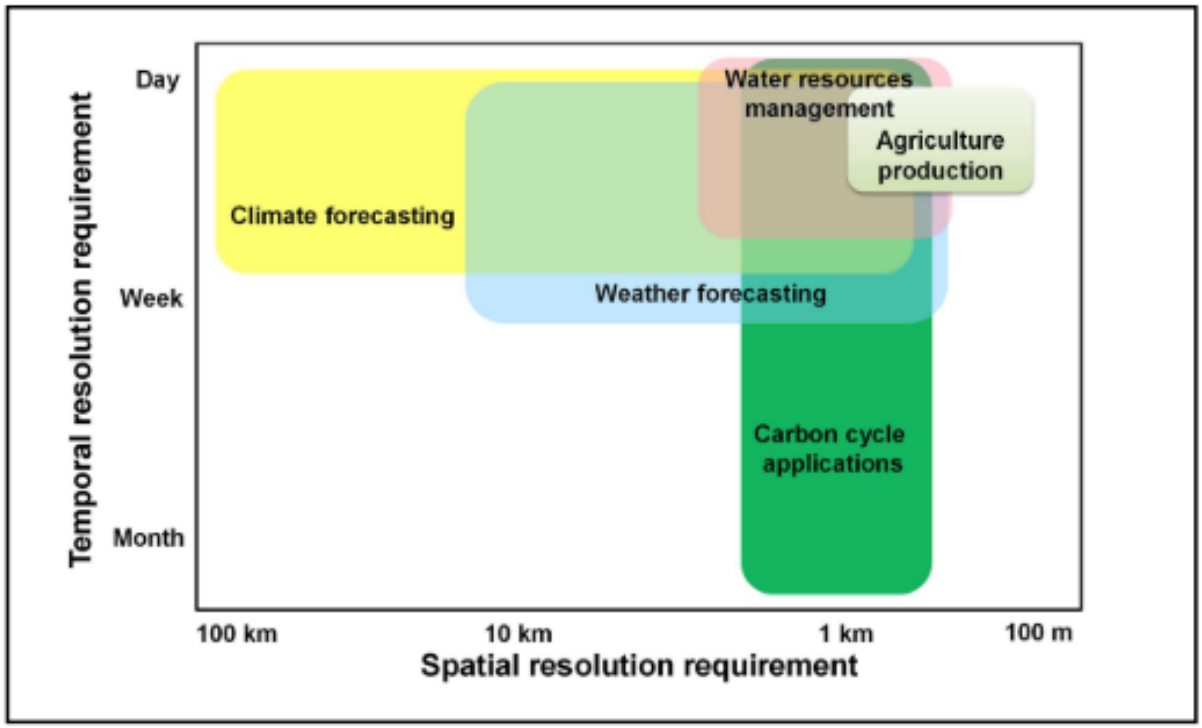


Figure 2.2: Summary of spatial and temporal resolution requirement of soil moisture for a range of applications (Sabaghy et al., 2018a)

3. STUDY AREA AND DATASETS

3.1. Alento River catchment description

It is located in the province of Salerno, Campania in Italy. It has a total area of 450 km^2 and an average elevation of 400m above sea level and is divided into sub-catchments (Romano et al., 2018). One of the sub-catchments is located near the village of Monteforte Cilento, the second sub-catchment is situated near the first one, and another sub-catchment is located near to the rural village of Gorga and characterised by forest (Nasta et al., 2013).

The main focus was on the Monteforte Cilento catchment, where UAV data acquisition had been done, and it has approximately 8.5 hectares. The geographical coordinates of the area are between 40^{0}15' and 40^{0}30' N and 15^{0}07'and 15^{0}22' E. It has the sub-humid Mediterranean mountains climate. The rainfall pattern is characterised by the cold season, with about 67 % of annual rainfall from October to April. From May to September is the hot season, whereas July is the hot month (M. C. Peel, B. L. Finlayson, 2002).

The wireless soil moisture network provides the soil moisture measurement into different depth (Romano et al., 2018), and Figure 3.1 represents the spatial distribution of the installed SoilNet sensors and the topography of the catchment. Monteforte Cilento catchment is mainly characterised by the agricultural activities and divided into a piece of lands which can be influenced by the hydrological processes in the catchment (Nasta et al., 2013).

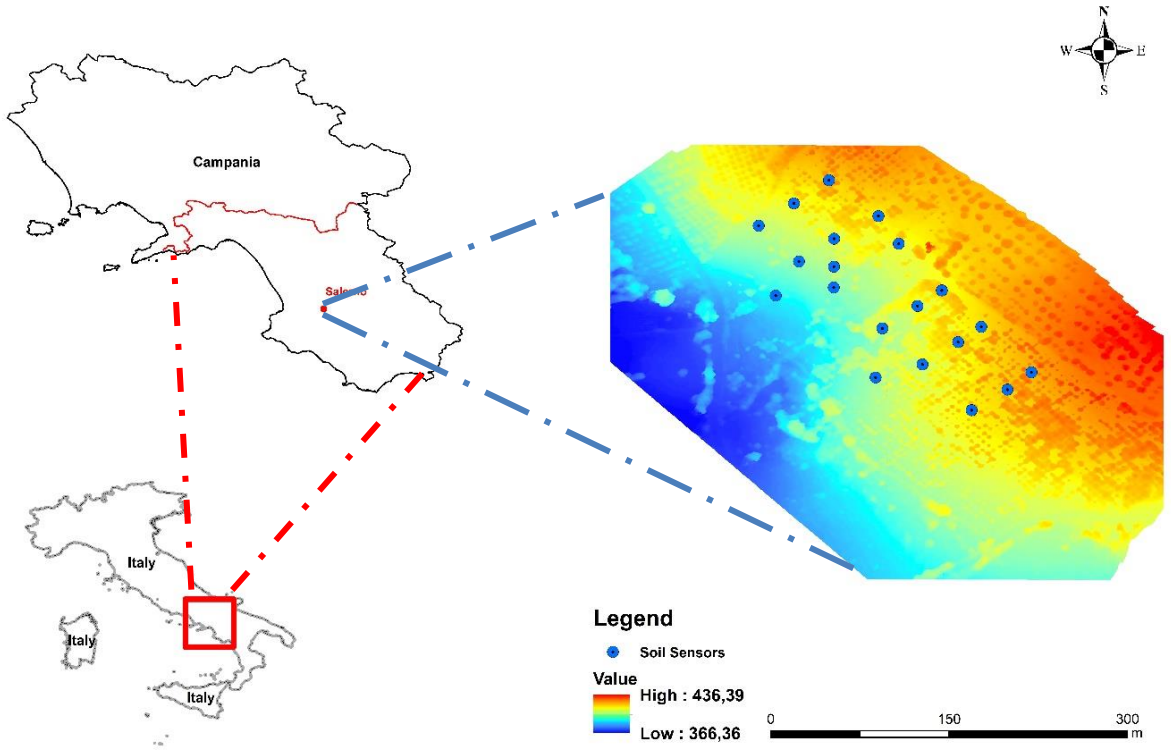


Figure 3.1: The topography of the study area and spatial distributions of the SoilNet soil sensors.

3.2. Data collection and description

The model inputs data were collected based on two spatial extents as 28km x 28km ($40^0$15'- $40^0$30' N and $15^0$07' - $15^0$22' E) and 78km x 78km ($40^0$26' - $40^0$51' N and $15^0$11' - $15^0$36' E). The study area lies inside the two spatial extents. The data of Sentinel -1 1 km SSM, NDVI, LST and DEM were acquired from 01 January 2015 till 31 December 2019 from various datasets. The UAV images were obtained using the Cubert hyperspectral, and FLIR Tau 2 thermal camera. The processed orthomosaic maps of land surface parameters (NDVI, LST, DEM) were available for 13 and 14 June 2019 at Noon and Sunrise and were provided as secondary data from other researchers. The next sections discuss the data description and source.

3.2.1. Sentinel-1 C-SAR Surface soil moisture (SSM) product

Sentinel-1 C-SAR surface soil moisture product is one of the European Space Agency developed mission. It is produced by Vienna University of Technology (TU Wien). It is a constellation of two satellites as Sentinel-1A and Sentinel-1B. It has the temporal resolution depending on the locations. For example, in the European continent, it has a temporal resolution of 3-8 days (January 2015 to October 2016) and 1.5-4 days from October 2016 (Bauer-Marschallinger & Schaufler, 2018).

It was derived at the top five centimetres of depth. It is expressed in % of saturation with range values between 0% (dry the soil) and 100% (wet soil). It can also be further converted to volumetric soil moisture by multiplying the porosity values (Bauer-Marschallinger & Schaufler, 2018). Level 2 Sentinel-1 C-SAR SSM products can be freely downloaded on https://land.copernicus.eu/global/ with 1-km of spatial resolution. The output product is a daily image which has the format of a netCDF4 file.

3.2.2. MODIS land products (LST and NDVI)

The land surface temperature (LST) and normalized difference vegetation index (NDVI) were downloaded from the NASA Land Processes Distributed Active Archive Center (LP DAAC)(Justice et al., 2002). All those land surface products have a spatial resolution of 1 km.

The daily land surface temperature product from Terra MODIS(MOD11A1) can be freely obtained from https://lpdaac.usgs.gov/products/mod11a1v006/. It is daily level 3 products (daytime and night time), but the main focus was on daytime LST as the previous downscaling study by Peng et al. (2015) showed that the LST has a high sensitivity to SSM due to its considerable variation in the daytime. Zhao et al. (2018) also found that the downscaling results using MODIS LST daytime had better performances than night time LST.

The normalized difference vegetation index from MODIS Terra (MOD13A2) can also be freely obtained from (https://lpdaac.usgs.gov/products/mod13a2v006/. Eventhough the MODIS images are collected daily, the NDVI products have the temporal resolution of 16 days (Didan et al., 2015). The files format of MODIS data is HDF-EOS.

3.2.3. GTOPO30 Product (Digital elevation model)

GTOPO30 topography data was used to give information about the elevation, and it is a global Digital Elevation Model developed by USGS Earth Resources Observation and Science Data Center in late 1996 (Abbaszadeh et al., 2019). This dataset has 1km of spatial resolution and can be freely accessed via https://earthexplorer.usgs.gov/. The product is distributed into tiles which one tile covers 50 degrees of latitude to 60 degrees of longitude.

3.2.4. Insitu measurement

The SoilNet wireless sensors provide the insitu soil moisture measurement data at different depths such as 15 cm and 30 cm. Sensors are installed in 20 different locations covering the study area. The GS3 sensor (Decagon Devices) wireless sensors measure soil dielectric permittivity which could be converted to volumetric soil water content (q, cm3 cm−3) using a calibration equation specific to the location of the study area (Romano et al., 2018). Figure 3.1 shows the study area map with spatial distributions of the soil sensors. The table below summarizes the descriptions on the in situ measurements collected in 20 locations for one single day (14 June 2019).

Soil sensor	East (m)	North (m)	Z (m)	e permittivity	q ($cm^3\ cm^{-3}$)	porosity
1	515684.92	4468257.90	513.44	18.00	0.22	0.52
2	515652.62	4468295.98	509.19	10.45	0.23	0.50
3	515627.21	4468326.45	505.29	8.46	0.31	0.54
4	515599.49	4468365.61	504.94	12.64	0.27	0.50
5	515586.37	4468388.73	507.32	19.40	0.29	0.61
6	515554.63	4468418.78	509.66	16.98	0.44	0.64
7	515669.53	4468243.32	510.78	9.93	0.30	0.57
8	515637.77	4468283.15	504.35	7.44	0.23	0.61
9	515611.62	4468313.42	501.67	8.70	0.29	0.53
10	515557.95	4468346.32	498.61	9.93	-	0.52
11	515557.95	4468369.68	501.51	15.04	0.50	0.60
12	515532.10	4468399.33	504.16	14.58	0.37	0.60
13	515646.51	4468226.07	507.53	10.98	0.23	0.52
14	515614.94	4468264.38	500.21	15.52	0.25	0.55
15	515589.14	4468294.60	495.73	15.36	0.39	0.61
16	515557.87	4468328.80	493.07	12.64	0.34	0.58
17	515535.47	4468350.57	494.45	16.32	0.30	0.57
18	515509.57	4468380.56	498.89	5.30	0.24	0.47
19	515584.89	4468253.41	495.79	18.34	0.38	0.56
20	515520.61	4468321.88	486.21	12.36	0.29	0.60

Table 3.1: Information about insitu measurement in the study area.

3.2.5. UAS measurements (LST, NDVI, DEM)

Cubert hyperspectral camera acquired the images with Band1 NIR; Band2 Red; Band3 Green at 13-June-2019; 15:42 PM with a resolution of 0.048 m., while the thermal images were obtained using the FLIR Tau 2 camera. The images were taken at noon (12:38 PM) and before sunrise (05:13 AM) with a resolution of 16 cm and 15 cm respectively at 14 June 2019. The land surfaces parameters as NDVI, LST and DEM were derived from the UAS measurements. The processed land surface parameters have 15 cm of the spatial resolution. The LST orthomosaic maps are available at sunrise and noon, and the files are in GeoTiff format.

4. RESEARCH METHODOLOGY

4.1. Research design

The random forest-based regression model was used to achieve the research objectives. Firstly, model inputs data were prepared, and among them, there were the ancillary land surface parameters, including land surface temperature (LST), normalized difference vegetation index (NDVI) and digital elevation model (DEM)for providing the topography information. The surface soil moisture products were accessed from the Copernicus data portal, and all ancillary model inputs data were at the same spatial resolution as surface soil moisture of 1km. Particularly, the unmanned aerial system (UAS) measurements acquired in the study area were used to derive land surface parameters (LST, NDVI and DEM) with a 15 cm of spatial resolution to be used as the predictors.

The RF algorithm was trained with the prepared ancillary inputs data at 1km of spatial resolution. The general concept of RF model training was to model the relationship between satellite soil moisture and land surface parameters at coarse resolution (1km). The accuracy of the trained RF models was assessed, and the best-trained model was applied to high spatial resolution land surface features data to predict the soil moisture at 15 cm spatial resolution

Lastly, the downscaled SSM data was validated with the in situ soil moisture from the study area. Figure 4.1 below summarises the main research design idea.

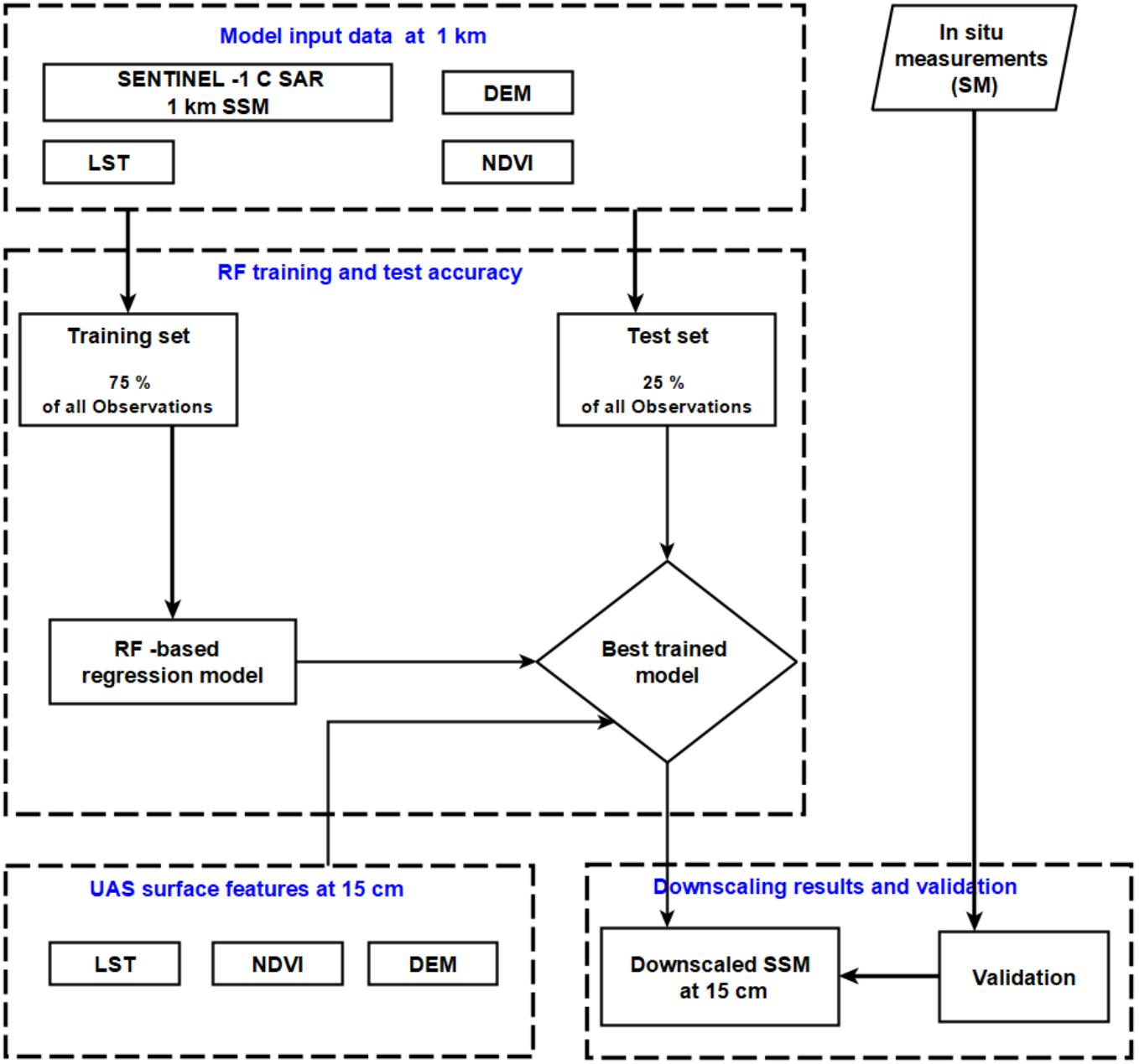


Figure 4.1: Flowchart of the research design for the study

4.2. Preparation of data

The data preparation was done to harmonize the downloaded data to the same file format, same projection and same study area extent. All data were converted into GeoTiff format, and all projections were converted into the same projections. For example, MODIS products have a sinusoidal projection and were reprojected to match the Sentinel -1 projection (WGS_1984).

The analysis was limited to the spatial extent of the study area, and the process of masking the data was done to the fixed two spatial extents mentioned in the data collection section to cover only the study area.

The linear scaling of the pixel values of the satellite images data was done by multiplying with a scale factor and adding offset. For example, the land surface temperature values were obtained in Kelvin by multiplying the data with a scale factor of 0.02 while 1/10000 was done to NDVI values (Didan et al., 2015; Wan, 2013)(e.g., according to MODIS land products user's guide documents). The Figure below represents an example of land surface temperature values on 14 June 2019 from MODIS Tile for the study area.

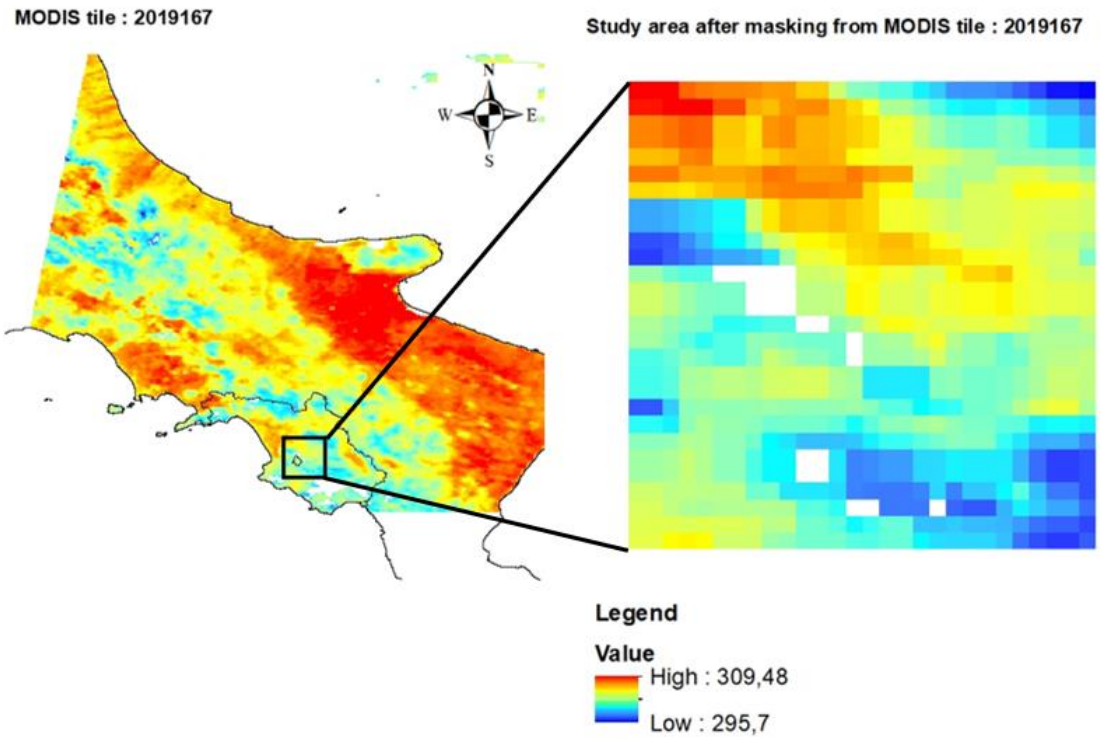


Figure 4.2: Example of land surface temperature values on 14 June 2019 from MODIS Tile for the study area.

On the other hand, linear scaling of pixel values representing the surface soil moisture in percentage was done by converting the physical values from digital values (DN); the multiplication was done with a scale factor of 0.5 on data (Bernhard Bauer, 2019) (e.g., according to Copernicus Global land service user's guide documents).

The figure below also shows an example of coarse resolution SSM on 14 June 2019, which was downscaled into a high-resolution SSM data.

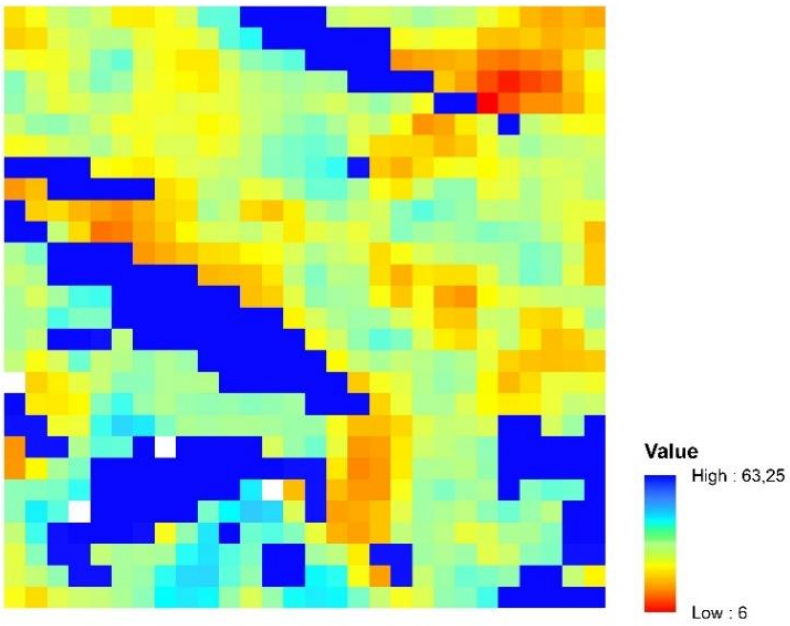


Figure 4.3: Example of coarse resolution surface soil moisture from Sentinel -1 1km on 14 June 2019

Identification of missing data and removing the invalid values were also made by looking into the values of the dataset. The missing data and invalid values were checked by using a statistical summary from the metadata. The invalid values were removed, and missing data were replaced as non-values.

Overview of processed datasets in this study

The table below summarised the processed model input data, consisting of time-series of images from 2015 to 2019.

Table 4.1: Summary of processed model input data

	Datasets (Sensor)	Short Description	Spatial resolution	Processing steps	Number of images
1	Sentinel-1 C-SAR	SSM	1 km	• Subsetting to the study area extent. • linear scaling of the DN values with scale factor	1799
2	MODIS Terra ,MODI11A	LST (Day)	1 km	• Reprojected data • Subsetting from MODIS tiles to Study area extent.	1817

				• Linear scaling of the DN values with scale factor	
4	MODIS Terra MODI12A	NDVI	1 km	• Reprojected data • Subsetting from MODIS tiles to Study area extent. • Linear scaling of the DN values with scale factor	115
5	GTOPO30	DEM	1 km	• Subsetting to Study area extent.	1
6	UAS measurements	LST,NDVI ,DEM	15 cm	▪ linear scaling of the DN values	4

4.3. Overview of surface features derived from the UAS measurements

UAS measurements were taken on the study area of 8.5 ha in the Monteforte Cilento sub-catchment and provided land surface parameters with 15 cm of spatial resolution. The orthomosaic maps were available after processing of the UAV images.

The normalized difference vegetation index (NDVI) was obtained using the images taken by a hyperspectral camera which has different bands. It was obtained as the ratio between near-infrared band 1 (NIR) and red light band 2 (R). UAS derived NDVI values map was available on 13 -June 2019 and was also used to represent the NDVI values for the next day of 14 June 2019.

The land surface temperature was derived from thermal images. Two orthomosaic maps were generated and provided the land surface values on 14 -June -2019 at noon (12:38 PM) and sunrise (05:13 AM). The units of LST values were converted to Kelvin from Celsius by adding 273.15 to the original values.

In addition to the above parameters, the digital elevation model (DEM) map was derived from UAV images at 95 m over the study area. The DEM represented the elevation information associated with each pixel.

4.4. RF-based regression modelling and testing performance

Preparation of the model inputs data was one of the essential steps in machine learning. It reduced the noise in the input data to improve model performance accuracy. In this study, the Random forest algorithm program was implemented in Python (Pedregosa et al., 2011).

The attention was on the quality of the model input data for training the RF model to achieve the good performance of the RF-based regression approach (Long et al., 2019). The quality of the surface parameters and soil moisture products were essential for obtaining good results. The main steps were as follows:

Step 1: Splitting model input datasets and training the RF model

After processing of model inputs data, the next step was to transform the data into RF machine learning understandable terms. Firstly, data were separated into targets and features which were known as the values

to be predicted, and values used to make a prediction respectively. The random forest regression algorithm tried to learn the relationship between the features and the target values during the training phase.
Different model inputs datasets were used for training the RF regression model. Randomly the dataset split up into two sets as train and test sets. During the training phase, the random forest algorithm divided the training set into many regression trees and making up the forest. The RF algorithm used 1000 trees as a maximum number of trees to improve the output results accuracy (Breiman, 2001). Each tree was built from a bootstrap sample which contains about 75 % of the input data. For each bootstrap sampling process, the left samples were 25 % of the input data and were not included in the training. They acted as the Out of bag samples which were used to test the performance of the trained models.
In this study, the training database was composed of different configurations of input datasets. The aim of training the RF model with different input configurations was to understand how much feature data is enough for training and at what spatiotemporal resolutions were needed for training RF. The training of the RF model was done with six different groups of input datasets with different spatiotemporal resolutions such as two spatial extents (28kmx 28km,78kmx 78km) and three periods (2015-2019, 2018 -2019 and 2019).

Step 2: Constructed SSM relationship model based on RF regression method

With the training dataset, the relationship between the land surface parameters data and surface soil moisture was obtained as a function **(F)** using the RF regression method at a coarse resolution of 1km (see below equation).

$$\mathbf{SSM}_{1km} = \mathbf{F}\ (\mathbf{LST}_{1km}, \mathbf{NDVI}_{1km}, , \mathbf{Elevation}_{1km},)\ \ \mathbf{(4.1)}$$

Step 3: Checking the performance of the constructed model

After training the model, the evaluation of RF downscaling models (with different input configurations) was done to assess the performances of the trained models to find out which one has a higher predictive capacity to capture the soil moisture variations in the study area. It was assumed that this RF trained model is spatial scale independent.
In this study, model performance evaluation was done to find out which model was well trained among others. The root mean square error (RMSE), Pearson correlation coefficient (R) and Coefficient of determination (R^2) were used to assess estimated SSM with the observed SSM from sentinel -1 data.
Generally, model performance can be improved by changing (tuning) the model parameters or model hyperparameters (Probst et al., 2019). The model parameters can be changed from model input datasets, while model hyperparameters can be done manually into model settings. For improving the model performance, some model hyperparameters kept at their default settings as from the findings of Probst et al. (2019), which concluded that the Random forest is an algorithm which could provide good results with default settings.

Table 4.2 shows the overview of the used hyperparameters of the random forest algorithm. In this study, the improvement of model performance was obtained by increasing the training set size as the previous study by Zappa et al. (2019) who suggested that it could provide the improvement in model performance accuracy. The RF model performance was improved by using a large enough study area to support the training samples required and to ensure the variability of soil moisture in the study area. Therefore, the model training samples were increased by increasing the spatial extent study area from 28 km by 28km to 78km by 78km.

Table 4.2: Overview of the different hyperparameter values of random forest algorithm (Probst et al., 2019)

Hyperparameter Description	**Description**	**Typical values**
mtry	Number of sampled observations in each split (training set and testing set)	75% for training and 25 % testing
Sample size	Number of observations that are drawn for each tree	n (the number of observations)
Replacement	Sampling the observations with or without replacement	with replacement
Node size	Minimum number of observations in a terminal node	5 for regression
Number of trees	Number of trees in the forest	1000
Splitting rule	Splitting criteria in the nodes	random

Step 4: Checking the role of the surface features on the performance of the RF model

On this step, the best-trained model was known, and the role of each input surface features could be identified from RF model outputs. The random forest algorithm can indicate the relative importance of a variable using commonly used an increased mean square error (MSE), which could be expressed in percentage (Grömping, 2009). It could be used to assess the contribution of different land surface inputs parameters (variables) on the performance of the trained model as well as on the downscaling accuracy. Therefore, the variable importance values were obtained by using the ranking of variable importance of the parameters/ features of the RF algorithm. The larger MSE value means higher the importance of the variable than others (Qu et al., 2019).

Step 5: Model prediction using the surface features derived from UAS measurements

The final step was to predict the soil moisture using the land surface features at a high spatial resolution to achieve the objective of the study. After all, there was the best-trained model to represent the surface soil moisture relationship model (**F**) among others. The best-trained model was applied to the surface parameters derived from UAV data to get the high-resolution SSM data at sub-meter of resolution. The prediction of surface soil moisture was done based on two imageries (Noon and Sunrise):

$SSM_{15\ cm}$ = F ($LST_{15\ cm}$, $NDVI_{15\ cm}$, , $DEM_{15\ cm}$) at Sunrise (4.2)
$SSM_{15\ cm}$ = F ($LST_{15\ cm}$, $NDVI_{15\ cm}$, , $DEM_{15\ cm}$) at Noon (4.3)

The outputs products were two downscaled surface soil moisture products with 15 cm of the spatial resolution. For this analysis, model combinations were based on available LST measurements taken at noon and sunrise.

4.5. Validation of the downscaled soil moisture using the ground measurements

Checking the accuracy of downscaled soil moisture was done using statistical metrics. The downscaled surface soil moisture ($SSM_{Downscaled}$) was compared with ground measurements values ($SSM_{in\ situ}$). The validation was done on 14 June 2019 insitu data. The unbiased root mean square error (ubRMSE), root mean square error (RMSE) and Pearson correlation coefficient (R) were used as the statistical metrics to validate the downscaled surface soil moisture. The correlation coefficient (R) measured the strength of a linear association between downscaled and in situ soil moisture, while the root mean square error (RMSE) indicated how close the in situ measurements were to the model's predicted values (downscaled soil moisture).

The bias was removed to get a better reliable estimation of RMSE, and it gives the ubRMSE. Mean difference (bias) give the average tendency of overestimation (positive value) or underestimation (negative value) of downscaled data. The RMSE and ubRMSE have the same units as the downscaled soil moisture.

- $RMSE = \sqrt{\frac{\sum_{i=1}^{n}(SSM_{Downscaled} - SSM_{In\ situ})^2}{n}}$ **(4.4)**

- $R = \frac{\sum(SSM_{Downscaled} - \overline{SSM}_{Downscaled})(SSM_{In\ situ} - \overline{SSM}_{In\ situ})}{\sqrt{(SSM_{Downscaled} - \overline{SSM}_{Downscaled})^2 (SSM_{In\ situ} - \overline{SSM}_{In\ situ})^2}}$ **(4.5)**

- $Bias = \frac{1}{n}\sum_{i=1}^{n}(SSM_{Downscaled} - SSM_{In\ situ})$ **(4.6)**
- $ubRMSE = \sqrt{(RMSE)^2 - (Bias)^2}$ **(4.6)**

Whereas:

- n= The number of observations
- $SSM_{Downscaled}$ = Downscaled surface soil moisture at 15 cm and 20 m spatial resolution.
- $SSM_{in\ situ}$ = The ground measurements values

5. RESULTS AND DISCUSSION

In this study, the results were discussed in four parts to address the specific objectives. Section 5.1 represented the short description of the relationship between the surface soil moisture and the land surface parameters used for training the models, as well as the derived land surface features from UAS measurements used for predicting high-resolution SSM. Section 5.2 presented RF downscaling results by focusing on the performance of the trained models and the impacts of land surface parameters on their performance, as well as spatial patterns of downscaled SSMs. Section 5.3 discussed the validation results of downscaled soil moisture against in situ soil moisture measurements collected from different locations in the study area. The last section 5.4 highlighted the limitations and opportunity of the research findings.

5.1. The relationship between land surface parameters and SSM

It was essential to use the land surface parameters that influenced surface soil moisture variability to build a good relationship between them. This study considered land surface parameters such as surface temperature (LST), normalized difference vegetation index (NDVI) and digital elevation model (DEM) being factors influencing surface soil moisture variability.

5.1.1. Land surface parameters influencing soil moisture variability

The relationship between LST and SSM was analyzed using the daily mean values from January 2015 to December 2019. Figure 5.1 represents the temporal variations of daily mean values of land surface temperature together with the SSM change. The surface soil moisture showed a visible response to the surface temperature.

Mostly, the surface soil moisture has a high mean SSM value, especially when there are low mean LST values while high values in land temperature representing the dry soil conditions. Other surface parameters could also influence soil moisture variability in the study area.

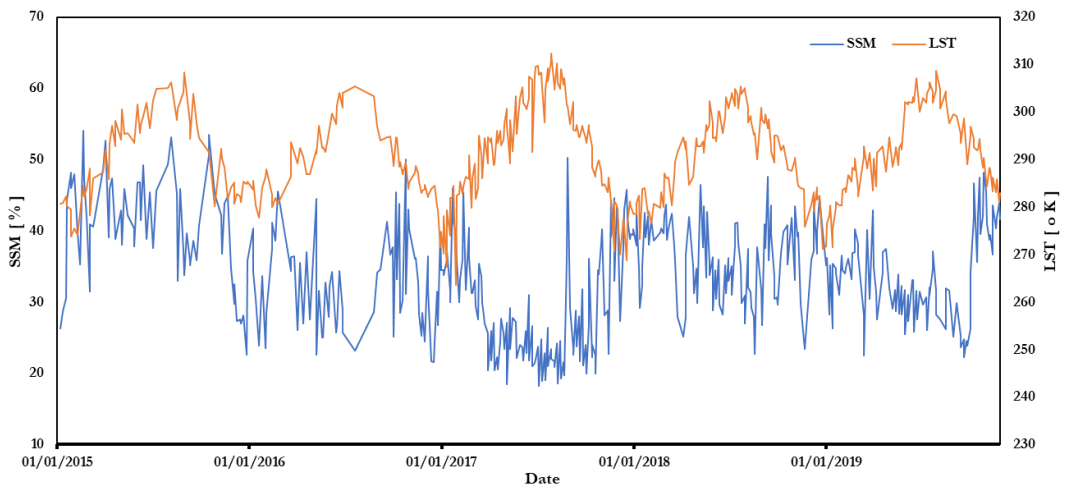


Figure 5.1: Daily mean surface temperature and Sentinel -1 1km daily mean surface soil moisture for available data

The normalized difference vegetation index has 16 days of temporal resolution and was considered as the values stay the same for 16 days period. Figure 5.2 represents the mean NDVI variations with mean SSM values along the study periods. Generally, higher values refer to healthy and dense vegetation, while lower NDVI values show sparse vegetation and bare soil. As mentioned before, Sentinel -1 1km Surface soil moisture was retrieved from radar backscatter measurements and experience low sensitivities to SSM in areas that are vegetated (Piles et al., 2009).

The presence of vegetation reduces the signal from the soil to the sensors, which results in the uncertainty of soil moisture estimation (Peng et al., 2016). The vegetation has an effect on surface soil moisture.

From Figure 5.2, when there are high values of NDVI result in the reduction of surface soil moisture values even though it does not always keep the same trend. There could be uncertainties in the remotely sensed SSM products as NDVI is not the dominant factor to control the SSM change.

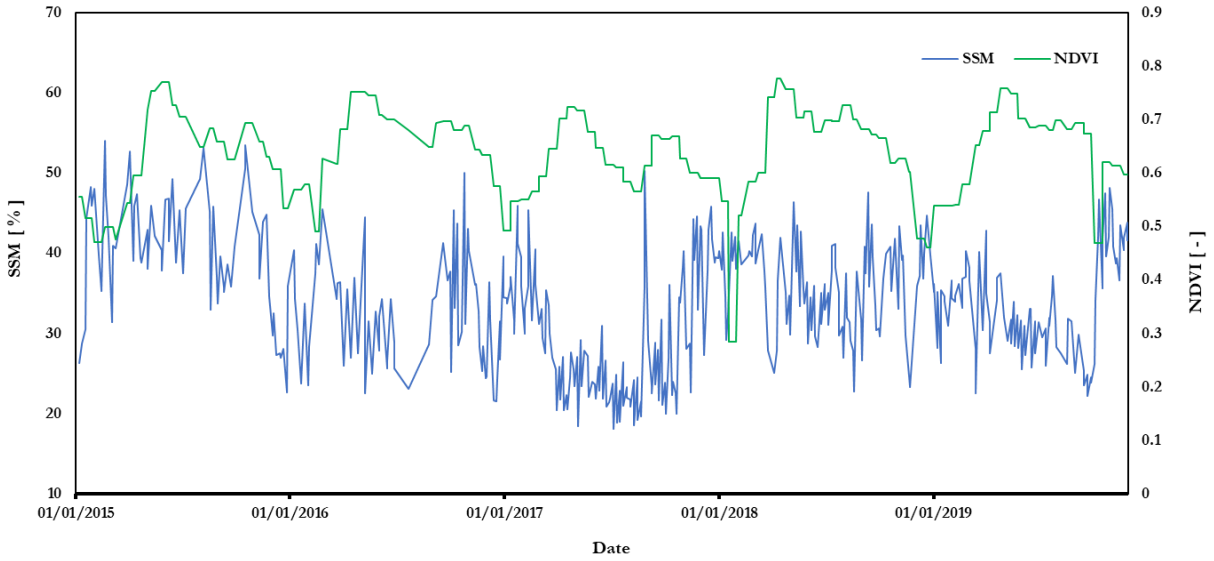


Figure 5.2: Daily mean NDVI values and Sentinel -1 1km Surface soil moisture for available data

The topography was represented by digital elevation model (DEM) including information of slope and elevation. It was also considered as one of the important factors that influence soil moisture, and it was a static parameter which remained the same for a specific time (Hawley, 1983) and was considered the same for the entire period of five years. Figure 5.3 represents the elevation range from 18 m to 1814 m. The mean elevation is approximately 630 m above the sea. The study of Hawley (1983) concluded that slope influences both infiltration and runoff; therefore, the areas with steep slopes are likely to have low soil moisture than flat areas due to lower infiltration and higher runoff rates.

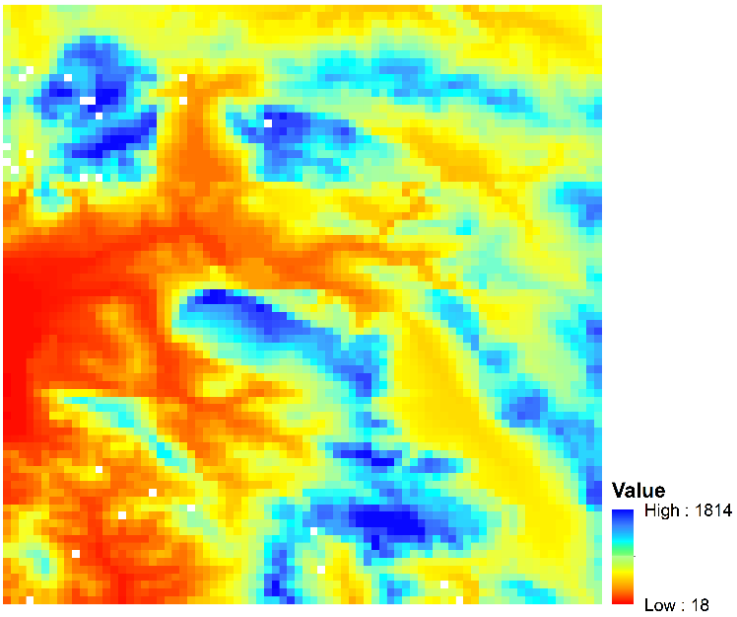


Figure 5.3: Topography representation from DEM data

5.1.2. The UAS derived land surface parameters

The UAS can provide high-resolution measurements of above land surface features. Figure 5.4 shows the spatial representation of derived UAS maps for NDVI, LST and DEM.

The NDVI values were acquired at the sunrise time and were visualized using the gradient of red to green. The map represented UAS-derived NDVI (higher values refer to healthy and dense vegetation, while lower values show sparse vegetation and bare soil). From the visual analysis, there was more vegetation in the study area as it is in the agricultural area.

The LST variations show clear patterns with respect to the types of surface cover such as bare soil or vegetation cover, and the changes in incoming solar radiation at sunrise and noon. Firstly, two LST maps showed detailed spatial information within different land covers with an inverse correlation relationship between LST and NDVI (Yue et al., 2007). Therefore the areas with the least and highest vegetation are experiencing the highest and lowest land surface temperatures respectively. The surface temperature also increases in response to an increasing amount of incident solar radiation (Malbéteau et al., 2018). LST noon figure represented high-temperature variations around solar noon while LSTsunrise shows the lowest variability as expected that in the morning there was minimum solar forcing radiation.

Digital elevation model (DEM) values show a detailed description in the elevation within the study area with a variation range of 70 meters. The east region of the study has a higher elevation, while the west part represents low elevation.

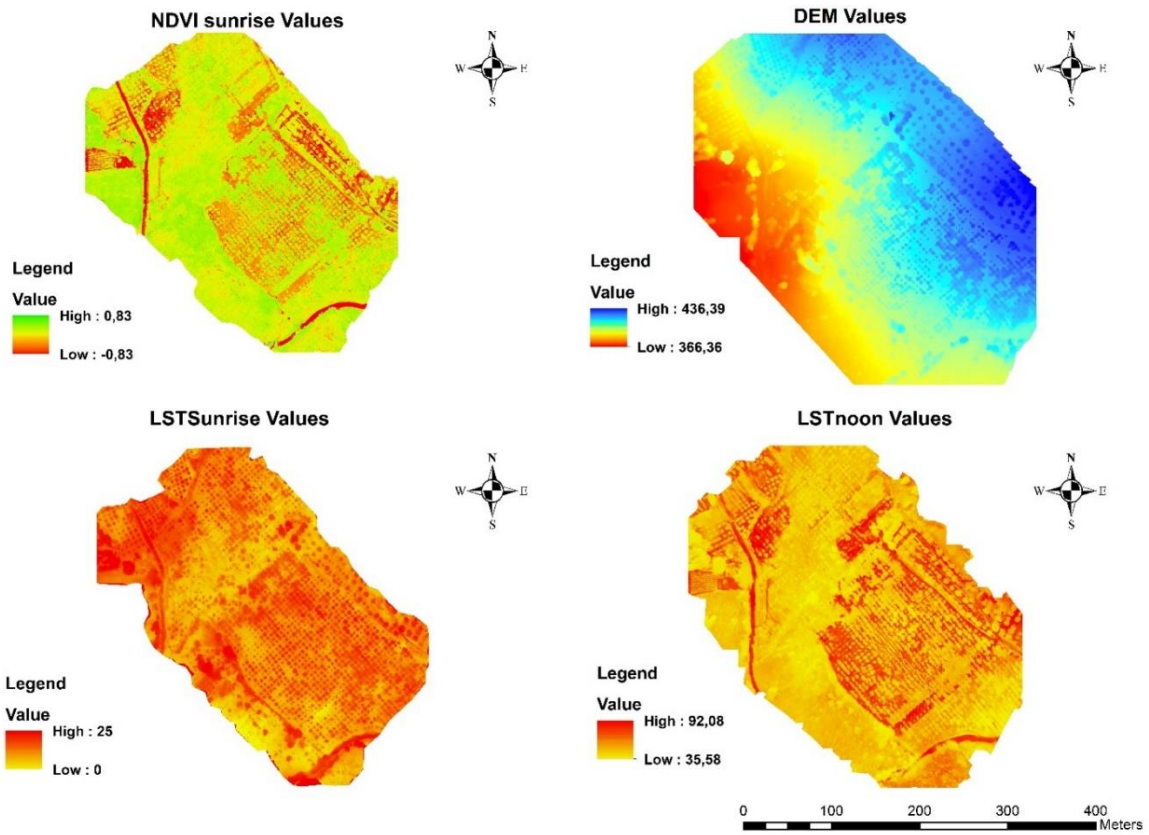


Figure 5.4: Derived land surface parameters from UAS measurements

5.2. RF models outputs and performance accuracy

5.2.1. Trained RF models

The RF models were trained with six different input configurations, based on the combination of NDVI, DEM, and LST (daytime), i.e., two spatial extents (28kmx 28km,78kmx 78km) and three periods (2015-2019,2018-2019 and 2019). Model inputs datasets had different study periods as well as climatic conditions and a considerable enough spatial extent to capture various spatial soil moisture variability and to support the training phase of the RF models. Table 5.1 summarized the used input datasets to establish a relationship model between SSM and land surface parameters. RF algorithm was trained with 75% of the input datasets and then evaluated the estimated results over the remaining 25% of observations.

5.2.2. Performance of the proposed downscaling algorithm

On this stage, there was the trained models and understandings of which one perform best. In this study, root mean square error (RMSE), Pearson correlation coefficient (R) and Coefficient of determination (R^2) provided a comprehensive description of the performance of the trained RF models. The statistical metrics were calculated between the model-predicted soil moisture and the satellite-observed soil moisture (i.e., Sentinel-1 C SAR 1Km SSM).

By varying configurations of training input data, different statistical metrics were obtained to assess the performance of the trained models. Table 5.1 represents the model accuracy results. In general statistical metrics showed that RMSE , R^2, R ranges were 12.3 % – 15.16 % (saturation degree), 0.6 – 0.68 , 0.81 - 0.83 respectively. Based on these findings, all trained models have good performance accuracy in modelling the Sentinel-1 C SAR 1Km SSM, no matter which model input data was used as the training set.

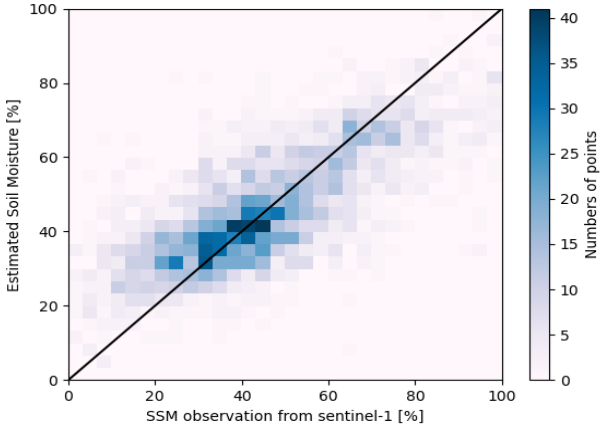
Table 5.1: Summary of model input dataset configurations and RF models outputs.

		Data splitting		**Model accuracy metrics**		
Model inputs datasets	**Spatial extent [km]**	**Trainings sets**	**Test sets**	**RMSE [Saturation Degree %]**	**R^2**	**Pearson correlation coefficient**
2015 -2019	28 x 28	1238	413	13.45	0.62	0.82
	78 x 78	18750	6251	15.16	0.61	0.81
2018-2019	28 x 28	499	167	13.3	0.63	0.82
	78 x 78	8016	2673	12.13	0.63	0.83
2019	28 x 28	345	115	13.99	0.6	0.81
	78 x 78	5279	1760	12.3	0.62	0.82

Comparing among six trained models with different input configurations, the result of 2018 2019 dataset with 78km x78km extent combination was better than others with a high degree of Pearson's correlation coefficient (R) of 0.83, R^2 of 0.63, RMSE of 12.13 %. The results were promising and confirmed the ability of random forest-based regression model to represent a relationship between the surface soil moisture and used ancillary land surface parameters.

Figures 5.5 represents the relationship between the estimated and the Sentinel-1 C SAR 1Km SSM data, and there is a good agreement between them as the distribution of data points is more concentrated on both sides of the 1:1 line. In general, model performance supposed to improve with an increase of training dataset (i.e., satellite observations and relevant land surface features) (Zappa et al., 2019). As a result, the model performance was improved from R of 0.82 to 0.83 by increasing the spatial extent from 28km x 28km to 78km x78km, which captured large area satellite observations and relevant land surface features. The appendix A shows the 2D histogram plots for all trained model scenarios, and there is a satisfactory degree of fitting for all trained models.

Results on 2018 -2019 datasets (78kmx78km extent)



Results on 2018 -2019 datasets (28kmx28km extent)

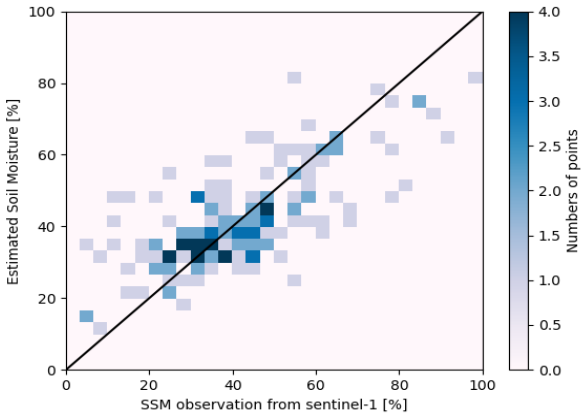


Figure 5.5: 2D histograms plots of estimated soil moisture results based on RF model and the original Sentinel-1 C SAR 1Km SSM for 2019 dataset

However, the accuracy of the trained models could be influenced by the model input parameters. Generally, the uncertainties can be due to different impacts of vegetation growth and the climatic conditions over the selected study area extent on the SSM variations. Additionally, the results obtained could also have uncertainty in the estimated soil moisture as the used land surface parameters do not capture all soil parameters (e.g., soil texture, rainfall, etc.), which mainly affect the spatial heterogeneity of the soil moisture in the study area within space and time (Abbaszadeh et al., 2019)

5.2.3. Role of land surface parameters on the downscaling performance

RF algorithm provided the relative importance of each input land surface features (surface parameters) as an increased mean square error (%IncMSE). The larger value of %IncMSE of a surface feature, the more that parameter contributed to the model. Analysis of the relative importance of different surface parameters (LST, NDVI, and DEM) was essential to understand their impacts on the performance of the trained RF model to predict the soil moisture. The results presented in Figure 5.6 indicate the surface features with variations the relative importance values for the best-trained model. There are also vertical error bars which show the standard deviation of the importance scores of each variable and indicate that there is no high variation of the values from the average values.

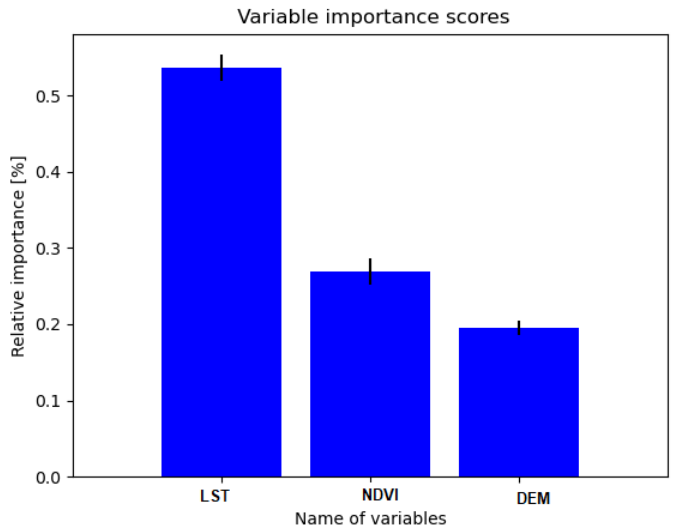


Figure 5.6: The variable importance scores for the RF-based approach

The land surface temperature (LST) has the highest value of 0.54 while normalized difference vegetation index (NDVI) with 0.26, which has a higher value than DEM with 0.20. The variations of SSM are mainly affected by LST and NDVI because of their impacts on surface energy fluxes and vegetation status (Long et al., 2019).

As the time series graphs in Section 5.1 showed the relationship changes between LST and NDVI with SSM data. Figure 5.1 shows a good inverse correlation between the SSM and land surface temperature (LST). Similar results found by Im et al. (2016) and Zhao et al. (2018) that land surface temperature are closely related to soil emission, which influences the soil moisture variability. The variation of soil moisture is also affected by vegetation and was validated by the findings of Gómez-Plaza et al. (2001) that the spatial variability of surface soil moisture is affected by vegetation. The above mentioned parameters are not the only ones to influence SSM variations, and other factors could change the soil moisture due to different growth periods and climate conditions in the study area (Hawley, 1983)

The digital elevation model (DEM) appeared to be the least essential parameter in this study. Previous studies by L. Brocca et al. (2007) proved that the influence of the DEM depends on the topography of the study area and there is less soil moisture variability at low elevation areas, which is the same as our study area. The fact that topography kept being the same for the entire study period influence less the spatiotemporal variability of soil moisture.

The importance of variable derived from RF algorithm is appropriate to understand the relationship between soil moisture product and land surface parameters, and it allowed to understand the potential physical mechanisms which influence the downscaled soil moisture variability in the study area. Generally, all parameters have a significant influence on the performance of the proposed RF model accuracy.

5.2.4. Spatial distribution of downscaled soil moisture products

After getting the best trained random forest model representing the relationship between surface soil moisture and the other surface parameters. The model was applied to the high-resolution land surface parameters derived from UAS measurement to obtain high-resolution surface soil moisture data over Monteforte Cilento Sub catchment.

The downscaled surface soil moisture maps were generated with two UAS imageries (i.e., at both noon and sunrise). Figure 5.7 &5.8 represent the spatial representation of downscaled SSM results on 14 June 2019 at two abovementioned specific times. The SSM products are expressed in saturation degree (%).

In general, the spatial pattern of two downscaled soil moisture products had been influenced by the surface features (LST, NDVI and DEM) in the study area. The results provide detailed information on spatial variations of surface soil moisture (5cm depth) distribution with 15 cm spatial resolution. The predicted surface soil moisture values are limited to the ranges of the coarse resolution surface soil moisture values covered by the training model input data (Zhao et al., 2018a). In other words, the downscaled SSM values were more related to the quality and values of the Sentinel -1 SSM.

The SSM values range between 37.49 % - 70.11 % and 30.86 – 47.01 % at sunrise and noontime respectively. Two downscaled products have differences between them as the downscaled SSM values at sunrise appeared to be higher than at noon values due to influences of LST values (i.e., LST at sunrise is much lower than that at noon). Such difference is supported by the fact that the land surface temperature was found as the most influencing parameter for trained RF model.

The results showed that the RF-based downscaling approach could generate high-resolution surface soil moisture by utilising the satellite product and UAS measurements. The downscaled soil moisture represents the sub-grid variation of soil moisture with the pixel of the satellite soil moisture datasets. The next step is to validate the downscaled SSM with in-situ measurements.

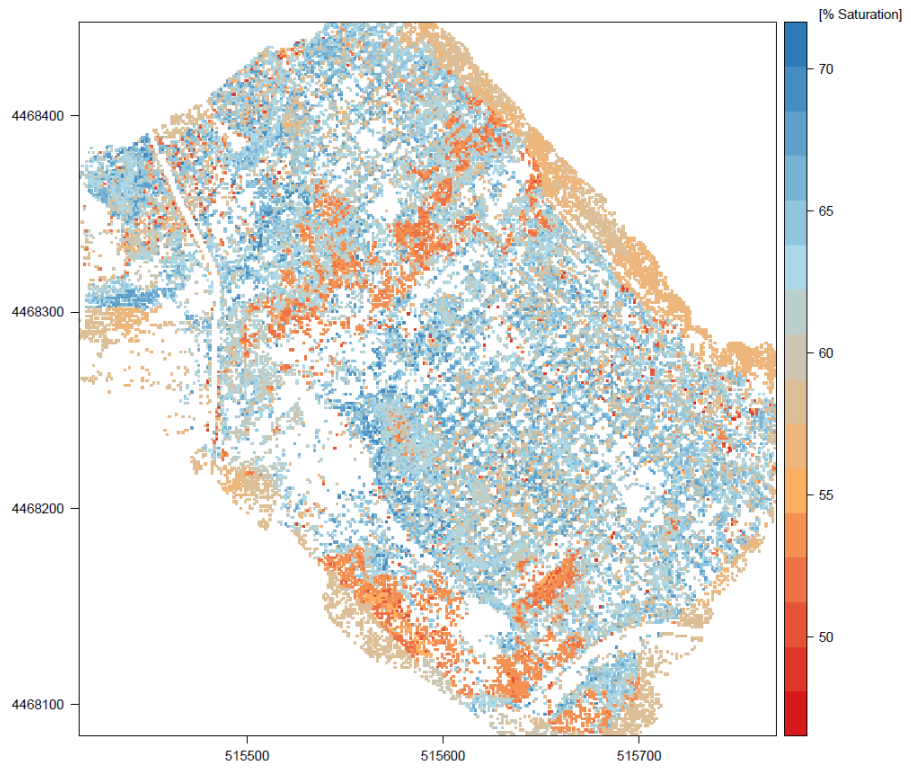


Figure 5.7: Spatial distribution of downscaled soil moistures on 14 June 2019 Sunrise time

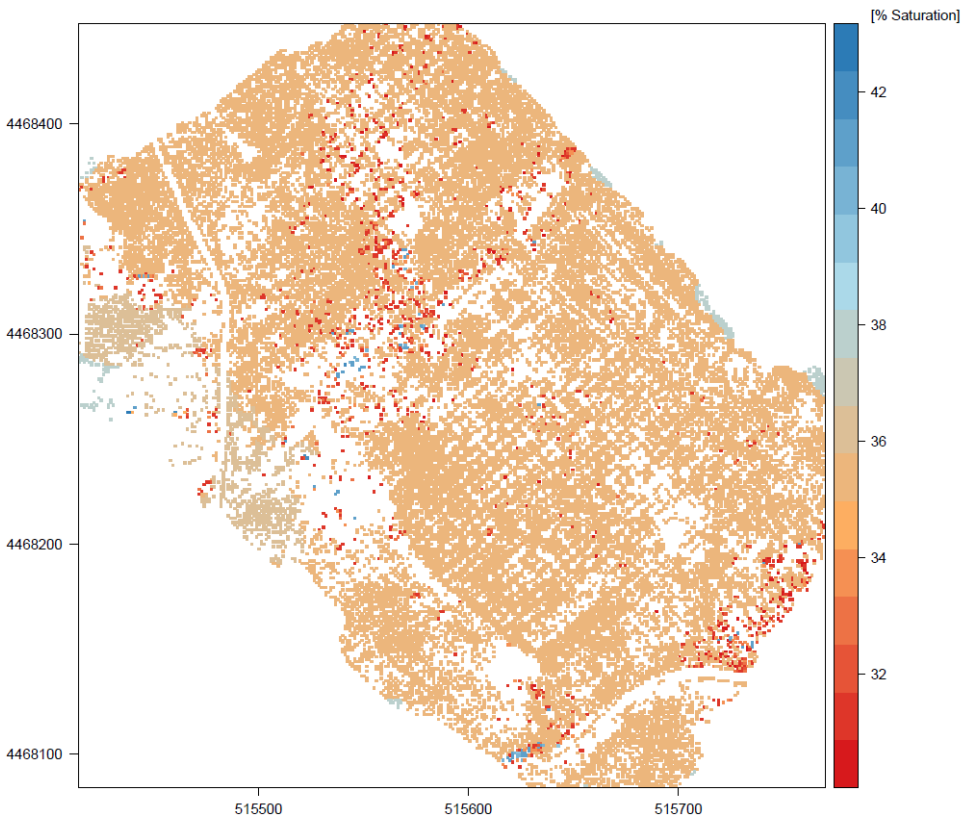


Figure 5.8: Spatial distribution of downscaled soil moistures on 14 June 2019 at Noontime

5.3. Validation of the downscaled surface soil moisture

The downscaled surface soil moistures at noon and sunrise were validated using the in situ measurements which were collected on 14 June 2019 from twenty field stations located in Monteforte Cilento Sub catchment (Figure 3.1). The downscaled SSM were extracted from the spatial maps, and the SSM values were converted into volumetric SSM by multiplying with the porosity values as in Table 3.1.

Table 5.3 showed validation results from the 14 points measurements. The average unbiased root mean square error (ubRMSE**)** was 0.07 cm^3/cm^3, and the Pearson correlation coefficient (R) ranged between 0.51 - 0.63 with the in situ measurements.

Table 5.2: Validation results of downscaled soil moisture based on RF-based regression model.

Downscaling results	Statistical metrics		
	RMSE [$cm3cm^{-3}$]	**ubRMSE [$cm3cm^{-3}$]**	**Pearson correlation coefficient [-]**
Sunrise	0.07	0.07	0.60
Noon	0.14	0.07	0.51

The accuracy level showed a good agreement based on the calculated correlation coefficient between the downscaled SSM and in-situ measurements. The points are closer to line 1:1 which indicate the predicted SSM values are in good agreement with the in-situ measured SSM and prove the success of the RF approach to generate high-resolution SSM that captures the spatial distribution of soil moisture in the study area.

Even though there is a correlation between them, the results still did not reach the required accuracy agreement which is similar to the target accuracy of ubRMSE = 0.04 cm^3/cm^3 in the top 5 cm suggested by the Global Climate Observing System (GCOS) requirement (Bauer-Marschallinger & Schaufler, 2018) and the SMAP science team (Entekhabi et al. .2010).

This disagreement between downscaled SSM and in-situ measurements can be due to the result of downscaled SSM products were generated at the topsoil layer (5 cm of depth) while the in situ measurements were measured at 15 cm of depth. Validation of downscaled SSM was compromised by the lack of comparable in situ soil moisture datasets with similar measurements at topsoil level as the downscaled soil moisture.

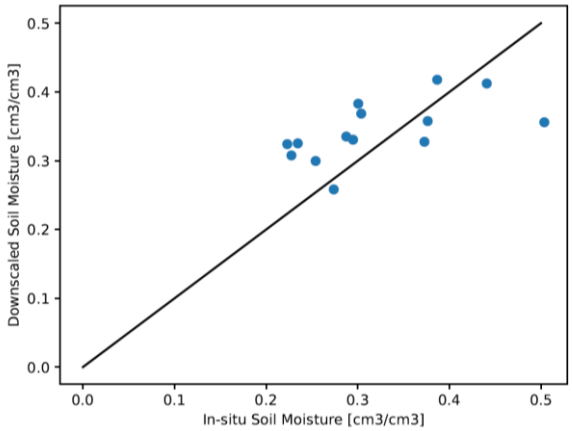


Figure 5.9: Validation results obtained with sunrise measurements

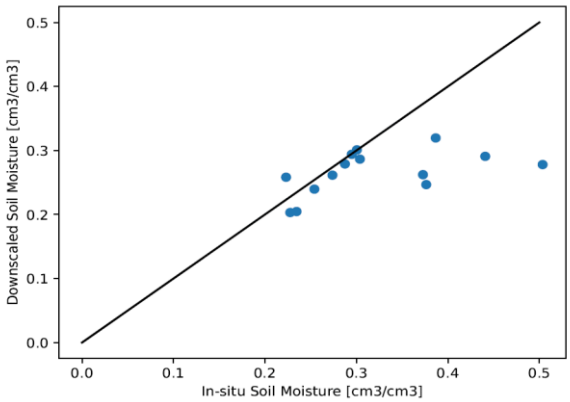


Figure 5.10: Validation results obtained with noon measurements

5.4. Advantages and limitations of this study

The proposed RF-based regression downscaling approach proved to be able to generate high-resolution SSM data using the UAS measurements. Regardless of the promising results, this study also has the limitations to achieve the abovementioned required accuracy.

The proposed downscaling approach was limited on a few land surface parameters as LST, NDVI, DEM. The performance of the model could have also been improved by including other different land surface parameters (e.g., soil texture, precipitation, etc.) which mainly affect the spatial heterogeneity of the soil moisture in the study area within space and time (Abbaszadeh et al., 2019). The other limitation was about the in-situ measurements which were taken at 15 cm depth while the downscaled SSM products were estimated at 5 cm, which compromised the validation results.

Even though the current study has the limitations, the implemented approach could provide sub-meter resolution soil moisture data for guiding the agricultural practices at the scale local farmer desired. It was to note that as there were only UAS measurements for two days over the study area (13 &14 June 2019), the current study was limited to that day. Nevertheless, this work served as a proof of concept study. It is expected that with the approach developed in this study, Sentinel-1 and Sentinel-2 data could be used to detect/monitor the field 'hotspots' (e.g. too much or too little water for plant/crop), which can be further examined with UAV flights when needed. Therefore, the proposed random forest regression model can be applied to coarse satellite data to generate high-resolution soil moisture using the derived land UAS land surface parameters.

6. CONCLUSION AND RECOMMENDATIONS

6.1. Summary and conclusion

In this research, the random forest method was used to downscale Sentinel -1 SSM from 1km to 15 cm spatial resolution in the Alento river catchment to provide valuable detail information for guiding the agricultural practices at the field scale. RF model was trained with various inputs data sets at 1km resolution, including SSM and land surface parameters, for establishing a relationship between SSM and the land surface parameters. The best-trained model was applied to the land surface features derived from UAS measurements to get high spatial SSM data at 15 cm resolution. This study considered LST, NDVI and DEM being essential predictors influencing surface soil moisture variability.

The analysis of trained RF models performance was tested using three statistical metrics between the estimated SSM and the observed Sentinel -1 SSM. It was shown that almost all the trained models have good performance. The results of model run with 2018-2019 data sets on 78km x78km spatial extent outperformed other trained models with high correlation (R = 0.83) and low values of RMSE (RMSE = 12.13 % saturation degree) and R^2 (R^2 = 0.63). The best trained RF-based model was used to generate high spatial resolution surface soil moisture.

Random forest algorithm provided the relative importance of the surface parameters to evaluate the impacts of surface parameters on SSM variabilities. The increased mean square error (%IncMSE) obtained from the RF model results was used to present relative importance scores. The results showed that all parameters have remarkable effects on predicting SSM. LST has a more significant impact than other with high %IncMSE (54 %), NDVI with %IncMSE= 26 % and topography were less than other input parameters with least %IncMSE = 20 %. The visual analysis of downscaled surface soil moisture maps showed sub-grid variability within the pixel of the original Sentinel -1 SSM. The downscaled surface soil moisture was affected by the LST values with high and low SSM value ranges at sunrise and noontime, respectively.

Furthermore, the downscaled SSM results were validated using the 14 in-situ SSM point measurements from SoilNet sensors installed at 15 cm (using the data collected 14 June 2019 at noon and sunrise time when there are UAS flights). Generally, validation results showed relatively good agreement with the in-situ measurements though did not reach GCOS requirement ubRMSE of 0.04 cm^3/cm^3. There are biases in validation results because the downscaled SSM products were generated at 5 cm of topsoil while the in situ measurements were measured at 15 cm.

The overall analysis indicates that RF method could be suitable to downscale the satellite-based soil moisture to get high spatial resolution soil moisture data with integration of UAS measurements and the accuracy of downscaled SSM data were mainly depends on the quality of model inputs data.

6.2. Recommendations

Generally, the results are encouraging, and the RF-based regression algorithm method downscaled coarse resolution satellite-based SSM products utilising the derived high spatial resolution land surface parameters from UAS measurements. The success of this study was mainly based on the accuracy of downscaled results with a high spatial resolution which depends on the quality of inputs datasets and the performance of the proposed method of downscaling.

In the future, further researches can be concentrated on the following :

- The same approach could also be applied with the Sentinel -2 data to do the downscaling. Sentinel-2 can provide the land surface parameters at 10 m resolution.
- The improvement of the model performance can also be made by tuning the hyperparameters of the RF algorithm.
- The spatial downscaling of coarse resolution surface soil moisture data could also be achieved by introducing other additional land surface features related to surface soil moisture (such as soil texture, rainfall, etc.).

LIST OF REFERENCES

Aasen, H., Burkart, A., Bolten, A., & Bareth, G. (2015). Generating 3D hyperspectral information with lightweight UAV snapshot cameras for vegetation monitoring: From camera calibration to quality assurance. *ISPRS Journal of Photogrammetry and Remote Sensing*, *108*, 245–259. https://doi.org/10.1016/j.isprsjprs.2015.08.002

Abbaszadeh, P., Moradkhani, H., & Zhan, X. (2019). Downscaling SMAP Radiometer Soil Moisture Over the CONUS Using an Ensemble Learning Method. *Water Resources Research*, *55*(1), 324–344. https://doi.org/10.1029/2018WR023354

Balasubramanian, A. (2017). Digital Elevation Model (Dem) in Gis. In *University of Mysore*. https://doi.org/10.13140/RG.2.2.23976.47369

Bartkowiak, P., Castelli, M., & Notarnicola, C. (2019). Downscaling land surface temperature from MODIS dataset with random forest approach over alpine vegetated areas. *Remote Sensing*, *11*(11), 1–19. https://doi.org/10.3390/rs11111319

Bauer-Marschallinger, B., & Schaufler, S. (2018). *Algorithm theoretical basis document surface soil moisture collection 1km version 1.0 issue 1.20*.

Bernhard Bauer, M. (2019). *Copernicus Global Land Operations,Product user manual* (Issue 1.30).

Bhandari, A. K., Kumar, A., & Singh, G. K. (2012). Feature Extraction using Normalized Difference Vegetation Index (NDVI): a Case Study of Jabalpur City. *Procedia Technology*, *6*, 612–621. https://doi.org/10.1016/j.protcy.2012.10.074

Borzuchowski, J., & Schulz, K. (2010). Retrieval of leaf area index (LAI) and soil water content (WC) using hyperspectral remote sensing under controlled glass house conditions for spring barley and sugar beet. *Remote Sensing*, *2*(7), 1702–1721. https://doi.org/10.3390/rs2071702

Breiman, L. (2001). Random forests. In Robert E. Schapire (Ed.), *Machine Learning* (pp. 1–122). Kluwer Academic Publishers. https://doi.org/10.1007/978-3-662-56776-0_10

Brocca, L., Morbidelli, R., Melone, F., & Moramarco, T. (2007). Soil moisture spatial variability in experimental areas of central Italy. *Journal of Hydrology*, *333*(2–4), 356–373. https://doi.org/10.1016/j.jhydrol.2006.09.004

Brocca, Luca, Ciabatta, L., Massari, C., Camici, S., Tarpanelli, A., Brocca, L., Ciabatta, L., Massari, C., Camici, S., & Tarpanelli, A. (2017). Soil moisture for hydrological applications: Open questions and new opportunities. *Water*, *9*(2), 140. https://doi.org/10.3390/w9020140

Chen, S., She, D., Zhang, L., Guo, M., & Liu, X. (2019). Spatial downscaling methods of soil moisture based on multisource remote sensing data and its application. *Water (Switzerland)*, *11*(7), 1–25. https://doi.org/10.3390/w11071401

Crow, W. T., Berg, A. A., Cosh, M. H., Loew, A., Mohanty, B. P., Panciera, R., De Rosnay, P., Ryu, D., & Walker, J. P. (2012). Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products. *Reviews of Geophysics*, *50*(2). https://doi.org/10.1029/2011RG000372

Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random Forests BT - Ensemble Machine Learning: Methods and Applications. In *Ensemble Machine Learning* (Vol. 45, pp. 157–175). Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-9326-7_5

Dash, P., Göttsche, F. M., Olesen, F. S., & Fischer, H. (2002). Land surface temperature and emissivity estimation from passive sensor data: Theory and practice-current trends. *International Journal of Remote Sensing*, *23*(13), 2563–2594. https://doi.org/10.1080/01431160110115041

Didan, K., Barreto Munoz, A., Solano, R., & Huete, A. (2015). *MODIS Vegetation Index User's Guide (MOD13 Series)*. http://vip.arizona.edu

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., Entin, J. K., Goodman, S. D., Jackson, T. J., Johnson, J., Kimball, J., Piepmeier, J. R., Koster, R. D., Martin, N., McDonald, K. C., Moghaddam, M., Moran, S., Reichle, R., Shi, J. C., … Van Zyl, J. (2010). The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE*, *98*(5), 704–716. https://doi.org/10.1109/JPROC.2010.2043918

EU Science HUB. (2019). *European droughts in 2018: a warning of things to come | EU Science Hub*. https://ec.europa.eu/jrc/en/news/european-droughts-2018-warning-things-come

Gillins, D. T., Parrish, C., Gillins, M. N., & Simpson, C. (2018). *Eyes in the Sky: Bridge Inspections With Unmanned Aerial Vehicles*. 1–180. https://www.oregon.gov/ODOT/Programs/Pages/Research-Publications.aspx

Gómez-Plaza, A., Martínez-Mena, M., Albaladejo, J., & Castillo, V. M. (2001). Factors regulating spatial distribution of soil water content in small semiarid catchments. *Journal of Hydrology*, *253*(1–4), 211–226. https://doi.org/10.1016/S0022-1694(01)00483-8

Grömping, U. (2009). Variable importance assessment in regression: Linear regression versus random forest. *American Statistician*, *63*(4), 308–319. https://doi.org/10.1198/tast.2009.08199

Hawley. (1983). SURFACE SOIL MOISTURE VARIATION ON SMALL AGRICULTURAL WATERSHEDS. *Journal of Hydrology*, *62*, 179–200. https://doi.org/10.1016/0022-1694(83)90102-6

Hillel, D., Warrick, A. W., Baker, R. S., & Rosenzweig, C. (1998). *Environmental soil physics* (1st editio). Academic Press.

Hsu, W.-L., & Chang, K.-T. (2019). Cross-estimation of soil moisture using thermal infrared images with different resolutions. *Sensors and Materials*, *31*(2), 387–398. https://doi.org/10.18494/SAM.2019.2090

Im, J., Park, S., Rhee, J., Baik, J., & Choi, M. (2016). Downscaling of AMSR-E soil moisture with MODIS products using machine learning approaches. *Environmental Earth Sciences*, *75*(15), 1120. https://doi.org/10.1007/s12665-016-5917-6

Jeziorska, J. (2019). UAS for Wetland Mapping and Hydrological Modeling. *Remote Sensing*, *11*(17), 1997. https://doi.org/10.3390/rs11171997

Jing, W., Yang, Y., Yue, X., & Zhao, X. (2016). A spatial downscaling algorithm for satellite-based precipitation over the Tibetan plateau based on NDVI, DEM, and land surface temperature. *Remote Sensing*, *8*(8). https://doi.org/10.3390/rs8080655

Justice, C. O., Townshend, J. R. G., Vermote, E. F., Masuoka, E., Wolfe, R. E., Saleous, N., Roy, D. P., & Morisette, J. T. (2002). An overview of MODIS Land data processing and product status. *Remote Sensing of Environment*, *83*(1–2), 3–15. https://doi.org/10.1016/S0034-4257(02)00084-6

Kim, D., Moon, H., Kim, H., Im, J., & Choi, M. (2018). Intercomparison of Downscaling Techniques for Satellite Soil Moisture Products. *Advances in Meteorology*, *2018*, 16. https://doi.org/10.1155/2018/4832423

Long, D., Bai, L., Yan, L., Zhang, C., Yang, W., Lei, H., Quan, J., Meng, X., & Shi, C. (2019). Generation of spatially complete and daily continuous surface soil moisture of high spatial resolution. *Remote Sensing of Environment*, *233*(August), 111364. https://doi.org/10.1016/j.rse.2019.111364

M. C. Peel1, B. L. Finlayson2, and T. A. M. (2002). Updated world map of the K¨oppen-Geiger climate classificatio. *Permafrost and Periglacial Processes*, *13*(3), 243–249. https://doi.org/10.1002/ppp.421

Malbéteau, Y., Parkes, S., Aragon, B., Rosas, J., & McCabe, M. F. (2018). Capturing the diurnal cycle of land surface temperature using an unmanned aerial vehicle. *Remote Sensing*, *10*(9). https://doi.org/10.3390/rs10091407

Manfreda, S., McCabe, M. F., Miller, P. E., Lucas, R., Madrigal, V. P., Mallinis, G., Dor, E. Ben, Helman, D., Estes, L., Ciraolo, G., Müllerová, J., Tauro, F., de Lima, M. I., de Lima, J. L. M. P., Maltese, A., Frances, F., Caylor, K., Kohv, M., Perks, M., … Toth, B. (2018). On the use of unmanned aerial systems for environmental monitoring. *Remote Sensing*, *10*(4). https://doi.org/10.3390/rs10040641

Nasta, P., Romano, N., & Chirico, G. B. (2013). Functional evaluation of a simplified scaling method for assessing the spatial variability of soil hydraulic properties at the hillslope scale. *Hydrological Sciences Journal*, *58*(5), 1059–1071. https://doi.org/10.1080/02626667.2013.799772

Njoku, E. G., Wilson, W. J., Yueh, S. H., & Rahmat-Samii, Y. (2000). A large-antenna microwave radiometer-scatterometer concept for ocean salinity and soil moisture sensing. *IEEE Transactions on Geoscience and Remote Sensing*, *38*(6), 2645–2655. https://doi.org/10.1109/36.885211

O'Neill, P. E., Chauhan, N. S., & Jackson, T. J. (1996). Use of active and passive microwave remote sensing for soil moisture estimation through corn. *International Journal of Remote Sensing*, *17*(10), 1851–1865. https://doi.org/10.1080/01431169608948743

Ochsner, E., Cosh, M. H., Cuenca, R., Dorigo, H., Draper, S., Hagimoto, Y., Kerr, Y. H., Larson, M., Njoku, E. G., & Small, E. (2013). State of the art in large-scale soil moisture monitoring. *Soil Scince Society of America Journal*, 1–32. https://doi.org/10.2136/sssaj2013.03.0093ï

Pablos, M., Martínez-Fernández, J., Piles, M., Sánchez, N., Vall-llossera, M., & Camps, A. (2016). Multi-temporal evaluation of Soil Moisture and land surface temperature dynamics using in situ and satellite observations. *Remote Sensing*, *8*(7). https://doi.org/10.3390/rs8070587

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. http://arxiv.org/abs/1201.0490

Peng, J., Loew, A., Merlin, O., & Verhoest, N. E. C. (2017a). A review of spatial downscaling of satellite remotely sensed soil moisture. *Reviews of Geophysics*, *55*(2), 341–366. https://doi.org/10.1002/2016RG000543

Peng, J., Loew, A., Merlin, O., & Verhoest, N. E. C. (2017b). A review of spatial downscaling of satellite remotely sensed soil moisture. *Reviews of Geophysics*, *55*(2), 341–366. https://doi.org/10.1002/2016RG000543

Peng, J., Loew, A., Zhang, S., Wang, J., & Niesel, J. (2016). Spatial Downscaling of Satellite Soil Moisture Data Using a Vegetation Temperature Condition Index. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(1), 558–566. https://doi.org/10.1109/TGRS.2015.2462074

Peng, Niesel, J., & Loew, A. (2015). Evaluation of soil moisture downscaling using a simple thermal-based proxy-the REMEDHUS network (Spain) example. *Hydrology and Earth System Sciences*, *19*(12), 4765–4782. https://doi.org/10.5194/hess-19-4765-2015

Piles, M., Entekhabi, D., & Camps, A. (2009). A change detection algorithm for retrieving high-resolution soil moisture from SMAP radar and radiometer observations. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(12), 4125–4131. https://doi.org/10.1109/TGRS.2009.2022088

Poullaouec, J., Bourbigot, M., Riccardo Piantamida, H. J., & Hajduch, G. (2016). *Sentinel-1 Products Overview*.

Probst, P., Wright, M., & Boulesteix, A.-L. (2019). *Hyperparameters and Tuning Strategies for Random Forest*.

Qu, Y., Zhu, Z., Chai, L., Liu, S., Montzka, C., Liu, J., Yang, X., Lu, Z., Jin, R., Li, X., Guo, Z., & Zheng, J. (2019). Rebuilding a Microwave Soil Moisture Product Using Random Forest Adopting AMSR-E/AMSR2 Brightness Temperature and SMAP over the Qinghai–Tibet Plateau, China. *Remote Sensing*, *11*(6), 683. https://doi.org/10.3390/rs11060683

Robinson, D. A., Campbell, C. S., Hopmans, J. W., Hornbuckle, B. K., Jones, S. B., Knight, R., Ogden, F., Selker, J., & Wendroth, O. (2008). Soil Moisture Measurement for Ecological and Hydrological Watershed-Scale Observatories: A Review. *Vadose Zone Journal*, *7*(1), 358–389. https://doi.org/10.2136/vzj2007.0143

Romano, N., Nasta, P., Bogena, H., De Vita, P., Stellato, L., & Vereecken, H. (2018). Monitoring hydrological processes for land and water resources management in a mediterranean ecosystem: The alento river catchment observatory. *Vadose Zone Journal*, *17*(1). https://doi.org/10.2136/vzj2018.03.0042

Rötzer, K., Montzka, C., & Vereecken, H. (2015). Spatio-temporal variability of global soil moisture products. *Journal of Hydrology*, *522*, 187–202. https://doi.org/10.1016/j.jhydrol.2014.12.038

Sabaghy, S., Walker, J. P., Renzullo, L. J., Akbar, R., Chan, S., Chaubell, J., Das, N., Dunbar, R. S., Entekhabi, D., Gevaert, A., Jackson, T. J., Loew, A., Merlin, O., Moghaddam, M., Peng, J., Peng, J., Piepmeier, J., Rüdiger, C., Stefan, V., … Yueh, S. (2020). Comprehensive analysis of alternative downscaled soil moisture products. *Remote Sensing of Environment*, *239*(November 2019). https://doi.org/10.1016/j.rse.2019.111586

Sabaghy, S., Walker, J. P., Renzullo, L. J., & Jackson, T. J. (2018a). Spatially enhanced passive microwave derived soil moisture: Capabilities and opportunities. *Remote Sensing of Environment*, *209*(January), 551–580. https://doi.org/10.1016/j.rse.2018.02.065

Sabaghy, S., Walker, J. P., Renzullo, L. J., & Jackson, T. J. (2018b). Spatially enhanced passive microwave derived soil moisture: Capabilities and opportunities. *Remote Sensing of Environment*, *209*(March), 551–580. https://doi.org/10.1016/j.rse.2018.02.065

Sagan, V., Maimaitijiang, M., Sidike, P., Eblimit, K., Peterson, K. T., Hartling, S., Esposito, F., Khanal, K., Newcomb, M., Pauli, D., Ward, R., Fritschi, F., Shakoor, N., & Mockler, T. (2019). UAV-based high resolution thermal imaging for vegetation monitoring, and plant phenotyping using ICI 8640 P, FLIR Vue Pro R 640, and thermomap cameras. *Remote Sensing*, *11*(3). https://doi.org/10.3390/rs11030330

Song, C., Jia, L., & Menenti, M. (2014). Retrieving high-resolution surface soil moisture by downscaling AMSR-E brightness temperature using MODIS LST and NDVI data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(3), 935–942. https://doi.org/10.1109/JSTARS.2013.2272053

Sun, D., & Pinker, R. T. (2004). Case study of soil moisture effect on land surface temperature retrieval. *IEEE Geoscience and Remote Sensing Letters*, *1*(2), 127–130. https://doi.org/10.1109/LGRS.2004.824749

Torbett, J. C., Roberts, R. K., Larson, J. A., & English, B. C. (2007). Perceived importance of precision farming technologies in improving phosphorus and potassium efficiency in cotton production.

Precision Agriculture, *8*(3), 127–137. https://doi.org/10.1007/s11119-007-9033-1

Tyralis, H., Papacharalampous, G., & Langousis, A. (2019). A brief review of random forests for water scientists and practitioners and their recent history in water resources. *Water (Switzerland)*, *11*(5). https://doi.org/10.3390/w11050910

Wagner, W., Hahn, S., Kidd, R., Melzer, T., Bartalis, Z., Hasenauer, S., Figa-Saldaña, J., De Rosnay, P., Jann, A., Schneider, S., Komma, J., Kubu, G., Brugger, K., Aubrecht, C., Züger, J., Gangkofner, U., Kienberger, S., Brocca, L., Wang, Y., … Rubel, F. (2013). The ASCAT soil moisture product: A review of its specifications, validation results, and emerging applications. *Meteorologische Zeitschrift*, *22*(1), 5–33. https://doi.org/10.1127/0941-2948/2013/0399

Wan, Z. (2013). *Collection-6 MODIS Land Surface Temperature Products Users' Guide*.

Yue, W., Xu, J., Tan, W., & Xu, L. (2007). The relationship between land surface temperature and NDVI with remote sensing: Application to Shanghai Landsat 7 ETM+ data. *International Journal of Remote Sensing*, *28*(15), 3205–3226. https://doi.org/10.1080/01431160500306906

Zappa, L., Forkel, M., Xaver, A., & Dorigo, W. (2019). Deriving field scale soil moisture from satellite observations and ground measurements in a Hilly Agricultural Region. *Remote Sensing*, *11*(22). https://doi.org/10.3390/rs11222596

Zhang, C., & Kovacs, J. M. (2012). The application of small unmanned aerial systems for precision agriculture: A review. *Precision Agriculture*, *13*(6), 693–712. https://doi.org/10.1007/s11119-012-9274-5

Zhao, W., Sánchez, N., Lu, H., & Li, A. (2018a). A spatial downscaling approach for the SMAP passive surface soil moisture product using random forest regression. *Journal of Hydrology*, *563*(April), 1009–1024. https://doi.org/10.1016/j.jhydrol.2018.06.081

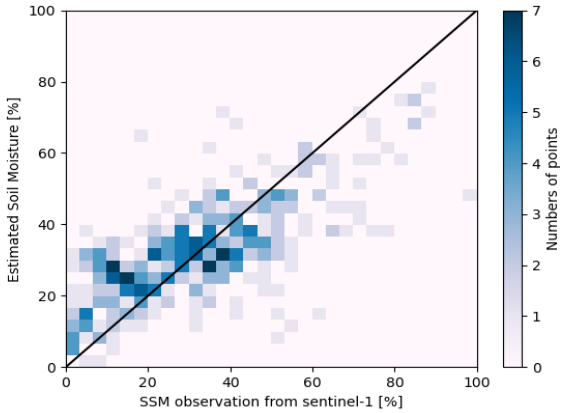
Zhao, W., Sánchez, N., Lu, H., & Li, A. (2018b). A spatial downscaling approach for the SMAP passive surface soil moisture product using random forest regression. *Journal of Hydrology*, *563*(June), 1009–1024. https://doi.org/10.1016/j.jhydrol.2018.06.081

Zhengming, W. (1999). *MODIS Land-Surface Temperature Algorithm Theoretical Basis Document (LST ATBD)*.

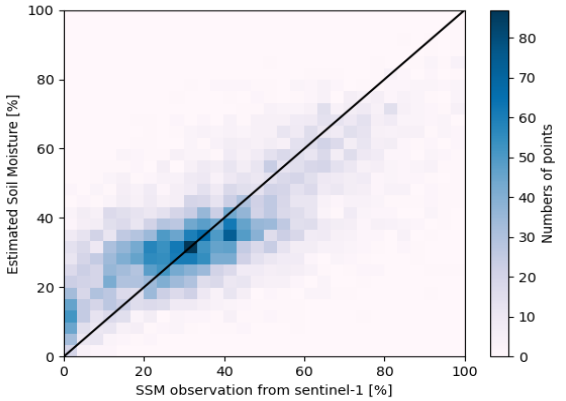
APPENDIX A

The 2D histograms represented the estimated soil moisture using the test sets vs the existing SENTINEL-1 C SAR 1Km SSM on various model input datasets

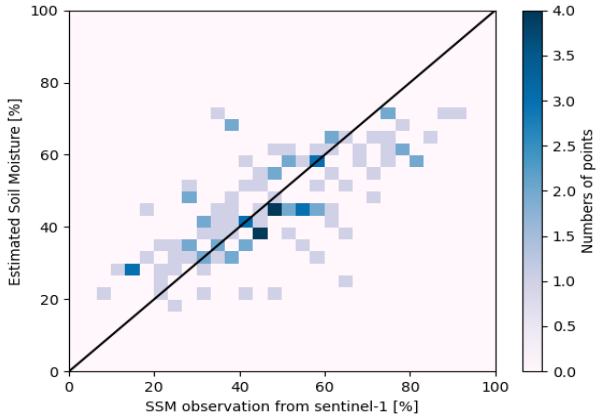
Model run on :2015-2019 at 28kmx28km



Model run on :2015-2019 at 78kmx78km



Model run on :2019 at 28kmx28km



Model run on:2019 at 78kmx78km