UNIVERSITÉ DE SHERBROOKE
Faculté de génie
Département de génie civil et de génie du bâtiment

DÉSAGRÉGATION DE L'HUMIDITÉ DU SOL ISSUE DES PRODUITS SATELLITAIRES MICRO-ONDES PASSIVES ET EXPLORATION DE SON UTILISATION POUR L'AMÉLIORATION DE LA MODÉLISATION ET LA PRÉVISION HYDROLOGIQUE

DOWNSCALING SATELLITE PASSIVE MICROWAVE SOIL MOISTURE AND EXPLORING ITS UTILITY FOR IMPROVING STREAMFLOW SIMULATION AND FORECASTING

Thèse de Doctorat
Spécialité: génie civil

Shimelis Asfaw WAKIGARI

Sherbrooke (Québec), Canada
March 2023

MEMBRES DU JURY

Robert LECONTE
Département de génie civil et de génie du bâtiment, Université de Sherbrooke
Directeur de thèse

Ramata MAGAGI
Département de géomatique appliquée, Université de Sherbrooke
Rapporteure, évaluatrice interne

Aaron BERG
Department of Geography, Environment and Geomatics, University of Guelph
Évaluateur externe

Alain PIETRONIRO
Department of Civil Engineering, University of Calgary
Évaluateur externe

Annie POULIN
Département de génie de la construction, École de technologie supérieure
Évaluatrice externe

RÉSUMÉ

De plus en plus de produits satellitaires en micro-ondes passives sont disponibles. Cependant, leur large résolution spatiale (25-50 km) n'en font pas un outil adéquat pour des applications hydrologiques à une échelle locale telles que la modélisation et la prévision hydrologiques. Dans de nombreuses études, une désagrégation d'échelle de l'humidité du sol des produits satellites micro-ondes est faite puis validée avec des mesures in-situ. Toutefois, l'utilisation de ces données issues d'une désagrégation d'échelle n'a pas encore été pleinement étudiée pour des applications en hydrologie. Ainsi, l'objectif de cette thèse est de proposer une méthode de désagrégation d'échelle de l'humidité du sol issue de données satellitaires en micro-ondes passives (Satellite Passive Microwave Active and Passive - SMAP) à différentes résolutions spatiales afin d'évaluer leur apport sur l'amélioration potentielle des modélisations et prévisions hydrologiques. À partir d'un modèle de forêt aléatoire, une désagrégation d'échelle de l'humidité du sol de SMAP l'amène de 36-km de résolution initialement à des produits finaux à 9-, 3- et 1-km de résolution. Les prédicteurs utilisés sont à haute résolution spatiale et de sources différentes telles que Sentinel-1A, MODIS et SRTM. L'humidité du sol issue de cette désagrégation d'échelle est ensuite assimilée dans un modèle hydrologique distribué à base physique pour tenter d'améliorer les sorties de débit. Ces expériences sont menées sur les bassins versants des rivières Susquehanna (de grande taille) et Upper-Susquehanna (en comparaison de petite taille), tous deux situés aux États-Unis. De plus, le modèle assimile aussi des données d'humidité du sol en profondeur issue d'une extrapolation verticale des données SMAP. Par ailleurs, les données d'humidité du sol SMAP et les mesures in-situ sont combinées par la technique de fusion conditionnelle. Ce produit de fusion SMAP/in-situ est assimilé dans le modèle hydrologique pour tenter d'améliorer la prévision hydrologique sur le bassin versant Au Saumon situé au Québec. Les résultats montrent que l'utilisation de l'humidité du sol à fine résolution spatiale issue de la désagrégation d'échelle améliore la représentation de la variabilité spatiale de l'humidité du sol. En effet, le produit à 1- km de résolution fournit plus de détails que les produits à 3- et 9-km ou que le produit SMAP de base à 36-km de résolution. De même, l'utilisation du produit de fusion SMAP/ in-situ améliore la qualité et la représentation spatiale de l'humidité du sol. Sur le bassin versant Susquehanna, la modélisation hydrologique s'améliore avec l'assimilation du produit de désagrégation d'échelle à 9-km, sans avoir recours à des résolutions plus fines. En revanche, sur le bassin versant Upper-Susquehanna, c'est le produit avec la résolution spatiale la plus fine à 1- km qui offre les meilleurs résultats de modélisation hydrologique. L'assimilation de l'humidité du sol en profondeur issue de l'extrapolation verticale des données SMAP n'améliore que peu la qualité du modèle hydrologique. Par contre, l'assimilation du produit de fusion SMAP/in-situ sur le bassin versant Au Saumon améliore la qualité de la prévision du débit, même si celle-ci n'est pas très significative.

Mots-clés : humidité du sol, désagrégation d'échelle, SMAP, forêt aléatoire, fusion, assimilation

ABSTRACT

The availability of satellite passive microwave soil moisture is increasing, yet its spatial resolution (i.e., 25-50 km) is too coarse to use for local scale hydrological applications such as streamflow simulation and forecasting. Many studies have attempted to downscale satellite passive microwave soil moisture products for their validation with in-situ soil moisture measurements. However, their use for hydrological applications has not yet been fully explored. Thus, the objective of this thesis is to downscale the satellite passive microwave soil moisture (i.e., Satellite Microwave Active and Passive - SMAP) to a range of spatial resolutions and explore its value in improving streamflow simulation and forecasting. The random forest machine learning technique was used to downscale the SMAP soil moisture from 36-km to 9-, 3- and 1-km spatial resolutions. A combination of host of high-resolution predictors derived from different sources including Sentinel-1A, MODIS and SRTM were used for downscaling. The downscaled SMAP soil moisture was then assimilated into a physically-based distributed hydrological model for improving streamflow simulation for Susquehanna (larger in size) and Upper Susquehanna (relatively smaller in size) watersheds, located in the United States. In addition, the vertically extrapolated SMAP soil moisture was assimilated into the model. On the other hand, the SMAP and in-situ soil moisture were merged using the conditional merging technique and the merged SMAP/in-situ soil moisture was then assimilated into the model to improve streamflow forecast over the au Saumon watershed. The results show that the downscaling improved the spatial variability of soil moisture. Indeed, the 1-km downscaled SMAP soil moisture presented a higher spatial detail of soil moisture than the 3-, 9- or original resolution (36-km) SMAP product. Similarly, the merging of SMAP and in-situ soil moisture improved the accuracy as well as spatial representation soil moisture. Interestingly, the assimilation of the 9-km downscaled SMAP soil moisture significantly improved the accuracy of streamflow simulation for the Susquehanna watershed without the need of going to higher spatial resolution, whereas for the Upper Susquehanna watershed the 1-km downscaled SMAP showed better results than the coarser resolutions. The assimilation of vertically extrapolated SMAP soil moisture only slightly further improved the accuracy of the streamflow simulation. On the other hand, the assimilation of merged SMAP/in-situ soil moisture for the au Saumon watershed improved the accuracy of streamflow forecast, yet the improvement was not that significant. Overall, this study demonstrated the potential of satellite passive microwave soil moisture for streamflow simulation and forecasting.

Keywords: soil moisture; downscaling; SMAP; Random Forest; merging; assimilation

REMERCIEMENTS

First and foremost, I would like to thank God for giving me the strength to complete this thesis and for his protection throughout these years.

Next, I would like to thank my thesis director Professor Robert Leconte for believing in me from the very beginning and for his continuous support during the past four years. I am indebted to his exemplary guidance and valuable feedback. Without his patient instruction, persistent encouragement and insightful criticism, the completion of this thesis would not have been possible.

I would also like to thank my thesis jury members: Professor Ramata Magagi, Professor Annie Poulin, Professor Alain Pietroniro and Professor Aaron Berg for their dedicated time in reading my thesis and for giving me constructive comments.

I also wish to acknowledge the Universite de Sherbrooke and the National Science and Engineering Research Council of Canada (NSERC) and the three industrial partners participating in the NSERC Industrial Research Chair (IRC): Hydro-Québec, Brookfield and Hydro-Sherbrooke

In addition, I also wish to thank the current and previous members of our industrial research chair including Mélanie Trudel, Didier Haguma, Yannick Duguay, Jean Bergeron, Khaled Mohammed, Renaud Jougla, Dipti Tiwari, Charles Whittaker, Sepehr Farhoodi, Senda Kouki, Vincent Tremblay and Isabelle Gallant for their valuable comments and ideas, and for all the memories.

I would like to extend a special thanks to Hassan Rouhani for his assistance when I came to Sherbrooke in 2018. I am also grateful to Renaud Jougla for his help in translation of some parts of this thesis into French.

Finally, my deepest gratitude goes to my parents (Asfaw Wakigari and Kore Bekele) for their unconditional love, prayer, and support throughout my life. Many thanks go to my siblings (Tsion, Jabesa and Minase) for their encouragement and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

Acronyms	Definition
AMSR-E	Advanced Microwave Scanning Radiometer
ASCAT	Advanced Scatterometer
BVC	Bilan Vertical à 3 Couches
CCI	Climate Change Initiative
CWSI	Crop Water Stress Index
ERS	European Remote Sensing satellite
ESA	European Space Agency
EVI	Enhanced Vegetation Index
GHz	Gigahertz
KGE	Kling-Gupta Efficiency
LST	Land Surface Temperature
ML	Machine Learning
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NSE	Nash-Sutcliffe Efficiency
RCM	RADARSAT Constellation Mission
RF	Random Forest
RMSE	Root Mean Square Error
SAR	Synthetic Aperture Radar
SMAP	Soil Moisture Active and Passive
SMOS	Soil Moisture and Ocean Salinity
ubRMSE	Unbiased Root Mean Square Error

Chapter 1 INTRODUCTION

1.1 Context and problematic

Soil moisture is an essential hydrological and climate variable which plays an important role in controlling the exchange of water and energy cycles between the land surface and the atmosphere (Entekhabi, 1995). In a water cycle, for example, it conditions the rainfall-runoff response of the watershed to a given precipitation event by partitioning into surface runoff and infiltration. This is particularly important in the case of the saturation excess runoff generation mechanism (Dunne et al., 1975), where the magnitude of the flood is determined by the pre-storm soil moisture state of the watershed. On the other hand, in an energy cycle, it redistributes the incoming net solar radiation into sensible and latent heat fluxes, thereby regulating near-surface variables such as temperature and humidity. In addition, soil moisture is also a key determinant of the global carbon cycle (Falloon et al., 2011) and most importantly, a number of hydrometeorological and agricultural applications such as hydrological forecasting, weather forecasting and irrigation planning are strongly dependent on soil moisture (Brocca et al., 2017; Seneviratne et al., 2010).

Soil moisture is highly variable both spatially and temporally because of its multifaced interactions with numerous environmental factors such as soil properties, topography, land cover and meteorological forcing (Crow et al., 2012). Thus, characterization of its spatial and temporal variability is important. Conventionally, in-situ measurements are used to estimate soil moisture. The common example is the gravimetric method (Reynolds, 1970), which is the oldest and the most well-known method for accurate estimation of soil moisture. It is usually used as the standard reference to validate soil moisture estimated by other methods, e.g., probes or model or remote sensing. However, the major downside of this method is that it is time consuming to implement over a large area and its destructive nature limits its usage at the same location or on the same soil sample.

Apart from gravimetric method, over the past few decades a plethora of soil moisture probe sensors have emerged for the purpose of soil moisture estimation. In contrast to the gravimetric method, these sensors indirectly estimate soil moisture from surrogate variables such as soil moisture potential or from the radioactive and electrical properties of the soil such as dielectric constant, capacitance or impedance (Lakshmi et al., 2014). A typical example of these sensors includes neutron probes, Time Domain Reflectometry (TDR), Frequency Domain Reflectometry (FRD), Amplitude Domain Reflectometry (ADR) and many others. Compared to the gravimetric method, probes allow continuous measurement of soil moisture (i.e., at a temporal resolution of minutes) at the same location on the same soil sample, yet they can be time consuming, labor-intensive and difficult to maintain if intended to establish a network over a large area, e.g., at watershed scale.

Overall, in-situ measurements provide an accurate estimate of soil moisture, but only at point scale and this limits their use for applications at watershed scale. For this reason, over the last few decades, remote sensing has become an attractive and feasible alternative for soil moisture estimation over a large area (Brown et al., 2013). Accordingly, the optical/thermal and microwave domains of electromagnetic spectrum are regions of greatest interest for soil moisture retrieval (Schmugge, 1976).

The potential of the optical/thermal remote sensing in the estimation of soil moisture has been demonstrated in many studies (Lobell & Asner, 2002; Wang et al., 2018; Weidong et al., 2002). They have the capability of providing soil moisture at high-spatial (e.g., in the order of hundreds of meters) and temporal resolutions (e.g., daily). However, they can be greatly affected by cloud contamination and vegetation interferences restricting their ability to generate both spatially and temporally continuous soil moisture maps. In addition, the shallow penetration depth (i.e., in the order of millimeters) of the soil is a major limitation for their use for operational applications. Because of these limitations, the interest in the use of the optical/thermal remote sensing for soil moisture retrieval is decreasing, notably since satellite passive microwave remote sensing has become available operationally (e.g., SMAP, SMOS, and AMSR-E).

Microwave remote sensing, notably at lower frequencies (e.g., 1-3 GHz), has become more appealing for soil moisture retrieval compared to optical/thermal remote sensing. This is primarily because of its sensitivity to soil moisture under all weather conditions and its penetration through sparsely to moderately vegetated canopies (Njoku & Entekhabi, 1996). It can be divided into active (radar and scatterometer) and passive (radiometer) remote sensing depending on their mode of operation. Active microwave remote sensing transmits its own source of energy to illuminate the Earth surface and measure portion of the transmitted energy that is redirected back towards the sensor, which is also known as backscatter coefficient. It can be further divided into imaging (e.g., Synthetic Aperture Radar (SAR)) and non-imaging (altimeters and scatterometers) sensors.

Both spaceborne imaging and non-imaging sensors have demonstrated potential for near-surface (i.e. few centimeters) soil moisture retrieval (Barrett et al., 2009). However, the non-imaging sensors (e.g., Advanced scatterometer (ASCAT)) are not primarily designed for soil moisture retrieval, but rather for wind speed and direction measurements over the ocean surface.

On the other hand, spaceborne imaging sensors (e.g., synthetic aperture radar (SAR)) have proven effective and widely used for retrieval of soil moisture at high-spatial resolution than non-imaging sensors (Abowarda et al., 2021; Paloscia et al., 2013). For example, Sentinel-1 with its C-band SAR is well known for its capability to provide soil moisture at high-spatial resolution (i.e., as high as tens of meters), but with temporal resolution of 6 to 12 days. This is primarily because of the smaller swath width of Sentinel-1 SAR, which increases the time it takes to cover the entire globe and subsequently increases the revisit time. On the other hand, the recently launched Canadian RADARSAT Constellation Mission (RCM) has the potential for near-surface soil moisture retrieval with temporal resolution of 4 days.

The principal basis for the spaceborne imaging active microwave remote sensing relies on the sensitivity of the backscatter signal to the dielectric properties of soil (Lin et al., 1994). However, the sensitivity of the backscatter signal to the soil moisture is affected by perturbing factors such as vegetation and surface roughness. Thus, reducing the impact of these factors on soil moisture retrieval is essential. Many studies have suggested the use of optimum configuration of sensor parameters (e.g., polarization, frequency, incidence angle) to reduce these impacts (Barrett et al., 2009; Petropoulos et al., 2015). For example, the combination of low incidence angle, low frequency (e.g., L-band) and HH\HV polarizations are suitable sensor configuration for optimal soil moisture retrieval by reducing the impact of vegetation and surface roughness.

An alternative to active microwave remote sensing is passive microwave remote sensing (i.e., radiometers), which have great potential for soil moisture retrieval. It measures the naturally emitted radiation from the land surface, which can be expressed as brightness temperature (Jackson & Schmugge, 1989). Similar to the active microwave remote sensing, they are of great interest for near-surface soil moisture retrieval because of their high sensitivity to variation of dielectric constant of the soil owing to soil moisture dynamics, notably at a lower frequency such as the L-band (~1.4 GHz) (Njoku & Entekhabi, 1996).

At the L-band frequency the effect of vegetation cover and surface roughness on the emitted radiation is reduced. However, at this frequency natural emission is very weak. Thus, the sensors must look over a larger area in order to detect this weak emission (Barrett et al., 2009). This, in turn, results in a coarser spatial resolution of satellite passive microwave soil moisture products, i.e., tens of kilometers, which limits them for local scale hydrometeorological applications where high-spatial resolution is required. In addition, penetration depth remains an issue. The L-band sensors have penetration depths that can reach up to 5-10 centimeters depending on vegetation conditions. It is possible to alleviate this issue by extrapolating near-surface soil moisture estimates to deeper soil layers relevant to agricultural and hydrological applications (Albergel et al., 2008).

Over the past decades, a number of passive microwave radiometers with different frequencies (e.g., L, C and X) onboard various satellite platforms have shown their potential for near-surface soil moisture mapping over a large area. For example, Advanced Microwave Scanning Radiometer (AMSR-E/2) at X and C-bands, Soil Moisture and Ocean Salinity (SMOS) and Soil Moisture Active Passive (SMAP) at L-band are widely used for soil moisture retrieval. Interestingly, the L-band passive microwave satellites such as SMOS and SMAP have received great attention for potential applications because of their prime devotion to global monitoring of surface soil moisture (Entekhabi et al., 2008; Kerr et al., 2001).

The SMOS is the first ever L-band passive microwave mission dedicated to monitoring of soil moisture and sea surface salinity at a global scale. It was launched by the European Space Agency (ESA) in 2009 and provides a global map of near-surface soil moisture every 2-3 days with resolution of about 40-km (Kerr et al., 2001). Similarly, the National Aeronautics and Space Administration (NASA) launched the SMAP on January 31, 2015, which carries both L-band radar and radiometer sharing a single platform, but the radar ceased operation on July 7, 2015 (Colliander et al., 2017). However, the radiometer component continues the generation of soil moisture maps at roughly 40-km spatial resolution.

Although substantial progress has been made in the estimation of soil moisture over a large area using passive microwave remote sensing during the last decades, there is still a long way to go in terms of soil moisture estimation at spatial and temporal resolutions desired by many hydrometeorological and agricultural applications, which is in the order of a few kilometers or less with decent temporal resolution. Putting it differently, the spatial resolution of the existing passive microwave soil moisture products is 25 to 50-km and does not commensurate with the resolution needed by many hydrometeorological applications (Figure 1.1).

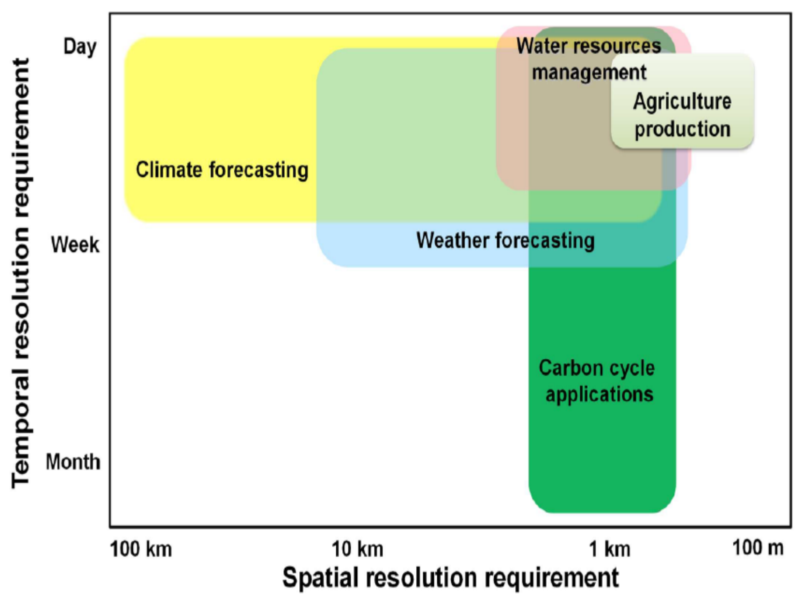

Figure 1.1 Spatial and temporal resolution requirements of soil moisture for a range of applications (Adapted from Sabaghy et al. (2018))

Accordingly to answer the ever-growing need for the high-spatial resolution of soil moisture, considerable efforts have been devoted to produce soil moisture maps at high-spatial resolution either through direct retrieval (Balenzano et al., 2021; Hajj et al., 2017) or through spatial downscaling (Abbaszadeh et al., 2019; Abowarda et al., 2021; Zhao et al., 2018).

A direct retrieval entails retrieval of high-spatial resolution soil moisture from synthetic aperture radars such as C-band SARs onboard Sentinel-1, RADARSAT-1/2, or Canadian RADARSAT Constellation Mission (RCM). Many studies have demonstrated the potential of these radars in mapping near-surface soil moisture at high-spatial resolution (i.e., in the order of tens of meters). For example, Ma et al. (2020) retrieved soil moisture at 100 m spatial resolution from combination of Sentinel-1 and 2 for precision agriculture. Similarly, Balenzano et al. (2021)

retrieved soil moisture at 1-km spatial resolution from Sentinel-1A over Yanco (New South Wales, Australia) and Little Washita (Oklahoma, USA) watersheds. Apart from Sentinel-1 SAR, notable progress have been made in soil moisture retrieval from other SAR sensors such as RADARSAT-2 (Chen et al., 2021; Xing et al., 2019) and ENVISAT ASAR (Santi et al., 2013). However, here it is worthwhile to recall that although current satellite C-band SAR sensors have spatial resolution suitable for watershed-scale studies, soil moisture retrieval is complicated by environmental conditions such as surface roughness and vegetation cover, which limits its use for operational applications. This suggests looking for an alternative approach which gives high spatial resolution soil moisture of good quality with decent temporal resolution and global coverage. Thus, an attractive alternative could be spatial downscaling applied to low resolution passive microwave soil moisture products (Peng et al., 2017; Sabaghy et al., 2018).

Spatial downscaling (also known as spatial disaggregation) is a technique that allows downscaling of the coarse-spatial resolution satellite passive microwave soil moisture products by synergistically using high-resolution (i.e., a kilometer or less) land surface variables or backscatter coefficient which have strong association with soil moisture such as land surface temperature and vegetation indices.

There exist a variety of spatial downscaling techniques, which can roughly be grouped as optical/thermal-based downscaling, radar-based downscaling and radiometer-based downscaling, based on the sources from which the high-resolution variables, indices or signals are derived for downscaling (Sabaghy et al., 2018). In addition, more recently, machine learning (ML) technique has received a great deal of attention (Zhao et al., 2018).

The optical/thermal-based downscaling relies on high-resolution land surface variables and indices derived from the optical\thermal remote sensing (e.g., MODIS) for downscaling of the satellite passive microwave soil moisture products. This method successfully demonstrated its capability in downscaling soil moisture products derived from, e.g., AMSR-E (Fang et al., 2013; Zhao & Li, 2013), SMOS (Djamai et al., 2015; Piles et al., 2011) and SMAP (Wen et al., 2020). It is a widely used method because of its simplicity, yet cloud and vegetation adversely affect the quality as well as the spatio-temporally continuity of the downscaled soil moisture by this method. In addition, its application is mostly limited to areas with uniform atmospheric condition (Chauhan et al., 2003; Piles et al., 2011).

Radar-based downscaling uses high-resolution SAR observations to downscale the satellite passive microwave soil moisture products. A number of studies have developed and tested different radar based downscaling algorithms, e.g., baseline downscaling technique (Das et al., 2011), the change detection method (Piles et al., 2009) and the Bayesian framework (Zhan et al., 2006). It has the potential to downscale satellite passive microwave soil moisture products and Tb over low to moderately vegetated areas when compared to the optical\thermal-based downscaling (He et al., 2018; Wu et al., 2016). This is because of the capability of SAR sensors to penetrate through vegetation to some extent, especially at lower frequencies. In addition, there is no effect of clouds on the radar-based downscaling as it operates under-all-weather conditions, but the generation of temporally continuous soil moisture is limited by the longer revisit time of SAR observations.

Radiometric-based downscaling uses high resolution land surface variables derived from radiometric sensor with high frequency for downscaling of satellite passive microwave soil moisture, which is mainly obtained from low frequency sensor. This downscaling technique is mainly convenient for the sensor with multi-frequency measurements such as AMSR-E with a range of frequencies (e.g., 6,925, 10,65, 18,7, 23,8, 36,5, and 89,0 GHz). Observation at high frequency (e.g., 89,0 GHz) can be used for downscaling of observation at lower frequency (e.g., 6.92GHz). Typical examples of radiometric-based downscaling include Smoothing Filter-based Intensity Modulation (SFIM) (Liu et al., 2011).

On the other hand, ML is becoming an appealing alternative for the downscaling of satellite passive microwave soil moisture products because of its capability to handle the complex non-linear relationship between soil moisture and varieties land surface variables, which has strong associations with soil moisture (Zhao et al., 2018). In addition, its effectiveness in handling a large dataset makes it more attractive compared to optical\thermal and radar-based downscaling. Indeed, it helps to consider a set of land surface and atmospheric variables that affects the spatio-temporal variability of soil moisture in the downscaling, in contrast to the optical\thermal and radar-based downscaling which are often limited to few variables. For example, the optical\thermal-based downscaling mainly relies only on the land surface temperature and vegetation indices, whereas the radar-based downscaling mainly relies only on the backscatter coefficients for downscaling of satellite passive microwave soil moisture.

Many studies have demonstrated the potential of different ML variants in the downscaling of satellite passive microwave soil moisture products (Abbaszadeh et al., 2019; Bai et al., 2019; Hu et al., 2020; Zhao et al., 2018, 2019). However, most of these studies have been conducted in areas favorable (bare to moderately vegetated areas often instrumented with in-situ soil moisture probes) for the downscaling of satellite soil moisture, e.g., Iberian Peninsula (semi-arid climate dominated by croplands, open shrublands and woody savannas) (Zhao et al., 2018), Tibetan plateau (continental climate dominated by grass and barren land) (Chen et al., 2020), and northern China (semi-arid region dominated with dry land and woodland) (Bai et al., 2019).

With regards to variables used in ML for downscaling, the combination of MODIS derived land surface variables such as land surface temperature, albedo, vegetation indices (VIs) and SRTM DEM derivatives such as elevation, slope and aspect are the most widely used for downscaling of satellite passive microwave soil moisture products (Abbaszadeh et al., 2019; Wen et al., 2019; Zhao et al., 2018). However, the use of SAR derived variables is still at its early stage for ML-based downscaling despite of the increase in the SAR observations.

On the other hand, over the last decades, many studies have assimilated satellite soil moisture products into different land surface and hydrological models with the aim of: 1) improving soil moisture estimation (Crow et al., 2017a; Lievens et al., 2017; Reichle, 2008), and 2) improving streamflow simulation and forecasting (Abbaszadeh et al., 2020a; Azimi et al., 2020a; Chen et al., 2011; Crow et al., 2017; Massari et al., 2015; Meng et al., 2017; Wanders et al., 2014). However, the latter have reported contrasting results. While some studies have found significant improvement in the simulation and forecasting of streamflow, a considerable number of other studies have shown little or no impact of satellite soil moisture on streamflow simulation and forecasting.

As an example, Jadidoleslam et al. (2021) found significant improvement in streamflow simulation after assimilation of the coarse-spatial resolution SMOS and SMAP soil moisture into a distributed hydrological model. Similarly, Azimi et al. (2020) reported marked improvement in discharge simulation following assimilation of Sentinel-1 and SMAP soil moisture into hydrological model. Conversely, Kumar et al. (2014) reported slight improvement in streamflow simulation after assimilation of soil moisture retrievals from AMSR-E. Lievens et al. (2015) found improvement only in peak flows after assimilation of SMOS soil moisture, whereas Parajka et al. (2006) reported no improvement at all in streamflow simulation after assimilation of soil moisture retrieved from ERS scatterometer.

The disagreement among these studies could be attributed to the complexity of hydrological model used (conceptual or physically-based distributed models), the assimilation techniques adopted (i.e., direct insertion, ensemble Kalman filter, or particle filter), characteristics of the watershed (e.g., location, size, land cover, soil type, and dominant hydrological processes), the vertical coupling strength between surface and subsurface physics of the model and the attributes of remotely sensed soil moisture data used such as spatial and temporal resolution, penetration depth to the soil and retrieval algorithm and its accuracy (Mao et al., 2019).

1.2 Statement of the problem and research question

Reconciling the spatial scale disparity between the passive microwave soil moisture products and local scale hydrometeorological and agricultural applications requirement is crucial. Considerable efforts have been made on this regard to downscale satellite passive microwave soil moisture products using high-resolution land surface variables and indices derived from the optical\thermal sensors. However, only a few studies have examined the utility of SAR observations such as backscatter intensity and textural information in the downscaling of satellite soil moisture products, notably using ML techniques (Bai et al., 2019; Karami et al., 2022).

Thus, in the light of increasing availability of SAR observation from the existing (e.g., Sentinel-1, Canadian RADARSAT constellation mission (RCM)) and upcoming SAR (e.g., NASA-ISRO SAR mission (NISAR)) sensors and the great capability of ML techniques in handling large datasets, exploring the utility of SAR observation for downscaling satellite passive microwave soil moisture along with the optical sensors derived variables is important.

On the other hand, the success of the downscaling of satellite passive microwave soil moisture products inherently depends on land surface and climate conditions of the selected study area. Interestingly, previous studies have applied the downscaling techniques mostly over bare to less vegetated areas and these techniques are site-specific and local in nature. Hence, further studies are needed by extending the downscaling techniques to different geographic regions with different land surface and hydro-climatic conditions, e.g., mixed to moderately forested areas with humid climate.

It is also important to note that the majority of previous studies have downscaled satellite passive microwave soil moisture with the objective to reduce the spatial scale disparity between the satellite soil moisture products (i.e., 25- to 50-km resolution) and in-situ soil moisture measurements (i.e., in the order of few centimeters) often for purpose of validation of satellite soil moisture. However, subsequent use of downscaled satellite passive microwave soil moisture product for hydrometeorological applications is still at its early stage.

Previous studies have focused on the assimilation of the original resolution (i.e., roughly 40-km) satellite passive microwave soil moisture products often for improving soil moisture estimation, whereas the studies that have focused on the assimilation of satellite soil moisture into hydrological models for improving the streamflow simulation and forecasting are relatively few (Abbaszadeh et al., 2020; Le et al., 2022). Ironically, they come-up with contrasting results. These contrasting results suggest the need for further investigations to clearly determine the impact of assimilation of the original and downscaled satellite passive microwave soil moisture on the streamflow simulation and forecasting under a broad range of land surface and hydro-climatic conditions. Especially, as stated before, the impact of assimilation of downscaled satellite passive microwave soil moisture on the model prediction has not yet been thoroughly explored.

Hence, based on the above insights, the main research question of this thesis is: Will the downscaling and subsequent assimilation of downscaled satellite passive microwave soil moisture to different spatial resolutions (i.e., 1, 3 and 9-km) into a distributed hydrological model improve the streamflow simulation and forecasting?

1.3 Objectives

Given the available low-resolution satellite passive microwave soil moisture products and a growing need for their incorporation into hydrological models, the main objective of this thesis is to downscale the coarse-spatial resolution SMAP soil moisture products into a range of spatial resolutions and subsequently assimilate into a physically-based distributed hydrological model for the purpose of improving streamflow simulation and forecasting. It is accompanied by three specific objectives:

1) To enhance the spatial resolution of the SMAP soil moisture products through spatial downscaling;
2) To explore the utility of spatially downscaled SMAP soil moisture products in improving streamflow simulation and forecast skills;
3) To evaluate the value of merged SMAP/in-situ soil moisture in improving streamflow forecast skills.

1.4 Original contributions

The originality of this thesis primarily lies in the assimilation of the downscaled SMAP soil moisture into a hydrological model. Hitherto, many studies have assimilated the original resolution SMAP soil moisture into different hydrological models and similarly other studies have made significant efforts in the downscaling of the SMAP soil moisture. Nonetheless, regardless of the many downscaling efforts, subsequent assimilation of the downscaled SMAP soil moisture products into hydrological model for improving its prediction skills have not yet been thoroughly explored. Thus, distinctively from previous studies, this thesis focuses on assimilation of downscaled SMAP soil moisture to a range of spatial resolutions (e.g., 1-, 3-, and 9-km) into a physically-based distributed hydrological model for improving streamflow simulation and forecasting.

The originality of this thesis also lies in the downscaling of SMAP soil moisture products. This study uses high-resolution land surface variables derived from the Sentinel-1 SAR along with those derived from the optical sensors of MODIS for downscaling of the SMAP soil moisture using ML technique as opposed to most of the previous downscaling studies which mainly relies only on land surface variables and indices derived from the optical sensors.

1.5 Thesis structure

This thesis is composed of seven chapters. The first chapter (**Chapter 1**) introduces the background and objectives of the thesis. **Chapter 2** presents an extensive review of literature pertaining to soil moisture importance and measurements, satellite soil moisture downscaling techniques and the role of soil moisture in improving hydrological model predictions. **Chapter**

3 describes the study area watersheds, datasets characteristics, and the downscaling method and framework used in this study. The downscaling of the SMAP soil moisture over the study area is presented in **Chapter 4.** A set of downscaling experiments conducted is presented in this chapter. This work has been published in the Remote Sensing journal. **Chapter 5** presents the second article, submitted to Journal of Hydrology: Regional Studies, which focuses on the assimilation of downscaled and native resolution SMAP soil moisture into hydrological model. In this chapter, a suite of experiments and procedure of implementations is presented. **Chapter 6** presents the third article, submitted to Hydrology journal (MDPI), which deals with merging of the SMAP and in-situ soil moisture for taking advantage of both data sources. The merged product is then used for updating the hydrological model to improve the streamflow forecast. Conclusions of the thesis and recommendations for future directions are discussed in **Chapter 7.**

Chapter 2 LITERATURE REVIEW

This chapter outlines the previous research studies devoted to soil moisture measurements and their applications. It begins with the general overview of the importance of soil moisture in the Earth system, followed by the description of soil moisture measurement techniques and the downscaling methods commonly used for enhancing the spatial resolution of remotely sensed soil moisture products. In the end, the benefit of remotely sensed soil moisture products for hydrological applications is presented.

2.1 Soil moisture importance and measurement

Soil moisture represents the water held in the pore space of the soil in a shallow layer of the Earth's upper surface (i.e., roughly up to the top 200 cm). It is very small in quantity compared to the total available global fresh water, yet plays a paramount role in the Earth system by controlling the exchange of water, energy and biogeochemical fluxes between the land surface and the atmosphere (Entekhabi, 1995; Seneviratne et al., 2010a). In the terrestrial water cycle, soil moisture regulates the fractionation of the incoming precipitation into surface runoff and infiltration, thereby modulating the magnitude of streamflow and ground water recharge. Similarly, soil moisture controls the partitioning of the incoming radiation energy into sensible and latent heat fluxes. In addition, a wide range of operational hydrometeorological and agricultural applications including weather forecasting, hydrological forecasting, crop yield forecasting and irrigation scheduling is heavily dependent on the soil moisture (Brocca et al., 2016; Peng et al., 2021).

Soil moisture is highly variable both spatially and temporally depending on the soil properties (e.g., texture and structure), surface topography, land cover and atmospheric forcing (Crow et al., 2012). Hence, characterization of its spatial and temporal variability has fundamental importance to better comprehend its role in the Earth's hydrological, energy and biogeochemical cycles and as well as in operational hydrological and agricultural applications. There are three common soil moisture estimation methods, which help in the characterization of the spatio-temporal variability of soil moisture. These are:1) in-situ measurements, 2) hydrological and land surface models and 3) remote sensing.

2.1.1 In-situ soil moisture measurements

The in-situ measurements provide soil moisture estimate at a point scale with a high-temporal resolution across the vertical soil profile. They generally can be divided into two broad categories: direct and indirect approaches. The direct approach involves direct estimation of soil moisture using the gravimetric method, which is the oldest and most reliable method. The soil moisture estimation by this method is based on weighting the difference of a soil sample before and after oven dried at approximately 105 ^{0}C for 24 hours (Dobriyal et al., 2012; Petropoulos et al., 2015). In contrast, the indirect approach measures soil moisture surrogate variables such as soil permittivity and capacitance instead of soil moisture content. These variables are then related to volumetric soil moisture content based on a predetermined physical or empirical relationships called calibration curves (Dobriyal et al., 2012; Vereecken et al., 2008). Many of the examples of indirect methods are probe based such as neutron probes, time domain reflectometry (TDR), frequency domain reflectometry (FDR) and tensiometer.

Compared to the indirect approach, the direct approach is economical, accurate, and often used as a standard to which the indirect approach is calibrated. However, it is laborious and time-consuming since it involves series of tasks including taking samples from the field by auger, transporting to the laboratory, weighting the sample and then waiting for several hours after putting the sample in the oven until all the moisture is removed through evaporation. In addition, its destructive nature makes it impossible to frequently measure soil moisture at the same location or on the same soil sample (Zazueta & Xin, 1994). On the other hand, the indirect approach has advantages such as high frequency of sampling because of the shorter response time of the sensors. In addition, it is not destructive and repetitive measurements on the same location or sample is possible (Muñoz-Carpena, 1969). However, this approach can be time consuming and costly to implement over a larger area.

During the last decades, many efforts have been made to establish in-situ soil moisture networks in different parts of the globe, for example, OZNET in Australia (Young et al., 2008), REMEDHUS in Europe (Martínez-Fernández & Ceballos, 2005), USDA-ARS in the United States (Jackson et al., 2010), McMaster Mesonet in Ontario, Canada (Kornelsen & Coulibaly, 2013), BERMS and Kenaston sites, Saskatchewan, Canada (Magagi et al., 2013) and RISMA in Manitoba, Canada (Ojo et al., 2015), to capture the spatial and temporal variability of soil moisture over a large area. However, most of these networks lack spatial density to accurately characterize the spatial variability of soil moisture at a scale required for hydrological applications (Dorigo et al., 2011), and establishing such networks over a larger area, for example at the watershed scale, is deemed costly and time consuming.

Overall, in-situ measurements offer an accurate estimate of soil moisture and still they are the backbone for the validation of model and remote sensing-based soil moisture data and for local hydrological and agricultural studies and applications. However, it is still challenging to measure the spatial and temporal distribution of soil moisture at regional and global scale using ground-based instrumentation. For this reason, over the last decades, a number of model and satellite remote sensing-based soil moisture estimation approaches have become increasingly available.

2.1.2 Hydrological and land surface models

Hydrological and land surface models offer an advantage of soil moisture estimation at different layers of soil profile across a large area ranging from the watershed scale to the continental scale with daily or sub-daily temporal resolution. Hydrological models, notably the physically-based distributed models, provide temporally continuous and spatially complete soil moisture at different layers based on water budgets. Typical examples of these models include HYDROTEL (Fortin et al., 2001), SWAT (Arnold et al., 2012) and WRF-Hydro (William et al., 2017). Alternatively, Land Surface Models (LSM) estimate the spatial distribution of soil moisture on regular grids on the basis of water and energy balance principles. Examples of these models include the Canadian Land Surface Scheme (CLASS) (Verseghy, 1991), the Interaction Soil Biosphere Atmosphere (ISBA) (Noilhan & Mahfouf, 1996) and the Biosphere-Atmosphere Transfer Scheme (BATS) (Dickinson et al., 1993).

Although models are a viable alternative source of soil moisture, they are heavily dependent on model parameterization, the representation of soil hydraulic properties in the model and the quality of meteorological forcing. The representation of soil hydraulic properties involves taking into account horizontal and vertical variation of soil properties (e.g., texture and structure) and representing the easiness with which water moves through the soil. This in turn affects the quality of simulated soil moisture and therefore, their sole use notably for operational applications should be based on caution. A number of studies have attempted to improve soil moisture estimated by model through merging with satellite soil moisture observation based on data assimilation techniques (Crow et al., 2017b; Lievens et al., 2017; Reichle, 2008). For example, Lievens et al. (2017) found improvement in soil moisture estimation after joint assimilation of the Sentinel-1 and SMAP soil moisture into the NASA Catchment Land Surface Model. Similarly, Renzullo et al (2014) assimilated AMSR-E and ASCAT-derived soil moisture products into the landscape model of the Australian Water Resources Assessment system (AWRA-L) to improved root-zone moisture estimation. Their results indicated improvement in the root-zone soil moisture estimates.

2.1.3 Remote sensing

Remote sensing is becoming the most viable technique for soil moisture estimation due to advantages such as a wider spatial coverage with decent temporal resolution (i.e., 2 to 3 days especially for the L-band passive microwave satellites such as SMAP). Remote sensing is generally based on measuring of emitted and/or reflected electromagnetic radiation from the terrestrial surface/near-surface (Palace et al., 2008). The different types of emitted and/or reflected energy forms electromagnetic spectrum, which ranges from the shorter wavelength (e.g., gamma and x-rays) to longer wavelength (e.g., microwave and radio wave)

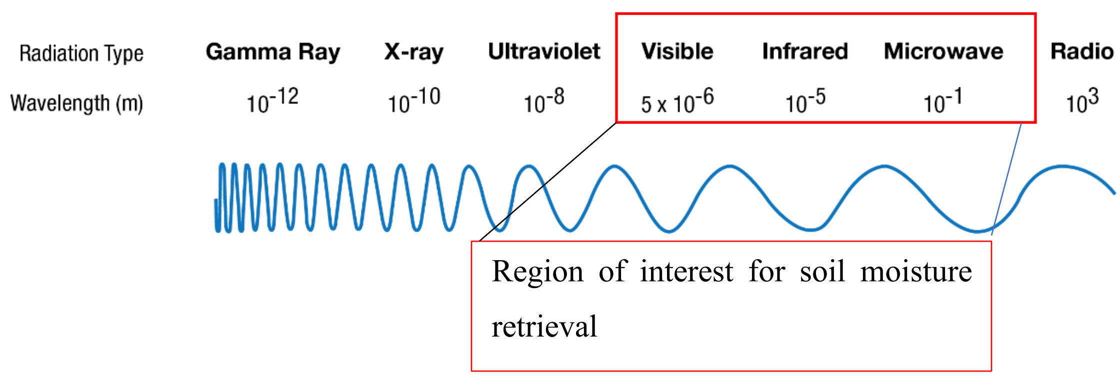

Figure 2.1 Electromagnetic spectrum range suitable for soil moisture retrieval (adapted from mynasadata. https://mynasadata.larc.nasa.gov/)

The optical (i.e., visible and near infrared), thermal (e.g., middle-wave infrared (MWIR) and long-wave infrared (LWIR)) and microwave are regions of electromagnetic spectrum with a demonstrated success in soil moisture retrieval (Figure 2.1). Notably, the microwave region of the spectrum at lower frequency is a region of greatest interest for soil moisture retrieval (Njoku

& Entekhabi, 1996; Schmugge et al., 1986). A brief overview of their theoretical background, advantages and disadvantages is presented in the following subsections, yet their comprehensive reviews can be found in Babaeian et al. (2019), Li et al. (2021) and Petropoulos et al. (2015).

Optical and thermal-infrared remote sensing

The optical remote sensing measures the reflectance from the Earth surface from which soil moisture can be estimated based on empirical relationships. Many studies have come up with contrasting results regarding the relationship between soil reflectance and soil moisture. While most of these studies have highlighted the linear relationship between soil moisture and reflectance, some other studies demonstrated nonlinear relationships. For instance, Lobell and Asner (2002) found a negative linear relationship between reflectance and soil moisture over bare soil, whereas Weidong et al. (2002) found a varying relationship depending on soil moisture level, notably for typical agricultural conditions. For example, for low soil moisture level they found negative relationship, whereas for higher soil moisture level after certain critical point which ranges between 0,15 to 0,40 g/cm^3 they found positive relationship between soil moisture and the reflectance. Moreover, Nocita et al. (2013) found a nonlinear relationship between the two variables. In the end, it is worth mentioning that most of these studies were conducted on bare soil.

On the other hand, over vegetated areas, the direct estimation of soil moisture is not possible as the observed reflectance describes the change in color and canopy water content of vegetation instead of the soil moisture variation. Therefore, soil moisture is indirectly estimated from empirical spectral vegetation indices such as Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1974), Normalized Difference Water Index (NDWI) (Gao, 1996) and Crop Water Stress Index (CWSI) (Xiao et al., 1994), which are basically derived from a combination of the spectral reflectance of two or more bands of optical remote sensing satellite sensors such as MODIS.

Apart from optical remote sensing, thermal-infrared remote sensing has been widely used for estimation of soil moisture based on the thermal properties of soil. It measures the thermal emission of the Earth surface from which the land surface temperature can be estimated. The land surface temperature can then be used to estimate soil moisture either solely using the thermal inertia method or in combination with vegetation indices using the temperature/vegetation index method (Wang & Qu, 2009; Zhao & Li, 2013).

The thermal inertia method is based on empirical relationships between diurnal temperature change and soil moisture (Wang & Qu, 2009), whereas the temperature/vegetation index method is based on a relationship between land surface temperature and vegetation indices (Li et al., 2021). The most well-known example of the temperature/vegetation index method is the universal triangle concept (Carlson et al., 1995; Petropoulos et al., 2009), which is a feature space with a shape of triangle formed from the scatterplot of land surface temperature and vegetation indices. The basis for this method is the sensitivity of soil moisture to land surface temperature over a range of land surface conditions.

Optical and thermal remote sensing have the advantages of high-spatial resolution, relatively mature technology, and availability of multi-bands observations. On the other hand, their main disadvantages include limited capability to penetrate through cloud and vegetation canopy, high attenuation by the atmosphere, shallow penetration depth of the land surface (i.e., in the order of few millimeters), and requirement of in-situ measurements to develop empirical relationships between reflectance and soil moisture (Li et al., 2021; Petropoulos et al., 2015). Moreover, confounding factors such as organic matter content and physical properties of soil could make interpretation of the relationship between soil moisture and soil reflectance more difficult (Bowers & Hanks, 1965; Sharma et al., 2018).

Microwave remote sensing

Microwave (MW) remote sensing measures emitted and reflected electromagnetic radiation from the surface and near-surface of the Earth. It relies on part of the electromagnetic spectrum with frequency ranging from 0,3 to 300 GHz (or 1 mm to 100 cm in terms of wavelength). From this range, the frequency bands designated with letters such as L (1-2 GHz), C (4-8 GHz), and X (8-12 GHz) are the most appealing bands for soil moisture retrieval. This is because of their high sensitivity to variations of the dielectric constant (or relative permittivity) of the soil as a function of the variation of moisture content of the soil. In addition, their capability to penetrate through clouds and to a varying degree vegetation canopy makes them attractive, notably at L-band (Njoku & Entekhabi, 1996). It can be classified into active (radar, scatterometer) and passive (radiometer) depending on their mode of operation.

Active Microwave remote sensing

Active microwave remote sensing sensors relies on their own source of energy to illuminate the Earth surface and measure the backscattered signal (Barrett et al., 2009). They are generally categorized into imaging and non-imaging sensors. Imaging sensors such as synthetic aperture radar (SAR) have the capability to estimate soil moisture with a spatial resolution in the order of meters to tens of meters, whereas non-imaging sensors such as the scatterometers (e.g., ASCAT) are originally designed to measure wind speed and direction, but their potential for soil moisture retrieval has been also proven (Magagi & Kerr, 1997; Wagner et al., 1999).

The backscattered signal from the land surface and near-surface is strongly dependent on the dielectric properties of the soil and other perturbing factors such as surface roughness and vegetation. Because active sensors at X, C and L-band have rather low signal penetration depths, they can only sense surface soil moisture, from 1-3, 3-5 and 5-10 cm depths respectively, depending on if the soil is dry/moist (higher/lower penetration depths). Surface roughness causes the incident energy from the radar to be scattered in many directions and affects the amount that returns to the radar sensor.

On the other hand, vegetation affects the signal by absorbing and scattering part of the incident energy from the radar. It also attenuates the reflected energy from the soil surface. Ideally, removal of the contribution of surface roughness and vegetation from backscattering signal is important so that soil moisture can solely retrieved, but their removal is difficult and complicated (Kolassa et al., 2013). This is because of the difficulties associated with the parameterization of surface roughness and vegetation. Also, dense vegetation may almost completely mask the backscattered signal from the underlying ground surface, thereby making it impossible retrieving soil moisture.

Many studies have suggested the optimum configuration of sensor parameters for the optimal retrieval soil moisture while minimizing the impact of surface roughness and vegetation (Barrett et al., 2009). For instance, Rao et al. (1993) found better estimate of soil moisture using multifrequency measurements instead of single frequency. In addition, Chen et al. (1995) found a reduced effect of surface roughness by making use of co-polarization ratios (HH/VV). Moreover, there is a general agreement from several studies that low incidence angle, low frequency (e.g., L-band), and either HH or VV polarization are the optimal set of sensor parameters for soil moisture retrieval (Barrett et al., 2009).

In addition, currently there is a great deal of interest in the use of polarimetric decompositions for retrieval of soil moisture from synthetic aperture radar sensors (SAR). This approach involves acquisition of SAR imagery using various and multiple polarization configurations and allows separation of contribution of different scattering mechanisms (Hajnsek et al., 2009; Pasolli et al., 2012; Wang et al., 2017). For example, Pasolli et al., (2012) retrieved soil moisture using polarimetric images derived from RADARSAT-2 in HH and HV polarizations over alpine meadows and pastures of Mazia Valley, South Tyrol, Italy.

Over the last decades, many soil moisture retrieval algorithms have been proposed to retrieve soil moisture from active microwave sensors and they are generally grouped into theoretical, empirical and semi-empirical models (Kornelsen & Coulibaly, 2013). Theoretical models simulate the backscatter coefficient from the dielectric constant of the soil, surface roughness and radar configurations and afterwards soil moisture is estimated by mathematical inversion of these models. The Integral Equation Model (IEM) is a typical example of such models. In general, theoretical models have a physical basis which makes them robust and applicable over a wide range of land surface conditions. Nevertheless, their application is limited, because of their complexity, large input data requirement, and high sensitivity to surface roughness (Petropoulos et al., 2015).

Alternatively, the empirical method is based on an empirical relationship between soil moisture and radar backscatter coefficients. Many studies have reported a linear relationship between the two (Kelly et al., 2003; Shoshany et al., 2000), whereas few studies have found a nonlinear relationship (Paloscia et al., 2013; Tomer et al., 2015). The models proposed by Oh et al. (1992) and Dubois et al. (1995) are the most widely known empirical models. Overall, this method is simple, but it is site specific and therefore not transferable to other places. In addition, it lacks physical basis and requires a good quality of in-situ soil moisture measurements for its calibration, which is often scarce (Karthikeyan et al., 2017).

Semi-empirical models combine the strength of empirical and theoretical models. This is simply to say that semi-empirical models have some physical basis while relying on experimental data. The main advantage of these models is that they are not site specific as far as the sites are within prescribed limits of the models. They are often valid over a bare to sparsely vegetated surface, but in more vegetated areas the semi-empirical water cloud model proposed by Attema and Ulaby (1978) can be an attractive option for soil moisture retrieval.

Passive Microwave remote sensing

Passive microwave remote sensing sensors (i.e., radiometers) measure the emitted energy from the land surface in the form of brightness temperature, which is proportional to the product of emissivity and the physical temperature of the emitting surface (Schmugge et al., 1986). They are widely recognized as a superior alternative for soil moisture retrieval notably at lower frequencies (e.g., L and P bands) because of their strong sensitivity to soil moisture dynamics (Njoku & Entekhabi, 1996; Schmugge et al., 1986). In addition, at these frequencies they have good capability to penetrate through cloud and vegetation canopy (at low to moderate vegetation density). Above all, the signal-to-noise ratio is high for these kind of sensors at lower frequencies which make them even more attractive for good quality soil moisture retrieval than the active microwave sensors.

Contrary to active sensors such as SARs, passive microwave remote sensing products are characterized by much coarser spatial resolution. This is mainly because of the low energy emission at lower frequencies which leads passive microwave sensors to look at larger field of view to detect this emission from the land surface resulting in coarse-spatial resolution (Karthikeyan et al., 2017). It is also important to note that they only measure near-surface soil moisture (i.e., roughly 5 to 10 cm for L-band sensors).

Similar to the active microwave sensors, the observed signal by passive microwave radiometers is affected by the land surface conditions such as surface roughness and vegetation, which means that over the bare soil, the sensitivity of the signal to soil moisture variation is higher, but with the presence of surface roughness and vegetation this sensitivity is reduced. Nevertheless, a number of theoretical and empirical studies have suggested that at lower frequency bands the influence of vegetation and surface roughness is greatly lessened (Njoku & Entekhabi, 1996; Petropoulos et al., 2015).

Soil moisture retrieval from passive microwave remote sensing involves two general steps. The first one involves relating the brightness temperature with the dielectric constant of the soil using the radiative transfer model (RTM), whereas the second step involves the use of dielectric mixing models to estimate soil moisture from the dielectric constant of the soil. However, the implementation of these steps varies depending on the land surface conditions such as bare soil and vegetation.

Over a bare and smooth soil, the thermal emission (i.e., in the form of brightness temperature T_B) can be expressed as the product of emissivity (e) and the physical temperature of soil surface (T_s) (equation (2.1)):

$$T_B = eT_s = (1 - r_s)T_s \tag{2.1}$$

where T_B is brightness temperature and r_s is soil reflectivity from smooth surface. The r_s can be expressed using the Fresnel equations (equation (2.2)) as function of the dielectric constant of the soil (ε) and viewing angle (θ) (Landau et al., 1961) :

$$r_s^H = \left|\frac{\cos\theta - \sqrt{\varepsilon - sin^2\theta}}{\cos\theta + \sqrt{\varepsilon - sin^2\theta}}\right|^2 \qquad r_s^V = \left|\frac{\varepsilon\cos\theta - \sqrt{\varepsilon - sin^2\theta}}{\varepsilon\cos\theta + \sqrt{\varepsilon - sin^2\theta}}\right|^2 \tag{2.2}$$

where r_s^H and r_s^V is reflectivity from smooth surface at horizontal and vertical polarizations, respectively.

In reality, natural soil surfaces are not generally smooth and therefore the soil roughness should be considered during the estimation of the surface reflectivity. Hence, a semi-empirical models formulated by Wang and Choudhury (1981) and shown in equations (2.3) and (2.4) are widely used to estimate the reflectivity from a rough soil surface (r_r) as a function of the reflectivity of a smooth surface (r_s), polarization mixing ratio (Q), and roughness parameter which characterizes surface roughness height (h).

$$r_r^H = [(1 - Q)\, r_s^H + Q\, r_s^V] * e^{\left(-hcos^N(\theta)\right)} \tag{2.3}$$

$$r_r^H = [(1 - Q)\, r_s^H + Q\, r_s^V] * e^{\left(-hcos^N(\theta)\right)} \tag{2.4}$$

where r_r^H and r_r^V are the reflectivity from a rough surface at horizontal and vertical polarizations, respectively.

A more robust model which considers the effect of vegetation such as absorption, attenuation and scattering on the emission is needed over a vegetated land surface. In this context, a zero order RTM has been widely applied and can be formulated as in equation (2.5):

$$T_{BP} = (1 - \omega_P)(1 - \gamma_p)(1 + \gamma_p r_{sp})T_c + (1 - r_{sp})\gamma_p T_s \tag{2.5}$$

where T_s and T_c are the effective soil and vegetation canopy temperatures, p is either horizontal (H) or vertical (V) polarization, r_{sp} is the soil reflectivity, ω_p is the single scattering albedo, and γ_p is the vegetation attenuation factor which can be obtained from the optical depth (τ_p) and viewing angle (θ) using equation (2.6):

$$\gamma_p = e^{\left(-\frac{\tau_p}{cos\theta}\right)} \tag{2.6}$$

The optical depth in the equation (2.6) can be estimated using equation (2.7) (Jackson and Schmugge (1991)):

$$\tau_p = b\, VWC \tag{2.7}$$

where b is a proportionality value which can be obtained from land cover-based look-up table and VWC is vegetation water content.

In general, the RTM is a premise for soil moisture retrieval algorithms of several passive microwave missions. For example, the single channel algorithm (SCA) of SMAP, the L-band Microwave Emission of the Biosphere (L-MEB) Model of SMOS and the land parameter retrieval model (LPRM) of AMSR2 are all dependent on the RTM.

Following the simulation of brightness temperature by the RTM, a dielectric mixing model is used to estimate soil moisture from the dielectric constant of the soil. The large contrast in the dielectric constant of dry and wet soil is the theoretical basis for dielectric mixing models. In addition, they are heavily dependent on soil textural information such as porosity, and percentage of clay and sand content of the soil. There are a number of dielectric mixing models including the Wang and Schmugge model (Wang & Schmugge, 1980), the Dobson model (Dobson et al., 1985) and the generalized refractive mixing dielectric model (Mironov et al., 2004). These models have been used in different passive microwave satellite missions. For example, the Wang and Schmugge model has been used in the SMAP soil moisture retrieval algorithm, whereas the Dobson model has been used in the AMSR-E/2 soil moisture retrieval process.

2.2 Enhancing the spatial resolution of passive MW soil moisture

Many of hydrometeorological and agricultural applications require soil moisture information at high-spatial resolution (e.g., at kilometer or sub-kilometer scale) and with good quality, but none of the abovementioned remote sensing techniques fulfill these requirements (Peng et al., 2021; Sabaghy et al., 2018). For example, optical/thermal sensors have the capability to estimate soil moisture at high-spatial resolution, but can be highly affected by the presence of clouds, dense vegetation and surface roughness. Similarly, SAR sensors, e.g., onboard Sentinel-1A/B, have presented an unprecedented opportunity in terms of providing of soil moisture at high-spatial resolution (i.e., in the order of tens of meters), but are profoundly affected by land surface characteristics such as vegetation canopy and roughness. On the other hand, satellite passive microwave radiometers, notably at L-band, provide soil moisture relatively with a decent temporal resolution (i.e., 2 to 3 days), but at coarse-spatial resolution ranging from 25 to 50-km.

Despite of these limitations, several efforts have been made to produce soil moisture maps at high-spatial resolution either through direct retrieval or spatial downscaling. Direct retrieval entails retrieval of soil moisture at high-spatial resolution notably from SAR sensors such as Sentinel-1 and RCM, which are currently the most prominent for generation of soil moisture at high-spatial resolution. A number of studies have attempted to retrieve soil moisture from Sentinel-1 at the field or the watershed scale with a broad range of land surface and hydro-climatic conditions. For example, Balenzano et al. (2021) have successfully retrieved soil moisture from Sentinel-1A at 1-km resolution over Yanco (New South Wales, Australia) and Little Washita (Oklahoma, USA) watersheds. Similarly, Hajji et al. (2017) retrieved soil moisture at 1-km resolution from Sentinel-1A over the Occitanic region of France.

An alternative approach to the direct retrieval is spatial downscaling, which involves disaggregation of the coarse resolution satellite passive microwave soil moisture products using high-resolution land surface and atmospheric variables (e.g., land surface temperature, and vegetation indices) derived from multi-sources including optical, thermal, radar and/or passive sensors. There are a host of downscaling techniques, which are generally classified as optical, radar and radiometer-based downscaling based on the source of high-resolution land surface variables used for downscaling. In addition, there are emerging downscaling techniques such as ML and data assimilation techniques (Peng et al., 2017; Sabaghy et al., 2018). Among these approaches, optical/thermal-based downscaling was identified as the most widely applied technique based on a review of studies conducted by Sabaghy et al. (2018) in different regions of the world under varying land surface and hydro-climatic conditions. Similarly, they found that radar-based downscaling and ML models are the two most promising approaches for soil moisture downscaling. These downscaling techniques are briefly described in the following sub-sections.

2.2.1 Optical/thermal based downscaling

The optical/thermal based downscaling entails disaggregation of the coarse-resolution satellite passive microwave soil moisture products such as SMOS and SMAP using high-resolution land surface variables retrieved from the optical and thermal sensors such as MODIS. This downscaling technique generally relies on the universal triangular/trapezoidal feature space established between vegetation indices and land surface temperature over heterogeneous land surfaces (Carlson, 2007; Moran et al., 1994). There are different varieties of optical/thermal based downscaling such as polynomial fitting, weighting factor, and physically-based downscaling methods.

The polynomial fitting downscaling technique involves establishing the polynomial function between the coarse-resolution satellite soil moisture and upscaled high-resolution land surface variables (i.e., aggregated to resolution of satellite soil moisture). The calibrated function at coarse-resolution is then applied to predict soil moisture at high-spatial resolution from high-spatial resolution land surface variables by assuming that the developed function is spatial scale invariant.

A number of studies have applied polynomial fitting technique to downscale SMOS (Piles et al, 2016; Piles et al., 2011; Sánchez-Ruiz et al., 2014), AMSR-E (Choi & Hur, 2012), Special Sensor Microwave Imager (SSM/I) (Chauhan et al., 2003), ASCAT (Montzka et al., 2017) to 1-

km or less using high-resolution land surface variables and indices derived from MODIS (onboard Aqua and Terra satellites), ASTER (Terra), TM (Landsat) or SEVIRI (Meteosat Second Generation - MSG)

The polynomial fitting technique is easy to apply with minimum data requirement, but its application is limited to bare and sparsely vegetated surfaces. In addition, its input can be affected by cloud hampering the generation of seamless soil moisture data. However, the use of geostationary satellites can alleviate this problem. For example, Zhang and Li (2013) and Piles et al (2016) disaggregated AMSR-E and SMOS soil moisture to 5- and 3-km, respectively, using MSG-SEVIRI radiometer data which enabled them to generate both spatial and temporally continuous soil moisture. However, MSG-SEVIRI provide data only for four specific regions (Europe, North Africa South Africa, and South America).

Alternatively, the weighting factor technique determines the variability of soil moisture in the grid cell of the coarse-resolution satellite passive microwave soil moisture using the soil wetness index developed at high-spatial resolution(Kim & Hogue, 2012a). In other words, this index is used as the scaling factor after normalizing by the mean of high-resolution soil wetness index in the coarse grid cell of satellite passive microwave soil moisture. The normalized soil wetness index is then multiplied with the coarse resolution satellite soil moisture to generate soil moisture at high-spatial resolution as shown in equation (2.8):

$$Soil\ moisture_{fine\ resolution} = Satellite\ SM_{coarse} \quad x\,\frac{SWI_{fine\ resolution}}{\overline{SWI}} \tag{2.8}$$

where SM is soil moisture and SWI soil wetness index

Several studies have applied this technique to downscale satellite passive microwave soil moisture. For instance, Kim & Hogue (2012b) downscaled AMSR-E soil moisture from 25- to 1-km using a soil wetness index named as UCLA. Their method improved the spatial detail of AMSR-E soil moisture, but with high uncertainty over densely vegetated areas. Similarly, Peng et al. (2015) downscaled the European Space Agency Climate Change Initiative (ESA CCI) soil moisture product from 25- to 5-km using the Vegetation Temperature Condition Index (VTCI). They demonstrated the feasibility of the method in downscaling of CCI soil moisture and suggested further application of the method in different land surface and climate conditions.

The weighting factor approach is simple to apply and requires minimum input data, yet the output of this method is highly dependent on the quality of soil moisture to be downscaled and the accuracy of the scaling factor. In addition, it is difficult to generate spatio-temporally continuous soil moisture because of the effect cloud and its application is limited from bare to sparsely vegetated areas.

The other popular variant of Optical/thermal based downscaling is Disaggregation based on Physical And Theoretical scale Change (DisPATCh) (Merlin et al., 2012). This technique has a physical basis in contrast to the abovementioned techniques. It uses the soil evaporative efficiency (SEE) derived from the universal trapezoid concept to downscale satellite passive microwave soil moisture products. For example, Merlin et al. (2012) have successfully applied this approach to downscale SMOS soil moisture from 40- to 1-km over semi-arid area located in Australia. They found that the DisPATCh algorithm performance varies based on the season. For instance, for summer season they found correlation of about 0,7 whereas during winter it dropped to zero, which might be attributed to lower coupling between evaporation and soil moisture.

Similarly, Djamai et al. (2015) applied the same algorithm to evaluate its performance under wet condition over agricultural site in the Canadian Prairies, near Kenaston. They found improvement in the spatial details of soil moisture with a correlation of 0,7 against in-situ soil moisture, yet they stated the difficulty in finding an optimum area over which atmospheric condition is uniform for the estimation of maximum and minimum temperature in SMOS pixel, which in turn affects the estimation of SEE and thereby affecting the soil moisture downscaling process. This is possibly because of the low spatial variability of soil moisture over wet areas.

Overall, the optical\thermal based downscaling has the advantage of using high-resolution auxiliary variables derived from the optical\thermal remote sensing (e.g., MODIS, TM) for downscaling of satellite passive microwave soil moisture. In the same manner, it is simple and relies only on few inputs, e.g., vegetation index and land surface temperature. However, the optical\thermal remote sensing signal can be contaminated by cloud and vegetation cover. This affects the quality of as well as the spatial and temporal continuity of auxiliary variables derived from them, which in turn reduces the possibility of generating spatio-temporally seamless downscaled passive microwave soil moisture with the desired quality. In addition, the optical/thermal based downscaling does not take into account other factors, which affect the spatial heterogeneity of soil moisture, such as topography, soil properties and meteorological forcing, such as precipitation. Furthermore, its application is limited to areas with uniform meteorological forcing and heterogeneous land cover with bare to low vegetated areas.

2.2.2 Radar-based downscaling

Radar-based downscaling uses high-resolution observations derived from Synthetic aperture radar (SAR) to downscale the coarse resolution satellite passive microwave soil moisture products. A well-known example of this approach is the SMAP baseline downscaling algorithm (Das et al., 2011) developed during the preparation for launch of NASA's SMAP mission which aimed to downscale the coarse-resolution (36-km) SMAP radiometer observations (i.e., brightness temperature) to an intermediate resolution (9-km) using high-resolution (3-km) radar observations (i.e., backscatter coefficient). The downscaled brightness temperature at 9-km is then inverted to estimate soil moisture using a forward RTM model. It must be recalled that the SMAP radiometer and radar are on the same platform. Unfortunately, the SMAP radar became inoperable on July 7th, 2015. Following the failure of the SMAP's L-band radar, a successful attempt was made to use Sentinel-1 C-band SAR images for continuation of the generation medium resolution soil moisture knowns as the SMAP\Sentinel-1 soil moisture using the same algorithm (i.e., SMAP baseline algorithm).

The global validation of the SMAP\Sentinel-1 soil moisture with in-situ measurements retrieved from ISMN indicated satisfactory results across a range of land surface conditions, with exception of forested areas (Mohseni et al., 2022). They also found that SMAP radiometer soil moisture has better performance than the SMAP\Sentinel-1 soil moisture. Overall, the SMAP\Sentinel-1 soil moisture has low temporal resolution because of the longer revisit time of Sentinel-1 and its spatial coverage is limited to the area covered by Sentinel-1 images.

Another method which is closely related with the SMAP baseline algorithm is the SMAP optional downscaling algorithm which directly downscales the SMAP radiometer soil moisture using the radar backscatter. In addition, the change detection method (Piles et al., 2009), the Bayesian framework (Zhan et al., 2006) and wavelet transformation (Gabriel & Virginia, 2017) are used for downscaling of satellite passive microwave soil moisture.

A number of studies have applied radar based downscaling techniques to downscale satellite passive microwave soil moisture and brightness temperature across a broad range of land and climatic conditions. For example, Wu et al. (2017) tested the baseline downscaling algorithm using airborne passive and active microwave observations collected over the arid landscape in Australia. They found that the accuracy of downscaled brightness temperature fulfills the requirements of the SMAP mission. Similarly, Li et al. (2018) applied wavelet transformation technique to downscale SMOS soil moisture from 40- to 25- and 1,25-km using Sentinel-1 SAR observations.

Radar-based downscaling has the advantage of downscaling satellite passive microwave soil moisture at very finer resolution because of the high-spatial resolution of SAR observations (i.e., in the order of tenth of meters). In contrast to optical/thermal remote sensing, SAR is capable of penetrating through low to moderately vegetated areas notably at lower frequencies, thereby properly reflecting the sensitivity of backscatter to soil moisture which in turn helps in generation of good quality downscaled passive microwave soil moisture.

However, this technique has limitations such as low temporal resolution because of the longer revisit time of radar unless both radiometer and radar are on the same platform like the SMAP mission, which is currently non-existent. In addition, the differences in radiometer and radar sensors properties such as frequencies and swath width are another limitation. For example, Sentinel-1 has C-band with swath width of 300-km, whereas SMAP radiometer has L-band with swath width of 1000-km. This difference limits the coverage of downscaled SMAP\Sentinel-1 soil moisture to the area covered by Sentinel-1 SAR images. However, Sentinel-1 images can be mosaicked along flight direction because of temporal proximity of the images. This is applicable for small and medium sized watersheds which can be covered by Sentinel-1 images along the direction of flight. However, for large watersheds many Sentinel-1 images with different temporal gaps are needed to cover the entire watershed. Merging such images with long temporal gaps (e.g., 12 days) distorts the information intended to infer.

Moreover, the success of the radar-based downscaling depends on land surface and hydro-climatic conditions of the study area. For example, previous studies have successfully applied this method in low vegetated areas with arid and semi-arid climates such as Murrumbidgee river watershed (He et al., 2018; Wu et al., 2016) and Berambadi watershed (South India) (Tomer et al., 2016). Similar studies have been conducted in areas dominated with agricultural crops e.g., Walnut creek watershed (Piles et al., 2009) and South of Ontario (Canada) (Li et al., 2018). However, still further studies are needed over a range of land surface and climate conditions such as over moderately to heavily forested area with humid climate.

Finally, the increase in the temporal record of SAR observations such as Sentinel-1 greatly supports future studies which need longer time series of SAR observation for downscaling passive microwave soil moisture as opposed to previous studies which based on shorter temporal record of SAR images.

2.2.3 Radiometric based downscaling

The radiometer-based downscaling relies on the use of high-resolution observations derived from passive microwave sensor at higher frequencies for downscaling of the coarse-resolution satellite passive microwave soil moisture often retrieved at lower frequencies. A well-known example of this approach is the Smoothing Filter-based Intensity Modulation (SFIM). This method was often applied to the data derived from the AMSR-E sensor which is accompanied with a range of frequencies (e.g., 6,925, 10,65, 18,7, 23,8, 36,5, and 89,0 GHz).

For example, De Jeu et al. (2014) downscaled the AMSR-E brightness temperature retrieved at lower frequency (6,9 GHz) with resolution of 56-km using high-resolution observation (11-km) retrieved at higher frequency (36,5 GHz). The downscaling was carried over the Australian Fitzroy watershed (Northeastern Australia), a watershed with sub-tropical and sub-humid climate and dominated by grass (78 %), using the Smoothing Filter-based Intensity Modulation (SFIM) method. The Land Parameter Retrieval Model (LPRM) algorithm was then used to retrieve soil moisture from the downscaled brightness temperature. They found that downscaling improved the spatial representation of soil moisture, yet they revealed that impact of precipitation on the observation at higher frequency affect the accuracy of the downscaled soil moisture.

2.2.4 Machine learning based downscaling

Machine learning (ML) is an essential part of artificial intelligence, which is totally data driven without relying on any physical basis (Xu & Liang, 2021). Recently, it received growing interest in many disciplines including hydrology, agriculture, remote sensing and many others. Notably, in remote sensing, in recent years it has received remarkable attention in downscaling of the coarse-resolution geophysical and atmospheric variables such as satellite passive microwave soil moisture, precipitation and land surface temperature (Belgiu & Drăgu, 2016).

ML has the capability to describe the complex non-linear relationship between soil moisture and soil moisture proxies such as land surface temperature, vegetation indices, topography and many others. Its basic implementation involves construction of a non-linear relationship at coarse-spatial resolution between satellite soil moisture and upscaled high-resolution land surface variables. The model constructed at coarse-spatial resolution is then used to predict soil moisture at finer resolution using high-resolution land surface variables as input, on the condition that the constructed relationship is scale invariant. A number of studies have demonstrated the capability of different variants of ML techniques such as artificial neural network (ANN) (Alemohammad et al., 2018), classification and regression trees (CART)(Liu et al., 2020), gradient boosting decision tree (GBDT) (Zhao et al., 2019), extreme gradient boosting (XGB) (Rao et al., 2022), and random forest (RF) (Abbaszadeh et al., 2019; Zhao et al., 2018) in the downscaling of satellite passive microwave soil moisture products. For example, Zhao et al. (2018) downscaled the SMAP L3 soil moisture from 36- to 1-km resolution over the Iberian Peninsula using the RF ML technique. Wei et al. (2019) downscaled the SMAP L3 soil moisture to 1-km resolution using the gradient boosting decision tree over Continental United States (CONUS). Similarly, Wen et al. (2021) downscaled the SMOS soil moisture from 40-to 1-km over an alpine mountains basin (Northwest of China), whereas Im et al. (2016) downscaled the AMSR-E soil moisture from 25- to 1-km resolution over two different regions with contrasting climate characteristics (Australia and South Korea).

A number of comparative studies have compared the performances of different variants of ML techniques. For example, Im et al. (2016) compared three ML techniques including random forest, boosted regression trees and Cubist over the entire country of South Korea and New South Wales (Australia). South Korea is dominated with forest land cover (80 %) followed by cropland (19 %), whereas New South Wales is dominated by shrub (64 %) followed by cropland (18 %) and forest (10 %). They found the outperformance of RF ML technique for both study areas. In addition, they showed that the RF showed better results for New South Wales than that of South Korea.

Similarly, Liu et al. (2020) compared six ML algorithms (i.e., ANN, BAYE, CART, KNN, SVM, RF) commonly used in the downscaling of satellite passive microwave soil moisture products. The study was carried over four selected locations across the globe with stable long-time series of in-situ soil moisture measurements. These are Okalahoma Mesonet (OKM) in the south of the high plains of the Mississippi River Basin, REMEDHUS (REM) in the northern part of the Iberian Peninsula, the Naqu network (NAN) in southeastern margin of the Tibetan Plateau and OZNNET (OZN) in coastal plains of southeast Australia. All of these sites are dominated by cropland and grassland\shrubland and the study illustrated the superiority of the RF approach for all the sites.

Many studies have used combination of land surface and atmospheric variables derived from different sources including MODIS, and SRTM for downscaling of the passive microwave soil moisture products (Abbaszadeh et al., 2019; Bai et al., 2019; Im et al., 2016; Zhao et al., 2018, 2019). Interestingly, predictors derived from MODIS along with topographic derivatives extracted from SRTM DEM are the most widely used. For example, Zhao et al. (2018) used LST, NDVI, enhanced vegetation index (EVI), Albedo (ALB), leaf area index (LAI), and normalized difference water index (NDWI) as predictors for downscaling of the SMAP soil moisture retrievals. Similarly, Abbaszadeh et al. (2019) used precipitation, LST, elevation (Elev), original SMAP soil moisture, and NDVI as predictors for downscaling of the SMAP soil moisture. On the other hand, Wen et al.(2021) used LST, NDVI, ALB, Elev, slope and Evapotranspiration (ET) for downscaling of the SMOS soil moisture, whereas Im et al. (2016) used ALB, LST, NDVI, EVI, LAI, and ET for downscaling of AMSR-E soil moisture.

However, there are only a few studies which used predictors derived from Sentinel-1A/B along with those derived from other sources such as MODIS. For example, Bai et al. (2019) successfully downscaled the SMAP enhanced soil moisture from 9- to 1-km using NDVI, EVI, NDWI, LAI, LST, ALB, Elev and radar backscatter over northern China. For downscaling they used random forest machine learning technique and reported high importance of Sentinel-1 backscatter followed by LST, LAI and NDVI. In addition , recently Karami et al. (2022) used radar backscatter, Elev, NDVI, land cover, % clay and % sand for downscaling the SMAP enhanced soil moisture from 9- to 1-km successfully. In their downscaling they used random forest machine learning technique and reported variation in the importance of predictors based on their study location characteristics. For example, for Murrumbidgee watershed (Australia) Sentinel-1 backscatter have the highest importance followed by NDVI and land cover, whereas for Firoozabad watershed (Iran) NDVI became the most important predictors followed by elevation and radar backscatter. These differences in timportance of predictors could be due to variation of land surface and climatic condition of the their study area, which they did not explain in detail. However, further studies are needed to confirm the suitability of SAR predictors in the downscaling of satellite passive microwave soil moisture over a range of land surface and climatic conditions.

Overall, the success of downscaling of the passive microwave soil moisture products strongly depends on land surface and climatic conditions of the study area besides the downscaling method used. This is mainly because of land surface conditions such as the density of vegetation cover determines the quality of satellite passive microwave soil moisture retrieval (Wu et al., 2015). Indeed, the quality of retrieval decreases with the increase of vegetation density. Thus, the lower the vegetation cover the better the quality of retrieved soil moisture. Interestingly, the same holds true for the predictors used for downscaling such as land surface temperature. The presence of vegetation hinders the retrieval of accurate land surface temperature as there is possibility of contribution of emission from vegetation. For these reasons most of the previous satellite soil moisture downscaling studies have been carried out in less vegetated regions where the effect of vegetation is minimal. However, further studies are needed to confirm their performance under varying land and climate conditions e.g., for moderately vegetated regions.

2.3 Hydrological Applications of satellite soil moisture

Hydrometeorological disasters, notably floods, remain the most detrimental hydrological disaster causing loss of life, destruction of property and infrastructures across the globe. Owing to its impact on the society, accurate and timely flood forecasting is decisive (Wanders et al., 2014). Hydrological models play a central role in this regard. However, they are subjected to a range of uncertainties including model structure, parameters, initial condition, meteorological forcing and observation data used for calibration (Liu & Gupta, 2007).

In the past decades, several studies have proposed different approaches of reducing these uncertainties (Beven, 2006; Liu & Gupta, 2007; Wagener, 2003). Among them one of the most promising approaches is an accurate representation of pre-storm soil moisture conditions of the watershed (i.e., also known as initial condition). This is in the light of the role soil moisture plays in partition of incoming rainfall into surface runoff and infiltration, thereby controlling the magnitude and timing of flood.

Over the last couple of decades, the assimilation of satellite soil moisture into hydrological and land surface models has gained much attention mainly because of the increased availability of spatially distributed soil moisture observations from microwave remote sensing satellites (e.g., AMSR-E, SMOS, SMAP and Sentinel-1). Accordingly, many studies have assimilated satellite soil moisture into land surface models for improving soil moisture estimation (Baldwin et al., 2017; De Lannoy & Reichle, 2016; Lievens et al., 2017; Zhao et al., 2013). For example, Baldwin et al. (2017) assimilated the AMSR-E soil moisture into the SMAR (Soil Moisture Analytical Relationship) model for improving the root-zone soil moisture estimation. Similarly, De Lannoy et al. (2016) assimilated SMOS soil moisture into the Goddard Earth Observing System Model for improving surface and root-zone soil moisture estimation.

The assimilation of satellite soil moisture for hydrological applications such as streamflow forecasting and drought monitoring has gained much attention recently. A number of studies have assimilated satellite microwave soil moisture derived from, e.g., AMSR-E (Alvarez-Garreton et al., 2014; López et al., 2016), SMOS (Leroux et al., 2016; Lievens et al., 2015), SMAP (Abbaszadeh et al., 2020; Jadidoleslam et al., 2021; Le et al., 2022), ASCAT (Brocca et al., 2010; Loizu et al., 2018; Massari et al., 2018) and Sentinel-1 (Azimi et al., 2020; Cenci et al., 2017) into a wide range of hydrological models, from simple conceptual models (e.g., GR4J) to physically-based fully distributed models (e.g., WRF-HYDRO).

Although these studies have assimilated soil moisture derived from different satellites, they have a similar objective, which is improving hydrological model predictions, e.g., improving streamflow forecasts (Massari et al., 2018b; Wanders et al., 2014), streamflow simulation (Azimi et al., 2020; Le et al., 2022) and drought monitoring (Bolten et al., 2010).

With regards to the impact of soil moisture assimilation on the accuracy of model prediction, studies have reported contrasting results possibly because of the assimilation schemes adopted, and physiographic characteristics of the watershed among many others. For example, Azimi et al. (2020) and Le at al. (2022) have found significant improvement in simulation of streamflow, while others have reported no to little improvement (e.g., Alvarez-Garreton et al., 2016; Kumar et al., 2014; Parajka et al., 2006).

Moreover, many of these studies have assimilated satellite soil moisture at its native resolution (i.e., 25 to 50-km). Indeed, this resolution might be sufficient for global and continental scales hydrological applications, but for local scale hydrological applications it cannot adequately represent the spatial heterogeneity of soil moisture. In this regard, the use of downscaled satellite soil moisture plays an important role for adequate representation of the spatial heterogeneity of soil moisture.

However, there are only a few studies that have assimilated downscaled satellite soil moisture notably from the passive microwave satellites which are fully dedicated to soil moisture measurements such as SMOS and SMAP. For example, Abbaszadeh et al.(2020) and Le et al. (2022) assimilated the 1-km downscaled SMAP soil moisture using Particle Filter and Ensemble Kalman Filter, respectively and found improvement in streamflow simulations. López et al. (2016) assimilated the 9-km downscaled soil moisture derived from AMSR-E and found improvement in streamflow simulation. Interestingly, in contrast to assimilation of the native resolution satellite soil moisture which resulted in differing results, the assimilation downscaled satellite soil moisture generally resulted in the improvement of streamflow simulation so far.

Moreover, the success of assimilation of the downscaled satellite passive microwave soil moisture highly relies on the physiographic and climate conditions of the study area. For example, the study by Abbaszadeh et al.(2020) has carried in the watershed with area 2577 km^2 located in the Southeastern part of Texas (USA) with warm and humid climate and moderately dense vegetation. On the other hand, Le et al. (2022) considered eight watersheds located across tropical Vietnam with area ranging from 267 to 6340 km^2. Three of the watersheds are dominated by forest land cover and the remaining five are dominated by shrubland and cropland. Similarly, the study by López et al. (2016) was conducted in the Murrumbidgee River basin (Australia), which is a dry watershed with sparse vegetation. In general, these studies have found improvement in the accuracy of streamflow simulation after assimilation of the downscaled satellite passive microwave soil moisture. However, the assimilation of downscaled satellite soil moisture is still at its early stages and further investigations are needed to confirm whether they continue to show similar results under different physiographic characteristics and climate conditions of the study area.

Chapter 3 METHODOLOGY

This chapter begins with description of the study watersheds followed by illustration of the attributes of dataset used. Afterwards, the hydrological model and the RF ML downscaling technique are briefly described. In the end, an overview of experimental procedures and frameworks is presented.

3.1 Study area

The Susquehanna and au Saumon watersheds located in eastern part of the United States and Southern of Québec (Canada), respectively, were selected for this study, see Figure 3.1.

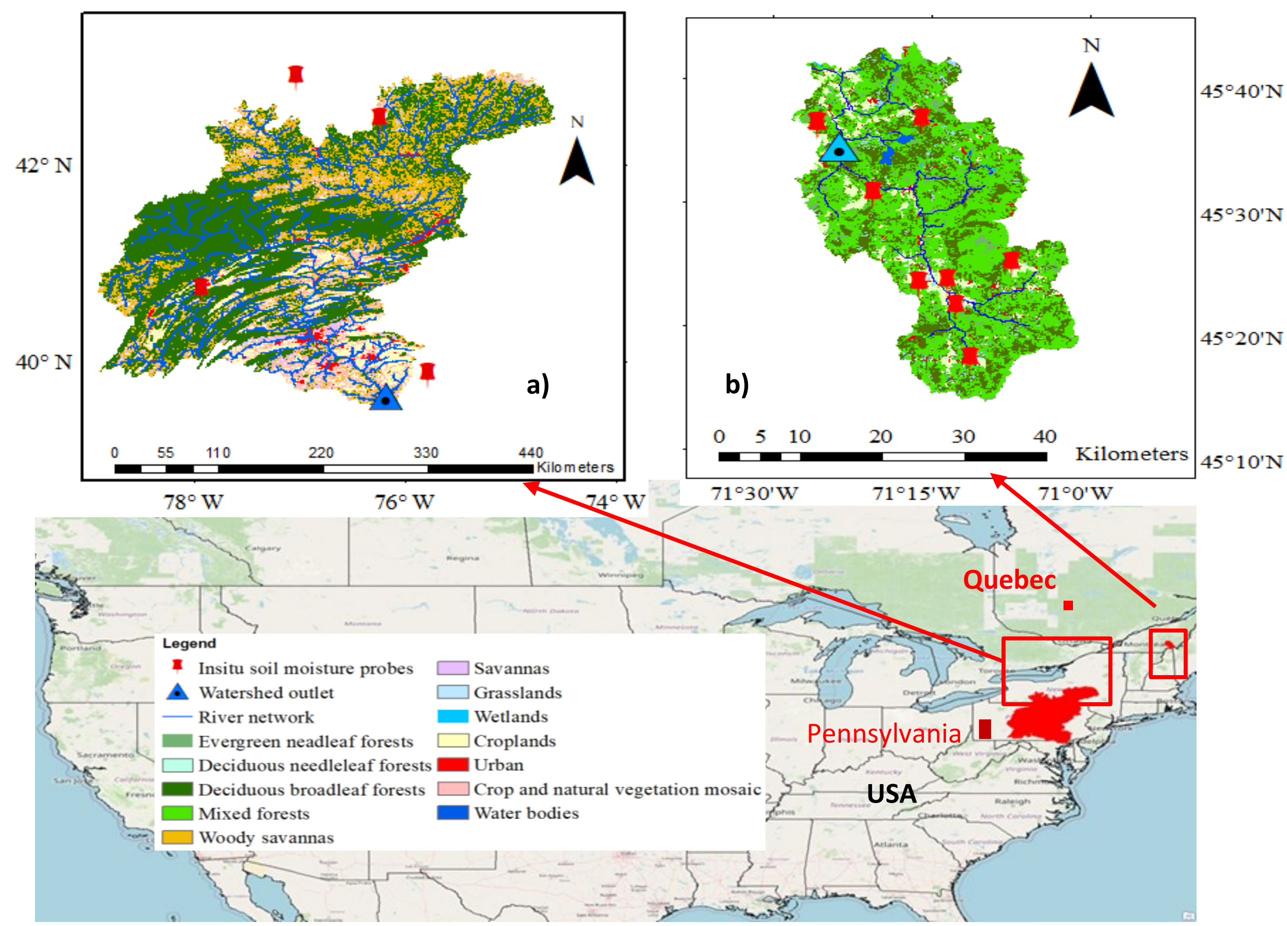

Figure 3.1 Location of the Susquehanna (left) and au Saumon (right) watersheds

The Susquehanna River watershed with an area of 71,225 km^2 is a large watershed dominated by forest (55%) and agriculture (33%) (DePhilip & Moberg, 2010). It has a humid continental mild summer with wet all year climate (Dfb class) according to Köppen-Trewartha climate classification (Belda et al., 2014). The mean annual temperature for the northern and the southern parts of the watershed is 6 and 12 ºC, respectively (Gan et al., 2022). It receives precipitation in the form of snowfall during the winter (November-April) and in the form of rainfall during the summer (May-November). Its average annual precipitation varies from 800 to 1250 mm. The high streamflow occurs during the spring season (March - May) due the combined effect of snowmelt and rainfall, whereas low streamflow occurs in late summer, see Figure 3.2. The elevation of this watershed varies between 4 and 1150 m.

The second watershed, au Saumon, has a drainage area of 1025 km^2 (Bergeron et al., 2014). Its main land cover is forest (i.e., 82 %). Similar to the Susquehanna River watershed, it is also located in the Dfb climate class with annual precipitation of roughly up to 1250 mm and average annual temperature of 4,5 ºC. The elevation of the watershed ranges from 277 and 1092 m. For this watershed, the high streamflow events occur in spring and fall seasons due to snow melt and rainfall, respectively, see Figure 3.2. Note that for the au Saumon watershed, streamflow is measured at a gauging station draining a watershed area of 769 km^2.

In this study, the Susquehanna watershed was chosen because of the diversity of its land covers. Knowledge of soil moisture variability is important for such an area to see its role in hydrological applications. In addition, the Susquehanna watershed is frequently affected by flood events. Thus, studying the impact of soil moisture wetness on flood generation was deemed appropriate for such a watershed. On the other hand, au Saumon watershed was selected due to the fact that we have in-situ soil moisture monitoring stations which deemed helps to condition the SMAP soil moisture which tends to overestimate soil moisture over au Saumon because of its forested land cover. Above all, both Susquehanna and au Saumon watersheds are our industrial research partners (e.g., Brook field and hydro-Quebec) watersheds, which must be considered in our study.

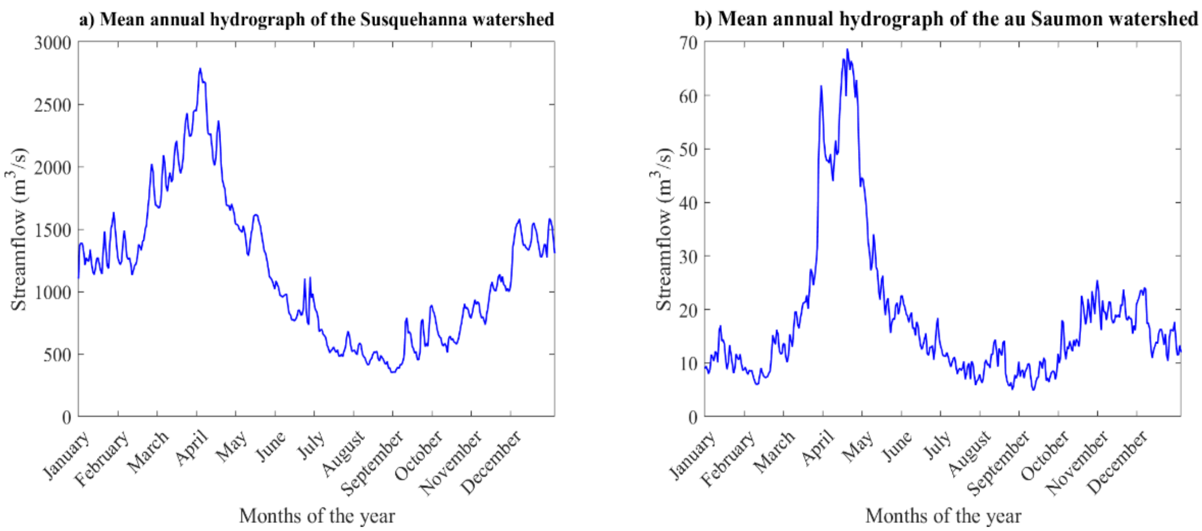

Figure 3.2 Mean annual hydrograph of the a) Susquehanna and b) au Saumon watersheds based on historical record of 1970 to 2021 and 1975 to 2022, respectively.

The historical streamflow data used for plotting of the mean annual hydrograph for Susquehanna and au Saumon in the Figure 3.2 was obtained from the United States Geological Survey (USGS) and Centre d'expertise hydrique du Québec, respectively.

3.2 Data

3.2.1 Satellite and in-situ observations

The SMAP level 3 (L3SMP) and SMAP enhanced (L3SMP_E) surface soil moisture with resolution of 36- and 9-km, respectively, were used in this study. The SMAP enhanced is essentially generated from L3SMP using Backus-Gilbert optimal interpolation. BG optimal interpolation technique is applied to the time-ordered T_B observations in L1B_TB. It uses the antenna gain pattern information specific to the 6-meter reflector assembly onboard the observatory to achieve optimal data interpolation on a fine grid from the original data. It is considered optimal due to its aim of producing interpolated values. The process of estimation involves the optimal weighting and linear combination of radiometric measurements that overlap in both the along-scan and across-scan directions (Poe & Conway, 1990).

In addition, the SMAP/Sentinel-1 (L2_SM_SP) soil moisture (Das et al., 2019) with a resolution of 1-km was collected and this product has a temporal resolution of 12 days over the study area. It is important to note that both the SMAP ascending (6 PM) and descending (6 AM) overpass was used in this study depending on the type of downscaling experiments implemented. The selection of was based on (1) temporal proximity to the Sentinel-1 overpass time (6 PM, ascending overpass) and (2) thermal equilibrium between vegetation canopy and land surface (6 AM, descending overpass). The SMAP soil moisture products can be downloaded freely from the NASA Snow and Ice Data Center Distributed Active Archive Center (NSIDC DAAC) (https://nsidc.org/data/smap/ (accessed date 30 May 2019)).

Sentinel-1 is a constellation mission consists of two C-band SAR satellites, i.e. Sentinel-1A and Sentinel-1B, launched by the European Space Agency (ESA) (Torres et al., 2012) respectively on April 3, 2014 and April 26, 2016 (Sentinel-1B end mission declared August 3, 2022) for the purpose of monitoring the land and ocean. They both have the capability to provide observation at spatial resolution as fine as 10 m in both VV and VH polarizations with a revisit time of 6 to 12 days. The interferometric wide swath (IWS) is the prime acquisition mode over the land among the available data acquisition modes such as stripmap, extra-wide swath and wave.

This study collected and pre-processed Sentinel-1A images corresponding to the ascending overpass (i.e., 10:59 PM). The ascending overpass was selected because of its temporal proximity to SMAP ascending overpass (6 PM). The Sentinel application platform (SNAP) tool was used to derive the backscatter intensities both in VV and VH polarizations (McGarragh et al., 2015; Torres et al., 2012). The pre-processing includes thermal noise removal, speckle filtering, radiometric calibration, terrain correction and normalization. Besides backscatter intensity, textural features such as contrast, entropy, correlation and variance were extracted using the grey-level co-occurrence matrix (GLCM) module in the SNAP. The Sentinel-1 images can be accessed by the users through ESA's Sentinel-1 internet data hub (https://scihub.esa.int/ (accessed date 25 June 2019)). The selection of textural features was based on previous studies which have used it for classification of land cover (Chen et al., 2020; Sarker et al., 2013). According to these studies contrast, entropy, correlation and variance are the most important features in land cover classification. It is worth noting that in this study, they were used to take into account the effect of vegetation on soil moisture variation.

The Moderate Resolution Imaging Spectroradiometer (MODIS) is onboard the Aqua and Terra satellites of the Earth Observing System (EOS) and launched in December 1999 and May 2002, respectively. The local equator descending and ascending overpass times are 10:30 AM and 10:30 PM, respectively, for Terra, while for Aqua it is 1:30 AM and 1:30 PM, respectively.

MODIS was launched by NASA primarily for monitoring of the spatial and temporal dynamics of land, ocean and atmosphere surfaces.

A suite of global land and ocean surface products with spatial resolution ranging from 250 to 1000 m can be collected from MODIS with temporal resolution of 1 to 2 days. In this study, the land surface variables including the 1-km resolution daily land surface temperature (MYD11A1, MOD11A1), 500-m daily surface albedo (MCD43A3) and 1-km 16-day normalized difference vegetation index (MOD11A2) were collected from the latest version of MODIS products (collection 6). These variables can be downloaded freely from the NASA Land Processes Distributed Active Archive Center (LP DAAC) (https://lpdaac.usgs.gov (accessed date 28 June 2019)).

The surface topography features including elevation, slope and aspect were extracted from the 30-m resolution digital elevation data derived from Shuttle Radar Topographic Mission (SRTM). This data was downloaded from the Land Processes Distributed Active Archive center (https://lpdaac.usgs.gov (accessed date 1 July 2019)).

In-situ soil moisture measurements recorded at depth of 5 cm were collected from 2 stations of the soil climate analysis network (SCAN) and 2 stations of the U.S. climate reference network (USCRN). These stations (i.e., Avondale, Ithaca, Geneva and Rocksprings) lay either in or near the Susquehanna watershed (Figure 3.1). Both USCRN and SCAN soil moisture has hourly temporal resolution. Additional characteristics of the in-situ soil moisture stations are provided in Table 3.1. This data was collected from the International Soil Moisture Network (ISMN) database (https://ismn.geo.tuwien.ac.at/en/ (accessed date 10 July 2019)). The number of stations of in-situ soil moisture in this watershed is not sufficient for rigorous validation of satellite soil moisture, but at least they can provide information on how soil moisture varies locally. It is recommended to have a number of in-situ soil moisture stations which are well spread within the pixel of satellite soil moisture. For example, Colliander et al., (2017) recommended to have at least 8 stations of in-situ for the 36-km resolution SMAP

Table 3.1 In-situ soil moisture stations properties

Station names	Network	Land cover	Elevation (m)	Precipitation (mm)
Avondale	USCRN	Mosaic herbaceous	122	1280
Geneva		Grassland	221	855
Ithaca	SCAN	Crop	374	1100
Rocksprings		Crop	372	1015

Similarly, for the au Saumon watershed in-situ soil moisture measurements that have been collected at representative locations since 2018 using soil moisture probes were used (Figure 3.1). The installation of these probes is supported by the research chair on valorization of Earth observations for water resources to which this PhD study is part of. These probes provide hourly measurements of soil moisture at 5 and 20 cm depths for each location. The measurements took place only during the summer and early fall seasons because the soil was likely to be frozen near the surface during the winter and spring seasons. The probes were installed both in open and forested areas (Figure 3.1).

3.2.2 Hydrometeorological data

For the Susquehanna River watershed, the daily meteorological dataset including precipitation and maximum/minimum temperature were collected from the Parameter-elevation Regressions on Independent Slopes Model (PRISM) (https://prism.oregonstate.edu/recent/) for the period of 1981 to 2020. The daily streamflow observation was obtained from the United States Geological Survey (USGS) (https://www.usgs.gov/products/data) for the same period.

On the other hand, for the au Saumon watershed, the daily deterministic estimates of precipitation and maximum/minimum temperature extracted from the Ensemble Meteorological Dataset for Planet Earth (EM-Earth) (Tang et al., 2022) for the period 1979 to 2019 was used to force the hydrological model to simulate streamflow. This data can be accessed from Federated research data respiratory (FRDR) of Canada (https://www.frdr-dfdr.ca/repo/dataset/), whereas the daily streamflow observation was obtained from the Centre d'expertise hydrique du Québec (CEHQ) (now the Direction de l'expertise hydrique du Québec) (https://www.cehq.gouv.qc.ca/).

Apart from the deterministic meteorological data mentioned above, the ensemble of meteorological data was extracted from EM-Earth for the purpose of ensemble streamflow simulations and forecasting. The EM-Earth is the recent ensemble meteorological dataset of precipitation, mean daily temperature, and daily temperature range developed at University of Saskatchewan (Tang et al., 2022). In a nutshell, for the EM-Earth ensemble dataset generation first they merged station-based Serially Complete Earth (SC-Earth) dataset with ERA-5 reanalysis estimates and then generated the ensemble members based on parametric probability distributions using spatiotemporally correlated random fields. It provides daily probabilistic estimates with 25 ensemble members. In addition, beside ensembles, it also provides both hourly and daily deterministic meteorological data. This dataset has spatial resolution of roughly 10-km and is available from 1950 to 2019 with a global coverage. In this study, the daily maximum and minimum temperature were estimated from the mean daily temperature and daily temperature range, as they were not provided in EM-Earth dataset. Interested readers can refer to Tang et al. (2022) for the detailed description of this dataset.

3.3 Hydrological model

A continuous and spatially distributed physically-based hydrological model called HYDROTEL is used for this study (Fortin et al., 2001). It was developed at the Institut National de la Recherche Scientifique, Centre Eau Terre Environnement (INRS-ETE), Québec (Canada). It is an appealing model for its easiness in incorporation of spatially distributed GIS and remotely sensed data such as soil type, soil moisture and snow water equivalent and for its minimal meteorological data requirement (i.e., only precipitation and max/min temperature). From an operational application perspective, HYDROTEL is being used for inflow and hydrological forecasting by CEHQ (Boucher et al., 2011).

HYDROTEL is accompanied with a GIS tool called PHYSITEL to spatially discretize the watershed into relatively homogeneous hydrological response units (RHHUs) based on river network, elevation, land cover and soil type. The model simulates hydrological state variables and fluxes for each RHHUs, and the fluxes are routed to the watershed’s outlet. Accordingly, in this study, the Susquehanna and au Saumon watersheds were discretized into 1025 and 205 RHHUs, respectively.

HYDROTEL integrates seven interconnected computational modules. These are: 1) interpolation of precipitation, 2) accumulation and melt of snowpack, 3) potential evapotranspiration estimation, 4) vertical water budget determination, 5) surface and subsurface flow generation, 6) Channel flow routing and 7) soil temperature and soil freezing. Each of these modules has two or more options for simulation of their respective hydrological process (Table 3.2).

Table 3.2 Computational modules of HYDROTEL (indicates methods selected for this study)*

Available modules	Simulation options
Interpolation of meteorological data	Thiessen polygons*, Weighted mean of nearest three stations
Snow accumulation and melt	Mixed (degree-day) energy budget model*, Multilayer model
Soil temperature and soil freezing	Rankinen*, Thorsen
Potential evapotranspiration estimation	Thornthwaite, Linacre, Penman, Priestley-Taylor, Hydro-Québec*, Penman-Monteith
Vertical water budget	BV3C*, CEQUEAU
Overland water routing	Kinematic wave equation*
Channel water routing	Modified Kinematic wave equation*, Diffusive wave equation

Herein only the vertical water balance module (i.e., soil moisture module) is described in detail as this study focused on the assimilation of SMAP soil moisture into the soil moisture module. An extensive description of the remaining modules can be found in Fortin et al. (2001) and Samuel et al.,(2019).

The vertical water balance module is built around two sub-modules: BV3C and CEQUEAU. It simulates vertical distribution of soil moisture in the soil column. In this study, BV3C (Bilan vertical 3 couches) was selected. BV3C vertically discretizes soil column into three layers. The top layer has a depth which is typically 5 to 10 cm, and it is commensurate with the penetration depth of L-band passive microwave satellites such as SMAP. The primary role of this layer is to redistribute the rainfall and/or snow melt into surface runoff and infiltration. Similarly, the second and third layers are responsible for generation of interflow and baseflow, respectively.

The water exchange between these layers is controlled by the one-dimensional Richards equation.

HYDROTEL was calibrated and validated using streamflow observation for the period of 1981-2006 and 2007-2016, respectively. Both calibration and validation were performed at daily time step. The dynamically dimensioned search (DDS) optimization algorithm (Tolson & Shoemaker, 2007) was used for the automatic calibration with objective of maximization of Nash-Sutcliffe efficiency (NSE) (equation (3.1)). In addition, Kling Gupta efficiency (KGE) (equation (3.2)) was used for evaluation of the calibration performance.

$$NSE = 1 - \frac{\sum_{i=1}^{n}\left(Q_{sim,i} - Q_{obs,i}\right)^2}{\sum_{i=1}^{n}\left(Q_{obs,i} - \bar{Q}_{obs,i}\right)^2} \tag{3.1}$$

$$KGE = \sqrt{(r-1)^2 + \left(\frac{\sigma_{sim}}{\sigma_{obs}} - 1\right)^2 + \left(\frac{\mu_{sim}}{\mu_{obs}} - 1\right)^2} \tag{3.2}$$

where Q_{obs} and Q_{sim} are the observed and simulated discharge, respectively, r is linear correlation between observations and simulations, σ_{obs} and σ_{sim} observation and simulated standard deviation, μ_{obs} and μ_{sim}, observation and simulation mean, respectively.

3.4 Experimental design

The framework of this study entails two major successive steps. The first one is downscaling of the SMAP soil moisture from 36-km to a range of high-spatial resolutions including 1, 3 and 9-km and the second one involves the assimilation of the downscaled SMAP soil moisture into HYDROTEL for the purpose of streamflow simulation and forecasting. Each one of them is detailed in the next sub-sections.

3.4.1 Downscaling the SMAP soil moisture

RF, primarily proposed by Breiman (2001), is an ensemble ML technique which is formed from the ensemble of many decision trees. It was selected for the downscaling of the SMAP soil moisture because of its proven performance compared to other variants of ML techniques and other downscaling methods, as discussed in Chapter 2. RF is mainly based on statistical techniques called bootstrap aggregation (bagging), which randomly resamples the training dataset with replacement to generate a set of subsets, also known as bootstrap samples. It utilizes about two-third of each bootstrap samples (in-bag samples) for independently training of each decision tree of the ensemble and the remaining one-third (out-of-bag (OOB) samples) for validation (or testing) of the algorithm. Lastly, the final predicted value was then obtained as the average of each tree's prediction.

In this study, the implementation of RF for downscaling of the SMAP soil moisture involves five sequential steps. These are: 1) Manipulation of the dataset (i.e., predictors and response variable), 2) generation of bootstrap samples, 3) splitting the bootstrap samples, 4) training and testing of RF model, and 5) application of the model. The general framework of these steps is presented in Figure 3.3.

In the **Manipulation** step, basic dataset preprocessing, such as reprojecting the predictors and response variables to the same coordinate system and aggregation of the high-spatial resolution predictors to the resolution of response variable (i.e., resolution of the SMAP soil moisture) has been done. The second step involves **generation of a number of bootstrap samples** by sampling with replacement. In the third step, **splitting the bootstrap samples**, each bootstrap sample was split into training (i.e., two-third of the sample (in-bag sample)) and testing (i.e., one-third of the sample (out-of-bag)) components. The fourth step, **training and testing of RF**, involves training of the RF model with in-bag bootstrap samples and testing with the use of out-of-bag bootstrap samples. The training and testing took place at coarse-spatial resolution (i.e., at resolution of SMAP soil moisture) and applied in the fifth step (i.e., **application of the model**) for predicting of soil moisture at high-spatial resolution from high-resolution predictors by assuming that the developed RF model is spatial scale invariant. In figure 3.3, the green shows the final outcome decided as the output from a given decision tree, whereas the blue shows possible outcome options.

It is also worth noting that 1 km is the highest resolution to which SMAP L3 was downscaled. This is mainly because most of the resolution of predictors are available at 1 km. The predictors with resolution higher than 1 km such as Sentinel-1 were aggregated to 1 km and those lower than 1 km such as SMAP Tb were resampled to 1 km.

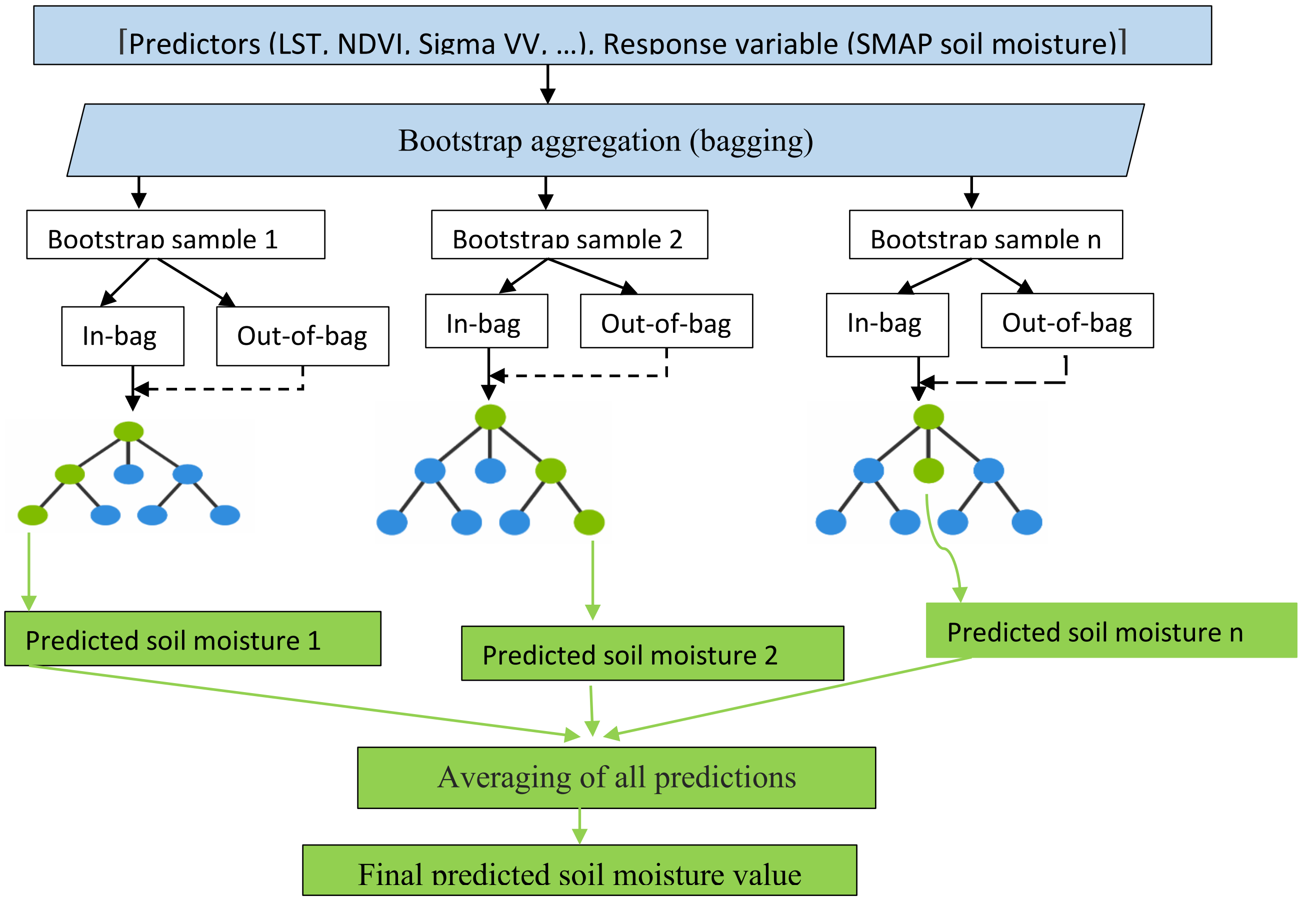

Figure 3.3 General framework for training and testing of RF

Overall, a total of 15 different land surface variables (i.e., predictors) was considered based on their strong association with spatial and temporal variability of soil moisture. These variables were derived from different sources including MODIS, Sentinel-1 and SRTM. Their general association with soil moisture is shown in Table 3.3.

In this study, RF was adopted to construct the relationship between the SMAP soil moisture (i.e., response variable) and a set of land surface variables (i.e., predictors) such as LST and NDVI shown in the Table 3.3.

This can simply be formulated as follows:

$$SM = f(LST, NDVI, Albedo, \dots) \tag{3.3}$$

RF involves tuning of few users defined parameters such as a number of decision trees to grow, maximum depth of the decision tree and maximum number of features selected at each split.

Accordingly, for this study the optimum number of decision trees, maximum depth of the decision tree and maximum number of features were found to be 150, 25 and 4, respectively. These are combination of hyperparameters that gave better performance of RF model. The scikit-learn Python library was used for the implementation of RF (Pedregosa et al., 2011).

It is also worth noting that soil moisture is also highly related to soil type. Soils with smaller particles (silt and clay) have a larger surface area than those with larger sand particles, and a large surface area allows a soil to hold more water. On the other hand, soil with coarse texture has less surface area which leads them to lower water holding capacity compared to fine textured soil.

Table 3.3 The relationship between soil moisture and varieties of land surface variables

Land surface variables	Data sources	General relationship with soil moisture
LST, NDVI, surface albedo	MODIS	• LST have negative relationship with soil moisture and NDVI have positive relationship with soil moisture, notably over sparse to moderately vegetated areas. • Albedo has indirect relationship with soil moisture
Backscatter coefficients (SigmaVV, SigmaVH)	Sentinel-1	• They have direct positive relationship with soil moisture, notably over sparse to moderately vegetated areas
Textural information (Contrast, Entropy, Correlation, Variance)		• To account for vegetation effects and roughness
Topographic derivatives (Elevation, slope and aspect)	SRTM	• To control the spatial distribution of soil moisture
Brightness temperature in V and H polarization	SMAP	• Provides integrated information of factors which affects soil moisture variation such as soil properties, soil temperature and land cover
Precipitation, Antecedent Precipitation Index (API)	PRISM	• Controls spatial and temporal dynamics of soil moisture

Three downscaling experiments were implemented in this study. The first experiment was based on the use of a combination of predictors derived from MODIS optical and Sentinel-1 active sensors. The second experiment was based on the use of predictors derived from MODIS optical sensor. Similar to Experiment 2, the third experiment was based on the use of predictors derived from MODIS optical sensor. However, their main difference is that the length of data record was controlled by availability of Sentinel-1 images (i.e., available every 12 days) for experiment 2, whereas for experiment 3 the length of data record was controlled by availability of SMAP soil moisture which is available every 2 to 3 days. Therefore, experiment 3 has large sample size.

For experiment 1, ascending overpass of SMAP and Sentinel-1 was used because of their temporal proximity, however in experiment 3 descending SMAP overpass was used because of the thermal equilibrium between vegetation canopy and land surface in the morning that result in better soil moisture retrieval quality. For experiment 2, SMAP ascending overpass was used since the main aim of this experiment was for comparative evaluation with experiment 1.

3.4.2 Merging the SMAP and in-situ soil moisture

Merging the strength of the SMAP enhanced (i.e., both the original 9-km and downscaled (1-km resolutions)) and in-situ soil moisture was also explored in this study for the au Saumon watershed which corresponds to the third objective of this thesis. This is because there are in-situ soil moisture measurements with fair spatial distribution in the au Saumon watershed, but for the Susquehanna the number of in-situ is soil moisture measurement stations are very small relative to the size of the watershed which restricted the merging of the SMAP and in-situ soil moisture.

Merging combines the strengths of SMAP enhanced and in-situ soil moisture while compensating for their respective shortcomings. For example, SMAP has the benefit of providing spatial distribution of soil moisture over a large area, while in-situ measurements are scarce spatially, but they are good at preserving the temporal dynamics of soil moisture. Hence, we adopted the conditional merging technique (Sinclair & Pegram, 2005), which is a spatial interpolation technique initially developed to merge radar and rain gauge rainfall measurements. Here, it was used to merge gridded SMAP soil moisture with in-situ measurements. The major advantage of this method is that it preserves the spatial covariance structure of the grid-based measurement (i.e., SMAP soil moisture) while maintaining the accuracy of in-situ measurements.

3.4.3 Assimilation of the SMAP and merged SMAP/in-situ soil moisture

Following the downscaling of the SMAP soil moisture (i.e., objective 1), the next major step was assimilation of the original and downscaled SMAP soil moisture products (i.e., 1-, 3-, and 9-km resolution) into HYDROTEL to evaluate their impact on the streamflow simulation (i.e., the second objective of the thesis). The direct insertion assimilation technique was used. This technique involves replacing the model simulated soil moisture with external observations, e.g., in-situ or remote sensing observations. Accordingly, in this study the model simulated top layer soil moisture was replaced with the SMAP soil moisture (i.e., the original and downscaled). Indeed, the top layer of the HYDROTEL model was adjusted to 5-cm depth to be commensurate with the penetration depth of the SMAP which is about 5-cm. First the model was checked for its sensitivity to change of depth of the first layer before assimilation of SMAP soil moisture. The results show only a very minimal impact on the model's performance. Thus, it is deemed that fixing the model's first layer has no significant impact on model performance compared to the impact of assimilated SMAP soil moisture.

Two cases were considered for the assimilation. The first case involves the assimilation of SMAP soil moisture whenever its available with good quality and full watershed coverage. The second case involves assimilation based on selected high rainfall events. In the second case, assimilation takes place prior to selected high rainfall events (i.e., to mimic operational forecasting), which results in lower frequency of updating as there are not many high rainfall events. Comparison between the two cases was performed to investigate the impact of frequency of updating. The selection of high rainfall events was based on a simple graphical visualization approach, where the watershed averaged rainfall events are plotted along with the corresponding magnitude of streamflow. In this context, only those rainfall events which generated high streamflow events are selected visually.

It is worthwhile to note that prior to assimilation, for both cases, the SMAP soil moisture (i.e., whether its downscaled or the original version) was bias corrected to the model simulated soil moisture using the cumulative distribution function (CDF) matching approach (Reichle & Koster, 2004). This is mainly because of the fact that estimation of the same variable (i.e., soil moisture) by different techniques results in systematics differences because of limitation of each sensing technology. The CDF remove the systematic differences between the two soil moisture data sources to have similar statistical moments such as mean and standard deviation.

The procedure for CDF matching is summarized as follows:

1) Rank the two data sets (i.e.,in-situ and SMAP soil moisture) in the ascending order
2) Calculate the difference between the two data set.
3) The third order polynomial fit was applied to the ranked in-situ soil moisture and calculated differences in step 2.
4) Finally, bias corrected SMAP soil moisture was estimated using polynomial parameters obtained in step 3.

In addition, the impact of assimilation of vertically extrapolated SMAP soil moisture was investigated. The SMAP inherently provides soil moisture information only for the top few centimeters (e.g., 5-cm), but the root zone soil moisture plays an important role in the generation runoff (Brocca et al., 2012). Thus, assimilation of vertically extrapolated SMAP soil moisture is important. This experiment was performed to assess if streamflow forecasts are improved by assimilating near-surface and root zone soil moisture when compared to assimilating near-surface soil moisture only and letting the model to propagate the information in the deeper soil layers. For vertical extrapolation a simple polynomial function which developed between HYDROTEL simulated top and intermediate layer soil moisture was used for the Susquehanna watershed due to paucity of soil moisture measurements. Then, the SMAP surface soil moisture was vertically extrapolated based on the developed polynomial relationship.

Vertical extrapolation methods are usually first calibrated on in-situ soil moisture and then used to vertically extrapolate satellite soil moisture. The exponential filter (ExpF), which is a semi-empirical method originally derived from water balance equation (Wagner et al., 1999), but later presented in a recursive form (Stroud, 1999) for surface soil moisture extrapolation was used for the vertical extrapolation of SMAP soil moisture (equation (3.4)) in the au Saumon watershed because of availability of in-situ measurements for this watershed.

$$SMrz_{tn} = SMrz_{tn-1} + K_n(SMsurf_{tn} - SMrz_{tn-1}) \tag{3.4}$$

$$K_n = \frac{K_{n-1}}{K_{n-1} + exp\left(\frac{-1}{T}\right)} \tag{3.5}$$

where SMsurf and SMrz are surface and root zone soil moisture, T is the optimal characteristic decay time, K is gain and tn is time step.

The HYDROTEL model was basically developed to accommodate remote sensing information such as soil moisture and snow water equivalent (Fortin et al.,1991). That is why the depth of the top layer was fixed to a range of 5-10 cm by design. In addition, the model did not consider vertical variation of hydraulic conductivity of the soil. This could be attributed to less sensitivity of the model to depth variation. In addition, our preliminary test on sensitivity analysis of the parameters shows that the model appears less sensitive to depth variation.

In another experiment, the impacts of the relation between the spatial resolution of SMAP soil moisture and watershed size on the accuracy of model streamflow simulation was examined. In this context, the Susquehanna (i.e., larger in size) and the Upper Susquehanna (i.e., relatively smaller in size) watersheds are considered. The HYDROTEL model was then updated with the original and downscaled SMAP soil moisture separately for both watersheds and the impact of assimilation of different resolutions of the SMAP soil moisture on the model performance is examined.

On the other hand, separate experiments were implemented for the au Saumon watershed, which corresponds to the third objective of this thesis. As opposed for the Susquehanna, for the au Saumon watershed the merged SMAP/in-situ soil moisture with resolutions of 1- and 9-km was assimilated into HYDROTEL model. Besides the merged SMAP/in-situ soil moisture, the downscaled SMAP enhanced (i.e., 1-km) and spatially interpolated in-situ soil moisture were separately assimilated into the model. In addition, the vertically extrapolate version of these products were assimilated.

The general methodological framework of this thesis is shown in Figure 3.4.

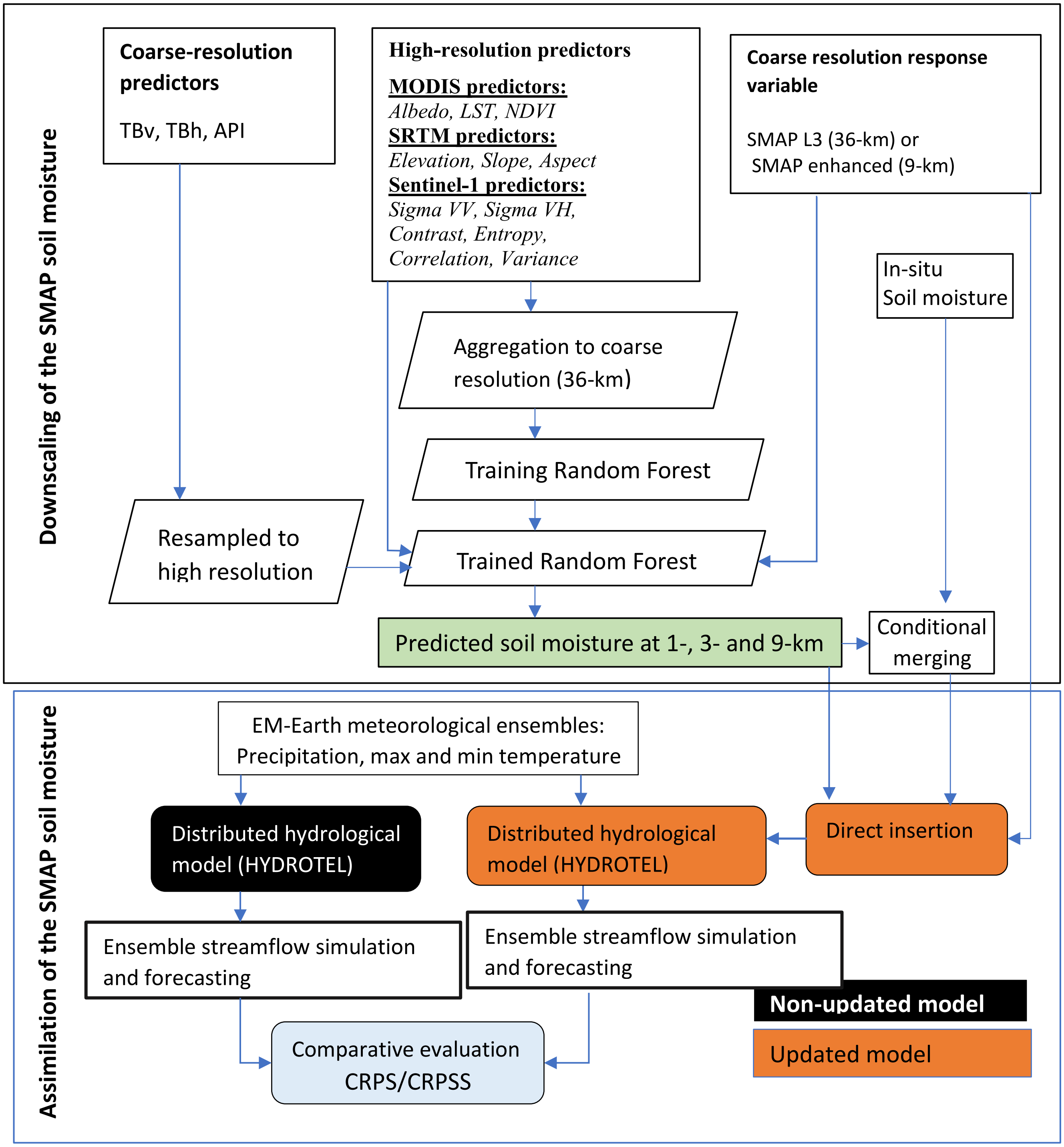

Figure 3.4 Methodological framework

3.4.4 Metrics

Validation of downscaled SMAP soil moisture

Besides the graphical comparison between the downscaled SMAP and in-situ soil moisture, several popular classical statistical metrics including the Pearson correlation coefficient (R) (equation (3.6)), bias (equation (3.8)), and unbiased root mean square error (ubRMSE) (equation (3.9)) were used to quantitatively evaluate the agreement of the downscaled SMAP soil moisture against in-situ soil moisture observations. For the Susquehanna watershed in-situ soil moisture observation collected from SCAN and USCRN networks were used, whereas for the au Saumon watershed in-situ soil moisture collected by our research team was used. The metrics are described by the following equations:

$$\mathrm{R} = \frac{\Sigma\left(\theta_{\mathrm{DSM,i}} - \overline{\theta_{\mathrm{DSM}}}\right)\left(\theta_{\mathrm{insitu,i}} - \overline{\theta_{\mathrm{insitu}}}\right)}{\sqrt{\Sigma\left(\theta_{\mathrm{DSM,i}} - \overline{\theta_{\mathrm{DSM}}}\right)^2\left(\theta_{\mathrm{DSM,i}} - \overline{\theta_{\mathrm{DSM}}}\right)^2}} \tag{3.6}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\theta_{DSM} - \theta_{insitu})^2}{n}} \tag{3.7}$$

$$Bias = \frac{\sum_{i=1}^{n}\theta_{DSM} - \theta_{insitu}}{n} \tag{3.8}$$

$$ubRMSE = \sqrt{(RMSE)^2 - (bias)^2} \tag{3.9}$$

where $\theta_{\mathrm{DSM,i}}$ and $\theta_{\mathrm{insitu,i}}$ is downscaled and in-situ soil moisture, respectively, n is length of soil moisture data record.

Evaluation of the benefit of soil moisture assimilation

The impacts of assimilation of SMAP soil moisture and merged SMAP/in-situ soil moisture on model performance in predicting the ensemble streamflow simulations and forecasting were evaluated by means of the efficiency index (Eff), the Normalized root mean square error (NRMSE) and the Continuous Ranked Probability Skill Score (CRPSS). The Eff measures the effect of assimilation with respect to the open loop run (equation (3.10)). A positive value of

Eff indicates improvement in the accuracy of the ensemble streamflow simulations because of the assimilation of the SMAP soil moisture, whereas the negative value indicates deterioration.

$$Eff = \left(1 - \frac{\sum_{t=1}^{T}\left(Q_{\overline{up},t} - Q_{obs,t}\right)^2}{\sum_{t=1}^{T}\left(Q_{\overline{ol},t} - Q_{obs,t}\right)^2}\right) \tag{3.10}$$

where Q_{obs} is observed streamflow, $Q_{\overline{up}}$ and $Q_{\overline{ol}}$ the mean of ensemble of streamflow by the updated and non-updated (ol) model, respectively.

The NRMSE represents the ratio of the root mean square error (RMSE) between the predicted ensemble streamflow by the updated model and observed streamflow to the RMSE between the open loop ensemble and observed streamflow (equation (3.11). A score greater than 1 indicates deterioration and vice versa. NRMSE provides information about the ensemble spread and the ensemble mean performance.

$$NRMSE = \frac{\frac{1}{n}\sum_{i=1}^{n}\sqrt{\sum_{t=1}^{T}\left(Q_{up,t} - Q_{obs,t}\right)^2}}{\frac{1}{n}\sum_{i=1}^{n}\sqrt{\sum_{t=1}^{T}\left(Q_{ol,t} - Q_{obs,t}\right)^2}} \tag{3.11}$$

where n and T are ensemble size and length of time series considered, respectively.

In addition, the CRPSS was used to assess the skill of the ensemble streamflow simulation (equation (3.12)). Its score varies between $-\infty$ and 1. A score greater than zero indicates an improvement in accuracy due to assimilation of the SMAP soil moisture and vice versa.

$$CRPSS = 1 - \frac{CRPS_{updated}}{CRPS_{non-updated}} \tag{3.12}$$

$$CRPS = \frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{\infty}\left[\,F_i(x)\; - F_i^0(x)\,\right]^2 dx \tag{3.13}$$

$$F^0(x) = \begin{cases} 0, x \leq \; observed\ streamflow\ value \\ 1, x > \; observed\ streamflow\ value \end{cases}$$

where the CRPS is the Continuous Ranked Probability Score, $F_i(x)$ and $F^o_i(x)$ is probability density function of the member of the ensemble simulations and observation, respectively.

NOTE TO THE READER

This is thesis in a « format par articles ». It's core is composed of three papers which are sequentially presented in Chapter 4, Chapter 5, and Chapter 6. The following diagram presents a general overview of each paper and their connection with each other.

Paper # 1

- involves downscaling of the SMAP L3 soil moisture from 36-km resolution to a range of resolutions (1-, 3-, and 9-km) over the Susquehanna watershed using a set of predictors derived from MODIS, Sentinel-1 and SRTM using random forest machine learning technique.

- In addition, SMAP enhanced soil moisture (9-km resolution) was downscaled to 1-km over the au Saumon watershed (results presented in the article 3)

Paper # 2:

- involves assimilation of the original and downscaled SMAP L3 soil moisture (i.e., 1-, 3-, and 9-km resolution) into HYDROTEL for improving the ensemble streamflow simulations.

- Two watersheds were considered: 1) Susquehanna (larger size) and 2) Upper Susquehanna (relatively smaller in size and its head sub-watershed of the Susquehanna watershed).

Paper # 3:

- involves merging of the SMAP enhanced and in-situ soil moisture

- assimilation of the merged SMAP/in-situ soil moisture into HYDROTEL for ensemble streamflow forecasting over the au Saumon watershed

Chapter 4 ENHANCING SPATIAL RESOLUTION OF SMAP SOIL MOISTURE PRODUCTS

Avant-propos

Auteurs et affiliation

- Shimelis Asfaw Wakigari : étudiant au doctorat, Département de Génie Civil, Faculté de Génie, Université de Sherbrooke.
- Robert Leconte : professeur, Département de Génie Civil, Faculté de Génie, Université de Sherbrooke.

Date d'acceptation : 2 février 2022.

État de l'acceptation : Version finale publiée

Revue : Remote Sensing

Référence : Wakigari, S. A., & Leconte, R. (2022). Enhancing Spatial Resolution of SMAP Soil Moisture Products through Spatial Downscaling over a Large Watershed: A Case Study for the Susquehanna River Basin in the Northeastern United States. *Remote Sensing*, *14*(3), 776; **https://doi.org/10.3390/rs14030776**

Titre français: Amélioration de la résolution spatiale de l'humidité du sol issue de SMAP par une désagrégation d'échelle spatiale sur un grand bassin versant : cas d'étude du bassin de la rivière Susquehanna au nord-est des États-Unis

Contribution au document : Cet article propose une désagrégation d'échelle à fine résolution des produits d'humidité du sol de SMAP. Des prédicteurs issus de différentes sources dont MODIS, Sentinel-1, SMAP et SRTM DEM sont assemblés en différentes combinaisons. Le modèle d'apprentissage automatique utilisé est une forêt aléatoire. Une série d'essais a été menée pour déterminer la meilleure combinaison de prédicteurs à utiliser pour la région d'étude. Le chapitre 4 de la thèse rapporte les résultats de cet article.

Résumé français

L'utilisation de données d'humidité du sol (HU) à haute résolution spatiale est primordiale pour de nombreuses applications en agriculture et en hydrologie tant à l'échelle locale que régionale. Grâce aux développements de produits satellitaires en micro-ondes passives en bande L, des mesures d'HU en surface sont disponibles à une couverture globale, chose que n'offrent par les mesures in-situ. Cependant, leur usage est souvent limité par leur faible résolution spatiale. Afin de pallier ce problème, cette recherche met en place des modèles de forêts aléatoires (FA) pour faire une désagrégation d'échelle à une résolution d'un km des produits d'humidité du sol SMAP niveau-3 (L3SMP, 36-km) et SMAP amélioré (L2SMP_E, 9-km). Les prédicteurs utilisés dans ces modèles de désagrégation d'échelle sont issus de Sentinel-1 bande C-SAR et de MODIS. Les arbres de FA sont entraînés et validés individuellement pour les deux résolutions spatiales initiales (36- et 9-km) selon deux approches: 1) en utilisant une combinaison de prédicteurs issus de MODIS, Sentinel-1 et autres tels que l'élévation et la température de brillance, 2) en utilisant les mêmes prédicteurs qu'en 1) sauf ceux issus de Sentinel-1. La comparaison entre les deux approches est faite uniquement aux dates pour lesquelles les produits Sentinel-1 sont disponibles. On observe alors que le retrait des données issues de Sentinel-1 dans la 2nde approche n'influence que peu le coefficient R qui passe de 0,84 pour l'approche 1) à 0,83 pour l'approche 2) pour les produits initialement à 36-km de résolution et de 0,91 à 0,86 pour ceux à 9-km de résolution. Il ressort aussi que la température de brillance en polarisation V est le prédicteur le plus important, suivi du NDVI, de l'albédo de surface et de l'API. Au contraire, les prédicteurs issus de Sentinel-1 jouent un rôle bien moins important, avec une très faible contribution dans l'amélioration du pouvoir prédictif des modèles de FA. Des limites sont communes aux deux expériences, dont notamment la faible taille de l'échantillon d'apprentissage qui vient du manque d'images Sentinel-1 (le temps de revisite du satellite étant de 10 jours). Par conséquent, pour contourner ce problème, une troisième approche est développée, dans laquelle aucun prédicteur de Sentinel-1 n'est inclus dans la phase d'apprentissage (comme dans l'approche 2)) mais où on utilise toutes les données à disposition, soit une par jour. Les résultats avec cette 3^{e} approche montrent une bonne correspondance entre les produits d'HU améliorés à fine échelle L3SMP et L2SMP_E issus des deux résolutions spatiales (36- et 9-km) et les mesures in-situ. De plus, l'HU à faible résolution des deux produits SMAP distinguent mieux les effets locaux tout en gardant les motifs spatiaux observés à leur résolution originale. Quelle que soit la résolution spatiale du produit SMAP utilisée pour initier la désagrégation d'échelle, les résultats à fine échelle sont de qualité similaire. Finalement, cette étude révèle des résultats prometteurs sur la désagrégation d'échelle des produits d'humidité d'HU issus de SMAP appliqués à des bassins humides à fortes dominantes forestières soumis à des étés chauds.

Mots-clés : humidité du sol; désagrégation d'échelle; validation; Sentinel-1; MODIS; SMAP; forêt aléatoire

Note : À la suite des corrections demandées par les membres du jury, le contenu de cet article diffère de celui qui a été accepté.

ENHANCING SPATIAL RESOLUTION OF SMAP SOIL MOISTURE PRODUCTS THROUGH SPATIAL DOWNSCALING OVER A LARGE WATERSHED: A CASE STUDY FOR THE SUSQUEHANNA RIVER BASIN IN THE NORTHEASTERN UNITED STATES

4.1 Abstract

Soil moisture with high-spatial resolution plays a paramount role in many local and regional hydrological and agricultural applications. The advent of L-band passive microwave satellites allowed for it to be possible to measure near-surface soil moisture at a global scale compared to in-situ measurements. However, their use is often limited because of their coarse-spatial resolution. Aiming to address this limitation, random forest (RF) models are adopted to downscale the SMAP level-3 (L3SMP, 36-km) and SMAP enhanced (L3SMP_E, 9-km) soil moisture to 1-km. A suite of predictors derived from the Sentinel-1 C-band SAR and MODIS is used in the downscaling process. The RF models are separately trained and verified at both spatial scales (i.e., 36- and 9-km) considering two experiments: 1) using predictors derived from the MODIS and Sentinel-1 along with other predictors such as elevation and brightness temperature and 2) using all predictors of the first experiment except for the Sentinel-1 predictors. Only dates when the Sentinel-1 images were available are considered for the comparison of the two experiments. The comparison of the results of the two experiments indicates that the removal of Sentinel-1 predictors from the second experiment only reduces the R value from 0,84 to 0,83 and from 0,91 to 0,86 for 36 and 9-km spatial scales, respectively. Among the predictors used in the downscaling, the brightness temperature in V polarization was identified as the most important predictor, followed by NDVI, surface albedo and API. On the contrary, the Sentinel-1 predictors play a less important role with no marked contribution in enhancing the predictive accuracy of RF models. In general, the two experiments have limitation, such as a small sample size for the training of the RF model because of the scarcity of Sentinel-1 images (i.e., revisit time of 12 days). Therefore, based on this limitation, a third experiment is proposed, in which the Sentinel-1 predictors are not considered at all in the

training of the RF models. The results of the third experiment show a good agreement between the downscaled L3SMP and L3SMP_E SM, and in-situ soil moisture measurements at both spatial scales. In addition, the temporal availability of the downscaled soil moisture increased. Moreover, the downscaled soil moisture from both SMAP products presented greater spatial detail while preserving the spatial patterns found in their original products. The use of the two SMAP soil moisture products as background fields for the downscaling process does not show marked differences. Overall, this study demonstrates encouraging results in the downscaling of SMAP soil moisture products over humid climate with warm summers dominated by vegetation.

Keywords: soil moisture; downscaling; validation; Sentinel-1; MODIS; SMAP; random forest

4.2 Introduction

Soil moisture as a key hydrological variable, is of paramount importance in modulating terrestrial water cycle components such as precipitation by partitioning into surface runoff and infiltration; subsequently, affecting the streamflow and ground water recharge (Tuttle & Salvucci, 2014). It also affects regional and global climate systems through its control of the interaction between land surface and atmosphere via evapotranspiration, thereby affecting air temperature and precipitation (Entekhabi, 1995; Seneviratne et al., 2010b). In addition, knowledge of soil moisture is important in agriculture for crop growth (Lakhankar et al., 2009; Vereecken et al., 2008) . Similarly, a large number of operational hydrological and agricultural applications, including flood forecasting (Massari et al., 2018a; Wanders et al., 2014), drought monitoring (Park et al., 2017; Velpuri et al., 2015) and irrigation planning (Sharif et al., 2015), are heavily dependent on soil moisture. Hence, the proper characterization of the spatial and temporal variations of soil moisture is imperative.

The spatio-temporal variability of soil moisture is often ascribed to multifaceted interactions of numerous environmental factors, including soil properties, surface topography, vegetation and atmospheric forcing (Crow et al., 2012a). Traditionally, this variability is monitored using in-situ measurements which offer an accurate estimate of soil moisture at different depths with a high-temporal resolution (Dobriyal et al., 2012; Dorigo et al., 2011). However, despite its accuracy, this method is costly, time consuming and, most importantly, suffers from low spatial representativeness (i.e., only in the order of a few centimeters). For these reasons, over the last few decades, a number of satellite remote sensing platforms have been launched to address the need for economically feasible global information on soil moisture with a temporal frequency of a few days to a few weeks.

Among satellites, remote sensing, active and passive microwave remote sensing satellites have demonstrated the potential for measuring near surface soil moisture under all-weather conditions depending on the underlying land surface conditions (Petropoulos et al., 2015; Schmugge et al., 1986b). Active microwave remote sensing has the capability to provide soil moisture at a high-spatial resolution (i.e., in the order of tens of meters) (Torres et al., 2012), but it is heavily affected by perturbing factors such as surface roughness and vegetation, thereby complicating soil moisture retrieval (Petropoulos et al., 2015). Comparatively, passive microwave remote sensing has long been recognized as a prominent tool for the retrieval and monitoring of surface soil moisture, owing to its sensitivity to near surface soil moisture dynamics and the capability to penetrate to a certain extent through the vegetation canopy, but suffers from its low spatial resolution (i.e., in the order of tens of kilometers) (Jackson & Schmugge, 1989; Njoku & Entekhabi, 1996).

Over the last few decades, several passive microwave satellites with different frequencies (e.g., L, C and X) have been launched, providing global coverage of surface soil moisture information. Among the several frequencies, the L-band passive microwave satellites (e.g., Soil Moisture and Ocean Salinity (SMOS) and Soil Moisture Active Passive (SMAP)) are the most promising and primarily devoted to global monitoring of surface soil moisture (Kerr et al., 2001; Narayan et al., 2006; Owe et al., 2001).

While the advent of L-band passive microwaves satellites has effectively addressed the difficulties associated with the continuous monitoring of near surface soil moisture over large areas, the spatial resolution of soil moisture products derived from these satellites is in the order of several kilometers (i.e., 25-50-km), which impedes their utilization for a large variety of local and regional applications (Wei et al., 2019; Zhao et al., 2018), such as agriculture and hydrology. Therefore, a detailed representation of sub-grid-scale spatial variability of soil moisture within the coarse-spatial resolution of passive microwave soil moisture products would significantly benefit these applications.

In order to address the need for spatially detailed soil moisture information, a growing number of studies have attempted to provide high-spatial-resolution soil moisture maps through downscaling by leveraging the complementary strength of passive microwave and radar and/or optical/thermal observations (Fan et al., 2018; Hu et al., 2020; Rudiger et al., 2016; Sánchez-Ruiz et al., 2014; Senanayake et al., 2021; Wei et al., 2019; Wu et al., 2011; Zhan et al., 2017; Zhao et al., 2018). In this context, downscaling takes advantage of high-spatial-resolution land surface variables derived from radar and/or optical/thermal sensors to disaggregate the coarse resolution of passive microwave soil moisture products.

A variety of downscaling techniques has been proposed and tested, which can be roughly grouped into three categories based on the sources of high-resolution land surface variables used in downscaling (Peng et al., 2017b; Sabaghy et al., 2018). These are (1) optical/thermal-based downscaling, (2) radar-based downscaling and (3) radiometer-based downscaling. In addition, there are also emerging techniques such as machine learning and data assimilation. Sabaghy et al.,(2018) summarized the performances of these downscaling techniques reported by many studies conducted in different regions of the world under varying climate and land surface conditions. Based on their comparison, optical/thermal-based downscaling was identified as the most widely applied technique to downscale passive microwave soil moisture, although with varying degrees of accuracy, whereas radar-based downscaling and ML models are considered the two most promising approaches for soil moisture downscaling with a higher degree of accuracy. An extensive review of these techniques can be found in Peng et al., (2017) and Sabaghy et al.,(2018).

The optical/thermal-based downscaling generally relies on the triangular/trapezoidal feature space established between vegetation indexes and land surface temperature derived from optical/thermal satellites (e.g., MODIS) over heterogeneous land surfaces (Carlson, 2007; Moran et al., 1994). The premise of this technique is that the sensitivity of soil moisture to land surface temperature varies under a wide range of land surface conditions (Peng et al., 2015). The empirical polynomial fitting and semi-physical evaporation-based approaches are among the techniques that fall under this category. A large number of studies have applied these techniques to improve the spatial resolution of passive microwave observations of soil moisture (e.g., Chauhan et al., (2003); Choi & Hur, (2012); Merlin et al., (2012)). For instance, Piles et al. (2011) applied the polynomial function to downscale SMOS soil moisture from 40-km to 10- and 1-km. Likewise, Djamai et al. (2015) applied a semi-physical model called DISPATCh (DISaggregation based on Physical And Theoretical scale CHange) to downscale SMOS soil moisture to 1-km over the Canadian Prairies.

Another alternative approach is the radar-based downscaling technique, which synergistically combines coarse-resolution passive microwave soil moisture or brightness temperature with high-resolution synthetic aperture radar (SAR) observations to enhance the spatial detail of soil moisture. A typical example of this technique is the SMAP baseline downscaling algorithm (Das et al., 2014) , which has been utilized by NASA on the SMAP mission to downscale the coarse-scale passive microwave brightness temperature (36-km) to an intermediate resolution (9-km) by making use of the high-resolution (3-km) SAR backscatter, which was then inverted to retrieve soil moisture at a 9-km spatial resolution. In addition, there are several other radar-based downscaling techniques, including the SMAP optional method (Das et al., 2011), the change detection method (Piles et al., 2009) and the Bayesian framework (Zhan et al., 2006). Despite the great potential of radar-based downscaling in improving the spatial resolution of soil moisture, the lower revisit frequency (e.g., 6-12 days) and lack of a temporally collocated passive microwave radiometer and SAR measurements is still a big challenge after the failure of the SMAP radar. However, in order to continue the production of high-spatial resolution soil moisture by the SMAP mission, significant efforts have been undertaken to substitute the SMAP radar with an active radar from other satellites (Das et al., 2016). Among the existing radar satellites, the European Space Agency's (ESA) Sentinel-1A/1B satellites were considered suitable due to the similarity of their orbital configuration with that of SMAP (Das et al., 2019; He et al., 2018).

Apart from the above downscaling methods, ML has recently received a great deal of attention for the downscaling of passive microwave soil moisture by learning the relationship between soil moisture and a set of land surface variables (Abbaszadeh et al., 2019; Zhao et al., 2018). It has become increasingly popular mainly due to (1) its ability to handle the complex non-linear relationship between soil moisture and a large set of land surface variables across a variety of heterogeneous surface conditions, (2) the rapid increase in computational power and public availability of remote sensing data and (3) the relatively good performance as compared to aforementioned downscaling techniques.

Numerous studies have reported successful applications of various ML techniques, e.g., artificial neural network (ANN), classification and regression trees (CART), gradient-boosting decision tree (GBDT) and random forest (RF), in the downscaling of passive microwave soil moisture products (Liu et al., 2018; Lv et al., 2021; Wei et al., 2019; Zhao et al., 2018). Among these techniques, a large number of studies have shown superiority of RF because of its potential to greatly reduce the problem of overfitting (e.g., (Abbaszadeh et al., 2019; Im et al., 2016; Liu et al., 2020)). For example, Zhao et al., (2018) downscaled SMAP soil moisture to 1-km over the Iberian Peninsula using RF and found that the RF had the capability to capture the spatial variability of soil moisture with great accuracy compared to the polynomial function fitting approach. Similarly, Liu et al.,(2020) compared six variants of ML methods and revealed that RF showed better performance than all others.

In general, the application of aforementioned downscaling techniques enhances the spatial resolution of passive microwave soil moisture with varying degrees of accuracy. However, most of these techniques are predominantly applied to the arid and semi-arid climatic regions, and their performances are often site-specific (e.g., Hu et al., 2020; Senanayake et al., 2017; Yin et al., 2020; Zhao et al., 2018), requiring further investigation across diverse geographical and climatic regions with a range of land cover types.

In the present study, RF models are applied for the downscaling of two SMAP soil moisture products (i.e., L3SMP (36-km) and L3SMP_E (9-km)) to a 1-km spatial resolution over the warm summer humid climate of the Susquehanna watershed. Studies have exploited the potential strengths of a suite of high-resolution land surface variables and indices derived from optical/thermal remote sensing such as the land surface temperature (LST), normalized difference vegetation index (NDVI), enhanced vegetation index (EVI), surface albedo and other auxiliary variables (e.g., precipitation, soil texture and surface topography) in the downscaling of passive microwave satellite soil moisture using RF (Bai et al., 2019; Hu et al., 2020; Zhao et al., 2018). However, the use of synthetic aperture radar (SAR) observations, such as those provided by the Sentinel-1 C-band SAR, as predictors notably in RF-based downscaling is often limited (Bai et al., 2019) . In other words, although a number of other approaches were proposed to merge C-band SAR and passive microwave soil moisture, the use of SAR information in RF-based downscaling is limited. As a result, it is necessary to clarify the contribution of SAR products (i.e., backscattering intensity and its textural information) along with other remotely-sensed products in the downscaling of SMAP soil moisture products through comprehensive assessments.

Considering the abovementioned premises, the purpose of the present study is to: 1) assess the performance of the RF-ML technique in the downscaling of SMAP soil moisture products over the warm summer humid climate of a large watershed with a variety of land cover types with a predominance of forests; 2) exploit the benefit of the high-resolution Sentinel-1 SAR backscattering intensity and its textural information along with MODIS products and other auxiliary variables in downscaling SMAP soil moisture products; 3) identify the best set of predictor variables for downscaling for underlying surface and climate conditions of the study area; 4) compare the downscaled and the existing 1-km high-resolution SMAP/Sentinel-1 soil moisture products against ground-based in-situ soil moisture and precipitation; 5) compare the relative merits of the SMAP level-3 (36-km) versus SMAP enhanced (9-km) soil moisture as background field in downscaling.

4.3 Materials and Methods

4.3.1 Study area

The study area, Susquehanna River Basin, is located in the eastern part of the United States, with an area of 71,255 km^2 (Figure 4.1).The basin experiences high flows during spring season due to combined effect of snowmelt and rainfall, while low flows occur in late summer and early fall. Its elevation varies between 4 and 1150 m (DePhilip & Moberg, 2010). The watershed is characterized by a mild, subtemperate and humid climate marked by cold winters with snow events and warm to hot summers. The average annual precipitation ranges from approximately 800 to 1250-mm, while the mean annual temperature is roughly 6 and 12 ^{0}C in the northern and southern portion of the basin, respectively. The basin is mainly dominated by forests (55 %), followed by agriculture (33 %) (DePhilip & Moberg, 2010). Its elevation varies between 4 and 1150 m.

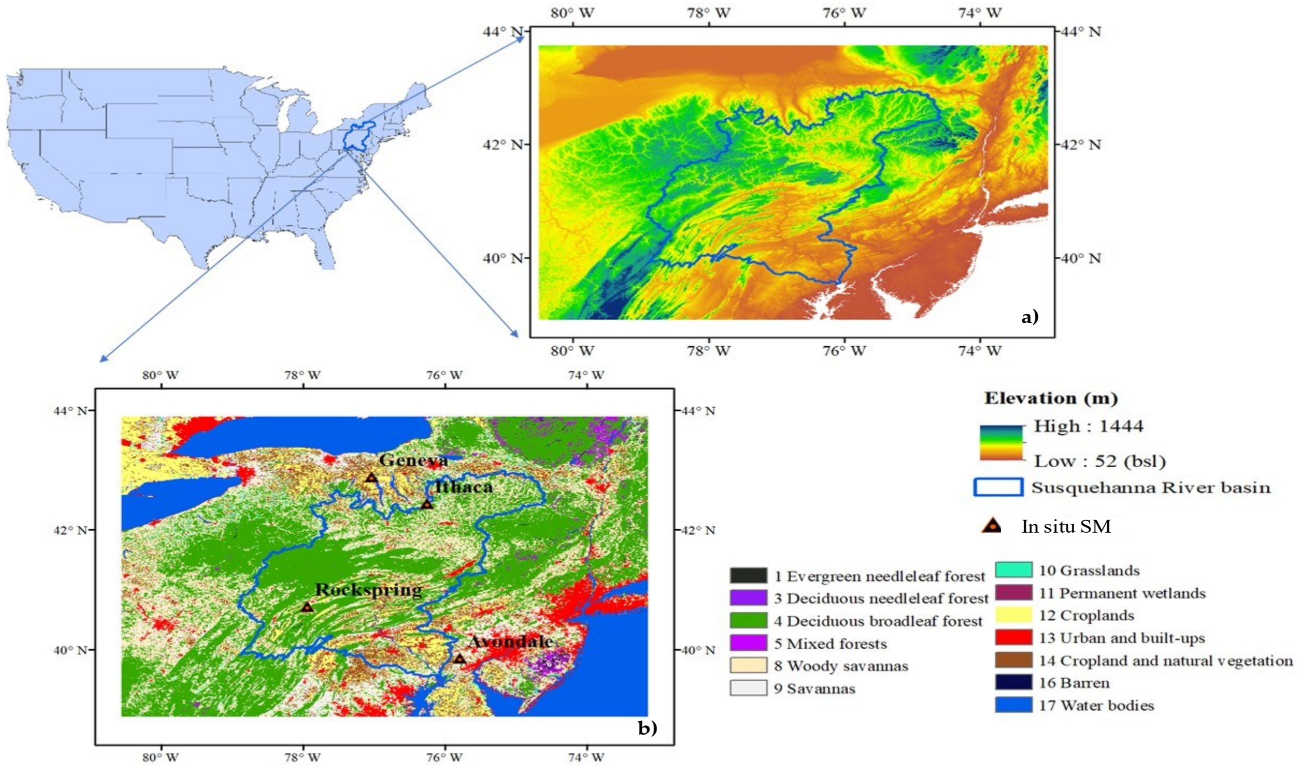

Figure 4.1 Study area: (a) DEM and (b) land cover

4.3.2 Description of data

The dataset used in this study was derived from a variety of sources, including: 1) remote sensing satellites, e.g., SMAP, MODIS, Sentinel-1; 2) models, e.g., PRISM, DEM; 3) in-situ measurements (e.g., ISMN). The data collected during unfrozen period (i.e., May to October) of 2017 to 2020 were used to circumvent the poor quality of SMAP soil moisture during the remaining months of the year because the soil was likely to be frozen near the surface during winter. A brief description of the main features of this data is presented as follows.

<u>SMAP soil moisture and brightness temperature</u>

The Soil Moisture Active Passive (SMAP) mission is the latest L-band satellite mission launched on January 31, 2015, by the National Aeronautics and Space administration (NASA). It was primarily dedicated to measuring near surface soil moisture and landscape freeze/thaw conditions below the ground surface (top ~5 cm). It is a sun-synchronized near-polar orbit satellite with local equator overpass times of 6:00AM (descending) and 6:00PM (ascending). SMAP incorporates both radar (active) and radiometer (passive) on the same platform, aiming to offer soil moisture at range of spatial resolutions, including 3-, 9- and 36-km. Nevertheless, due to SMAP radar failure that occurred on July 7, 2015, the generation of the 3-km (active) and 9-km (active–passive) soil moisture products was halted. Consequently, in order to continue the generation of soil moisture at aforementioned resolutions, the SMAP mission proposed: 1) to replace the 9-km active–passive soil moisture with SMAP enhanced soil moisture, which was retrieved from the inversion of interpolated coarse-resolution SMAP L3 brightness temperature using the Backus–Gilbert optimal interpolation technique; 2) to replace the original SMAP 3-km soil moisture products with downscaled SMAP enhanced soil moisture using Sentinel-1 SAR backscatter.

In this study, the SMAP level-3 (L3SMP) and SMAP enhanced (L3SMP_E) soil moisture and brightness temperature with a resolution of 36- and 9-km, respectively, were used. These products had temporal resolution of 2-3 days, posted on the 36- and 9-km Equal-Area Scalable Earth (EASE) grid, respectively. In addition, 1-km high-spatial-resolution SMAP/Sentinel-1 (L2_SM_SP) images were collected to compare with the downscaled SMAP soil moisture product in this study. L2_SM_SP had a temporal resolution of 12 days. These data can be downloaded freely from the NASA Snow and Ice Data Center Distributed Active Archive Center (NSIDC DAAC) (https://nsidc.org/data/smap/ accessed date 30 May 2019).

Sentinel-1 backscatter and textural features

Sentinel-1 is a constellation mission composed of two satellites (i.e., Sentinel-1A and Sentinel-1B), each carrying an imaging C-band (5,405 GHz) synthetic aperture radar (SAR) instrument in a near-polar sun-synchronous orbit to monitor the land and ocean (Torres et al., 2012) . It has four data acquisition modes: stripmap, interferometric wide (IW) swath, extra-wide (EW) swath and wave. The IW swath is the prime acquisition mode over land surfaces, which provides observations at spatial and temporal resolution of 10-m and 6-12 days, respectively, in both VV and VH polarizations. The incidence angle of Sentinel-1 was normalized to have a similar incidence angle as that of the SMAP incidence angle, which is 40^o. This is done to correct the effects of the unequal incidence angles.

In the present study, the Sentinel-1 images collected in IW mode during ascending overpass were used. The post-processing of the images, including thermal noise removal, speckle filtering, radiometric calibration, terrain correction and normalization, was performed using the

Sentinel application platform (SNAP) tool. In addition, four less correlated textural features, namely, contrast, entropy, correlation and variance, were extracted from the Sentinel-1 images using the grey-level co-occurrence matrix (GLCM) provided by SNAP. These features were extracted with a moving window of 5x5 and average orientation of all directions (0°, 45°, 90° and 135°). The Sentinel-1 imagery is openly available to users through ESA's Sentinel-1 internet data hub (https://scihub.esa.int/ accessed date 25 June 2019).

MODIS products

The Moderate Resolution Imaging Spectroradiometer (MODIS) sensor on board NASA's Earth Observing System (EOS) Terra and Aqua satellites is widely used for the monitoring and understanding of the spatiotemporal dynamics of land, ocean and atmosphere at the surface (Justice et al., 2002). It operates on a solar synchronous polar orbit with local equator overpass times of around 10:30AM (descending) and 10:30 PM (ascending) for Terra and 1:30 AM (descending) and 1:30PM (ascending) for Aqua. It provides a suite of global products every 1-2 days with spatial resolution ranging from 250-m to 1000-m. In this study, the latest version (i.e., Collection 6) of MODIS land surface variables which had a strong connection with soil moisture, including 1-km daily Terra and Aqua LST (MYD11A1, MOD11A1), 1-km 16-day vegetation index (MOD11A2) and 500 m daily surface albedo (MCD43A3), were retrieved from the NASA Land Processes Distributed Active Archive Center (LP DAAC) (https://lpdaac.usgs.gov/ accessed date 28 June 2019).

Surface topography

Topographic attributes, including elevation, slope and aspect, were extracted from the digital elevation data obtained from the NASA Shuttle Radar Topography Mission (SRTM) digital elevation model at 30-m resolution. These data were downloaded from the Land Processes Distributed Active Archive center (https://lpdaac.usgs.gov/ accessed date 1 July 2019).

Precipitation and antecedent precipitation index (API)

The spatiotemporal variability of soil moisture is highly affected by precipitation patterns. In addition, precipitation that occurred over preceding days has a considerable effect on soil moisture (i.e., soil moisture memory) of a given day. In this study, daily precipitation data obtained from Parameter–elevation Relationships on Independent Slopes Model (PRISM) (Di Luzio et al., 2008) at a resolution of 4-km were used to calculate the modified antecedent precipitation index (modified API), aiming to include the effect of precipitation of the past few days on downscaled SMAP soil moisture of a given day. This data is available covering the entire conterminous United States (http://www.prism.oregonstate.edu/ accessed date 3 July 2019).

In-situ soil moisture

In-situ soil moisture measurements of the top 5 cm depth of the soil column and precipitation data were collected from the soil climate analysis network (SCAN) and the U.S. climate reference network (USCRN). There were about 4 soil moisture probes which lay within the study area, including Avondale, Ithaca, Geneva and Rocksprings (Figure 4.1). The probes provided hourly values of volumetric soil moisture. The dominant land cover types for Avondale and Geneva were mosaic herbaceous and grassland, respectively, whereas for Ithaca and Rocksprings it was crop. The average annual precipitation for the four stations varied from 850 to 1280 mm (i.e., Geneva (855 mm), Rockspring (1015 mm), Ithaca (1100 mm) and Avondale (1280 mm)), whereas average annual temperature ranged from 8 to 12 ^{0}C (i.e., Ithaca (8 ^{0}C), Geneva (9 ^{0}C), Rockspring (10 ^{0}C) and Avondale (12 ^{0}C)). The elevation of the probe locations varied from 122 to 374 m (i.e., Avondale (122 m), Geneva (221 m), Rockspring (372 m) and Ithaca (374 m). These data were collected from the International Soil Moisture Network (ISMN) database (https://ismn.geo.tuwien.ac.at/en/ accessed date 10 July 2019).

The basic characteristics of dataset, such as spatial and temporal resolutions, used in this study were summarized in Table 4.1.

Table 4.1 Summary of data used for the downscaling process in this study.

Datasets (or product name)	Variables/Indices	Temporal Resolution(day)	Spatial Resolution (km)
SMAP L3 SM	Soil moisture		36
SMAP enhanced	Brightness temperature	1-3	9
SMAP/Sentinel-1	Soil moisture	12	1/3
Sentinel-1 SAR	Sigma VV/VH, Contrast, Correlation, Variance, Entropy	12	10 m
MOD11A1 MYD11A1	LST	1	1
MOD13A2	NDVI	16	1
MCD43A3	Albedo	1	500 m
SRTM	Elevation, Slope, Aspect	-	90 m
PRISM	Precipitation, API	1	4
ISMN	Soil moisture, Precipitation	1	-

4.3.3 Data pre-processing

All the data used in this study were reprojected to the same coordinate system (i.e., WGS 84) and rigorous data quality control was applied to remove poor-quality pixels which might have been affected by the presence of clouds or aerosols. In addition, to solve the spatial resolution differences among the predictors and response variables, two methods were adopted: simple arithmetic averaging and the nearest-neighbor interpolation. The simple arithmetic averaging was used to aggregate the high-resolution predictors (e.g., LST, Sigma VV/VH) to coarse resolution, whereas the nearest-neighbor interpolation was adopted for resampling of coarse-resolution predictors (e.g., SMAP brightness temperature and precipitation) to fine resolution. The open water bodies in the study area were also masked out using MOD44W, which is a 250 m resolution static global water map provided by MODIS.

4.3.4 Random Forest

In this study, random forest (RF) ML technique was adopted for downscaling of L3SMP (36-km) and L3SMP_E soil moisture (9-km) soil moisture to 1-km. It has increasingly gained popularity within the remote sensing community to solve problems related to classification and regression (Belgiu & Drăgu, 2016). It is a non-parametric tree-based model formed from an ensemble of decision trees (Breiman, 2001). It has great capability to efficiently describe the complex non-linear relationship between soil moisture and a suite of land surface variables as shown in equation (4.1) (Hutengs & Vohland, 2016).

$$SM = f(p_1, p_2, p_3 \dots p_n) + \varepsilon \tag{4.1}$$

where *SM* is soil moisture; ε is the model estimation error, $p_1, p_2, p_3 \dots p_n$ represent predictors (i.e., NDVI, albedo, LST); n represents the total number of predictors.

RF is primarily based on the bootstrap aggregation (bagging), which is a statistical technique used for random resampling of the training dataset with a replacement in order to generate a number of subsets (i.e., bootstrap samples). In this work, about two-thirds of each bootstrap sample (i.e., in-bag (IB)) were used for independent training of each decision tree of the ensemble, whereas the remaining one-third (i.e., out-of-bag (OOB)) was used for verification of the model. In the end, the final predicted value was obtained through averaging of the prediction of each regressor.

RF has a few user-defined parameters to tune in order to increase its predictive power. These are the number of decision trees to grow, maximum number of features selected at each split and the maximum depth of the decision tree. In addition, another attractive feature of RF is its ability to provide the relative importance of predictors by using out-of-bag data. The scikit-learn Python library, which has several built-in functions, was used for implementation of RF algorithm (Pedregosa et al., 2011).

In the present study, the construction of RF model was based on using all available data during the summer seasons of 2017 to 2020 over the Susquehanna watershed, as opposed to construction for each individual day. This was primarily because of the lack of sufficient training data if the model was constructed for each individual day for a small size of study area such as ours. For example, only 130 pixels of SMAP soil moisture covered our study area. Thus, if the model was constructed based on each day, we would only have 130 data points of SMAP soil moisture with their corresponding aggregated predictors for training of RF model at coarse-spatial resolution, which is really very small. Thus, to circumvent the lack of sufficient training data, we used all available data during the study period for training of RF model. This agreed with the previous studies conducted in relatively small-sized study areas in different regions of the world (Abbaszadeh et al., 2020; Chen et al., 2019; Hu et al., 2020).

In this study, the implementation of RF model for downscaling of SMAP soil moisture entailed five general steps: 1) **aggregation:** involved aggregation of high-resolution land surface variables (i.e., predictors) to SMAP soil moisture resolution using simple arithmetic averaging; 2) **splitting of dataset:** involved dividing of aggregated dataset into training and testing sets; 3) **developing the linking model**: this entailed training the RF using SMAP soil moisture as response variable and aggregated land surface variables as predictors (Figure 4.2); 4) **model evaluation**: involved assessment of the model performance built in the third step using out-of-bag and testing datasets; 5) **model application**: this involved applying the developed model to high-resolution land surface variables to predict soil moisture at 1-km.

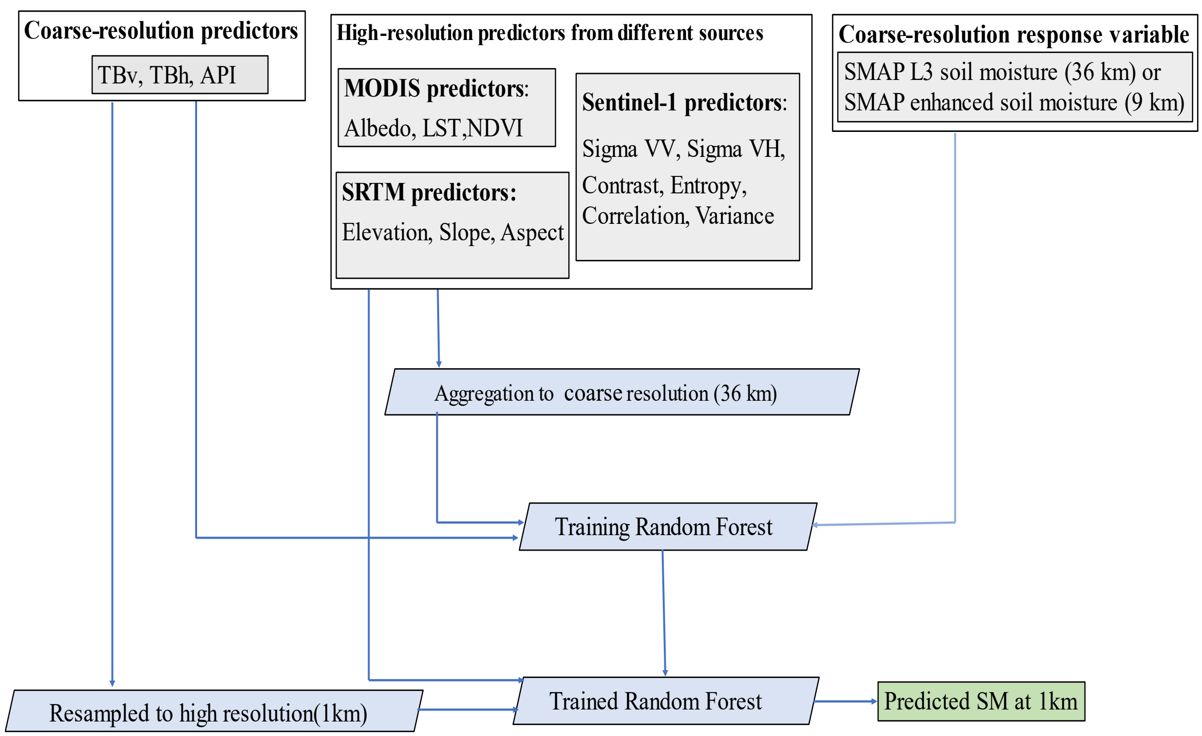

Figure 4.2 Flow chart of downscaling procedure.

It is important to note here that the coarse resolution variables such as brightness temperature and precipitation were resampled to fine resolution during prediction of soil moisture, since they were often not available at high-spatial resolution.

4.3.5 Predictors selection

In the present study, a total of 15 different predictors, including LST, NDVI, surface albedo, elevation, slope, aspect, antecedent precipitation index, brightness temperature, and Sentinel-1 co- and cross-polarized backscatter and textural information (i.e., contrast, entropy, variance and correlation), was used for downscaling of SMAP soil moisture products using RF. The selection of these predictors was based on their strong association with soil moisture variability.

For example, LST is strongly interlinked with soil moisture through thermal inertia and evapotranspiration (ET), among others. Thermal inertia describes the resistance of soil to temperature change and it has positive relation with soil moisture, which implies that wetter soil has higher thermal inertia which in turn leads to decrease in diurnal LST range (Pablos et al., 2018) , whereas, in the context of ET, the relationship between LST and soil moisture varies depending on ET regimes: water- or energy-limited. In water-limited environments, increase in soil moisture leads to decrease in LST, whereas in energy-limited environments, their relationship is not significant (Alemohammad et al., 2018; Sun et al., 2019). Similarly, brightness temperature is among the predictors which influence soil moisture variability. It is dependent on emissivity and surface temperature of soil which, in turn, is determined by dielectric property of the soil. Interestingly, dielectric properties of the soil are primarily dependent on soil moisture content, among many other factors (Ye et al., 2015) . On the other hand, precipitation is the main source of soil moisture and controls its spatial and temporal variations.

Apart from above-mentioned predictors, Sentinel-1 co- and cross-polarized backscatter coefficients, which represent normalized radar return from targeted surface, have a direct positive relationship with soil moisture, most notably over sparse to moderately vegetated areas (Hajj et al., 2017). It is worth to note that this relationship could be affected by perturbing factors such as vegetation and surface roughness. Similarly, topographic characteristics such as elevation, slope and aspect control the spatial distribution of soil moisture. For example, flat terrain tends to be wetter than sloping hillside and vice versa (Meerveld & McDonnell, 2006).

Moreover, surface albedo is among the important variables that affect soil moisture variability. It represents the proportion of incoming solar radiation reflected by the earth surface and plays a crucial role in the energy balance of land surface, thereby affecting near-surface climate. It can be affected by many factors, including soil moisture, soil texture, elevation angle and vegetation cover. For example, in the context of soil moisture, wet soils tend to have a darker color than dry soils, which in turn reduces surface albedo, implying negative relationship between the two variables (Guan et al., 2009).

NDVI is also one of the predictors strongly related to soil moisture. It reflects the greenness status of vegetation, which in turn depends on the availability of soil moisture. This indicates a positive correlation between NDVI and soil moisture. Similar to NDVI, Sentinel-1 image textural features are used to describe the spatial pattern of pixel values of an images and are often used for land cover classification. Here, as predictors, they were used to provide information about spatial variation of vegetation, which is heavily dependent on soil moisture (Haralick et al., 1973).

Overall, in this study, predictors such as LST, antecedent precipitation index and brightness temperature were used with the aim to preserve temporal dynamics of downscaled SMAP soil moisture, whereas static topographical information comprising elevation, slope and aspect were included to maintain the spatial patterns of downscaled SMAP soil moisture. In addition, vegetation information, e.g., NDVI and Sentinel-1 textural information were used to be able to elucidate the impact of vegetation on the spatio-temporal patterns of downscaled soil moisture. Sentinel-1 backscatter in VV polarization was included as additional predictor because of its high sensitivity to soil moisture change (notably in sparse to moderate vegetation cover), while VH polarization was considered to account for the effect of vegetation dynamics on soil moisture.

In this study, three experiments were implemented. The first two experiments (i.e., Exp1 and 2) were implemented to evaluate the contribution of predictors derived from Sentinel-1. In Exp1, Sentinel-1 predictors, along with other predictors, were used in the downscaling of SMAP soil moisture products (i.e., L3SMP and L3SMP_E), but in Exp2 they were removed and only the remaining predictors were used in the downscaling (see Table 4.2). For comparison purposes, both Exp1 and 2 had the same data length, which was determined by the availability of Sentinel-1 images (i.e., available every 12 days over the study area). On the other hand, Exp3 was proposed, in which the Sentinel-1 predictors were not considered. The exclusion of these predictors increased the training data sample size due to the frequent availability of all the predictors (Table 4.1) and SMAP soil moisture as a response variable (i.e., available every 2 to 3 days). This implied that Exp3 could produces downscaled soil moisture every 2 to 3 days as opposed to Exp1, which produced downscaled soil moisture only every 12 days because of the temporal constraint imposed by the Sentinel-1 revisit time. Overall, it is worth to mention that the downscaling was implemented for spatio-temporally collocated predictors and response variables.

Table 4.2 Combination of different predictors for different experiments (Sentinel-1 predictors)*

Experiments	Predictors	Response variable	Remarks
Exp1	LST, NDVI, Albedo, TB_V, TBh, API, elevation, aspect, slope Sigma VV/VH*, Contrast*, Correlation*, Variance*, Entropy*	L3SMP L3SMP_E	Length of data was determined by availability of Sentinel-1 images
Exp2	LST, NDVI, Albedo, TB_V, TBh, API, elevation, aspect, slope		Sentinel-1 variables were removed from Exp2
Exp3	LST, NDVI, Albedo, TB_V, TB_h, API, elevation, Aspect, Slope	L3SMP L3SMP_E	Relatively large sample size data for training of RF models

4.4 Results

4.4.1 Random Forest model performance

In this study, two SMAP soil moisture products, namely, the SMAP level-3 (L3SMP, 36-km) and SMAP enhanced (L3SMP_E, 9-km), were used as background fields to downscale to 1- km spatial resolution using RF models. A total of 84 images of SMAP soil moisture which were spatio-temporally collocated with another 84 images of each predictor available for summer seasons of 2017 to 2020 was used in the RF model development (Exp1,Table 4.2). The number of available images was limited due to the low temporal availability of Sentinel-1 images. About two-thirds of these images were used to train RF models, whereas the remaining third was used for testing. Before applying RF models for prediction of soil moisture at 1-km, their performances were evaluated at coarse-spatial resolution (i.e., at 36- and 9-km) by constructing the models using the predictors aggregated at these resolutions. Figure 4.3 a and b presents the scatter plots of the predicted soil moisture as a function of observed SMAP soil moisture at 36-

and 9-km, respectively, using test data of Exp1. In addition, their respective performance metrics are also displayed in the figure.

As can be seen in Figure 4.3 a and b, the predicted soil moisture generally varied linearly with the observed SMAP soil moisture; however, with a slight tendency to overestimate and underestimate low (0,2) and high (0,4) soil moisture values, respectively, for both L3SMP and L3SMP_E soil moisture. In addition, the statistical performance metrics presented in the figure indicated a good agreement between the predicted and observed SMAP soil moisture at both spatial scales. The RF model developed at the 9-km spatial scale performed better with R, ubRMSE and bias values of 0,91, 0,03 m^3/m^3 and -0,001 m^3/m^3, respectively, compared to the RF model built at 36-km with R, ubRMSE and bias values of 0,84, 0,04 m^3/m^3 and -0,001 m^3/m^3, respectively.

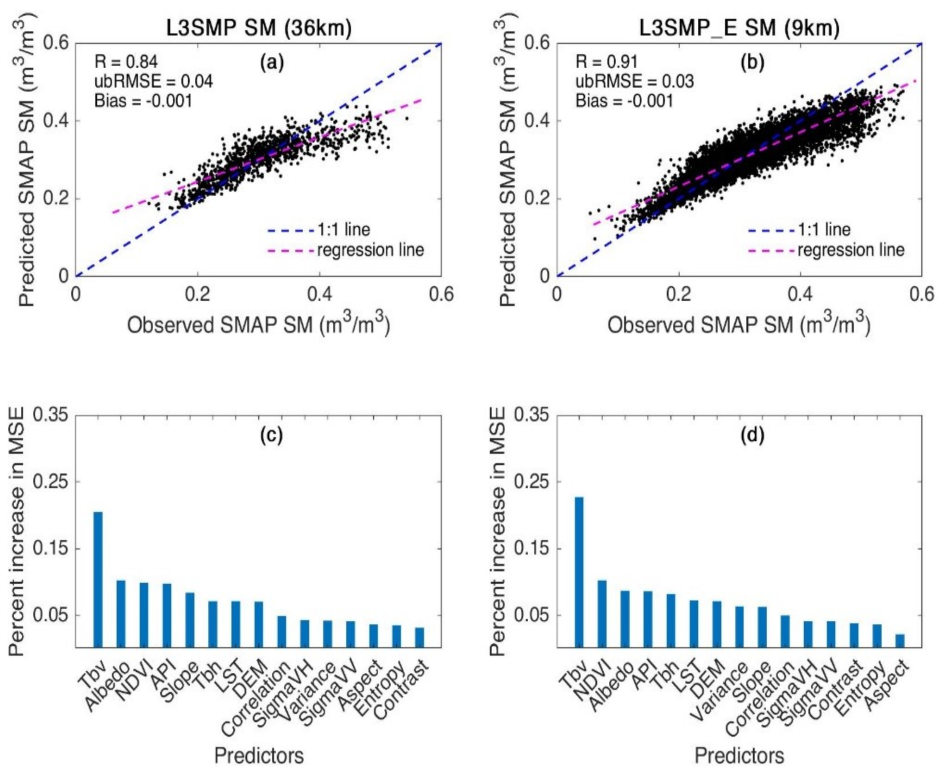

Figure 4.3 RF models verification at (a) 36-km, (b) 9-km and their predictors importance scores at (c) 36-km and (d) 9-km.

The evaluation of RF models at both spatial scales generally suggests their capability to predict soil moisture at a coarse-spatial resolution with acceptable accuracy. Therefore, these models were implemented for the prediction of soil moisture at 1-km spatial resolution, assuming that the RF model built at coarse-spatial resolution was also valid for the prediction of soil moisture at high-spatial resolution using high-spatial resolution predictors. In other words, it assumed that the developed RF models were spatial scale invariant.

4.4.2 Relative importance of predictors

Whenever RF models are employed in the downscaling of geophysical variables, including soil moisture, it is of interest to identify the relative importance of the predictors used. Relative importance is often computed by determining the percentage of the increase in the mean square error (MSE) when a given predictor is randomly permuted, while keeping the remaining predictors unaltered. Generally, higher values indicate the higher importance of a given predictor in enhancing the predictive power of the RF model. Figure 4.3c and d shows the relative importance of each predictor for both RF models. According to the figure, among the input predictors, the TBv, Albedo, NDVI, API, Tbh and slope were identified as the most important predictors for both models, while the predictors derived from Sentinel-1 had less importance as shown in Figure 4.3 c.

4.4.3 Validation of downscaled soil moisture with in-situ observations

Figure 4.4 shows the time series comparisons of the downscaled L3SMP soil moisture, the most recent 1-km SMAP/Sentinel-1 existing soil moisture product (L2_SM_SP) and the original L3SMP soil moisture, along with in-situ soil moisture observations and precipitation collected from four selected sites (i.e., Avondale, Ithaca, Geneva and Rockspring) for the summer seasons (May to October) of 2017 to 2020. The downscaled L3SMP soil moisture matched well with the temporal dynamics of the in-situ soil moisture across the four sites. Likewise, L2_SM_SP and the original L3SMP soil moisture also captured well the temporal variability of in-situ soil moisture. Compared to L2_SM_SP and L3SMP soil moisture, the downscaled L3SMP soil moisture more closely corresponded with in-situ soil moisture observations, yet a slight systematic underestimation or overestimation of extreme soil moisture values of all the products was observed; see, for example, the Geneva site. Moreover, all the three products, including the downscaled L3SMP, L2_SM_SP and the original L3SMP soil moisture, responded quite well to the temporal change of rainfall events as can be seen in Figure 4.4, but in-situ soil moisture better correlated with the rainfall events, since both were point measurements located at the same site as opposed to all other products, which give an average soil moisture over a large area (i.e., 1x1 km and 36 x36 km grid cell).

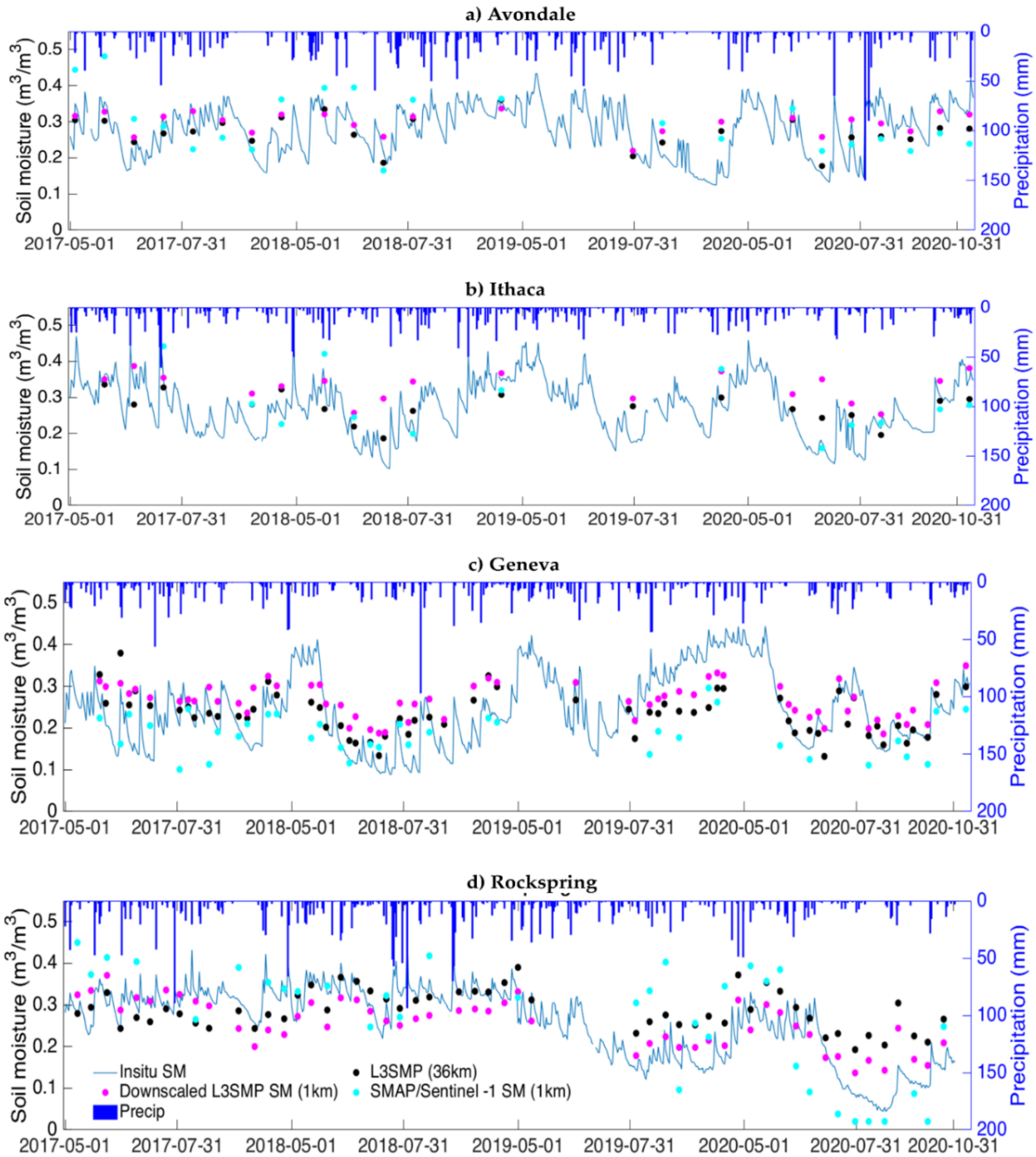

Figure 4.4 Time series of downscaled L3SMP soil moisture, original SMAP (L3SMP), L2_SM_SP, daily precipitation and in-situ soil moisture for (a) Avondale, (b) Ithaca, (c) Geneva and (d) Rockspring sites

In a similar way, time series plots of the downscaled L3SMP_E soil moisture, the original L3SMP_E soil moisture and L2_SM_SP along with in-situ soil moisture observations and precipitation for the four in-situ soil moisture stations in the Susquehanna watershed are shown in Appendix Figure A.0.1 . It can be observed that both the downscaled L3SMP_E and its original counterpart showed a good agreement with in-situ soil moisture observations. Comparing Appendix Figure A.0.1 to Figure 4.4, it was found that both the downscaled L3SMP and L3SMP_E soil moisture had similar temporal behaviors in capturing the wetting and drying patterns of in-situ soil moisture in response to rainfall events.

4.4.4 Quantitative assessment of downscaled soil moisture

Figure 4.5 presents the scatter plot of downscaled L3SMP, the original L3SMP and L2_SM_SP soil moisture against in-situ soil moisture observations at the Avondale, Ithaca, Geneva and Rockspring sites, whereas Table 4.3 shows their respective performance metrics.

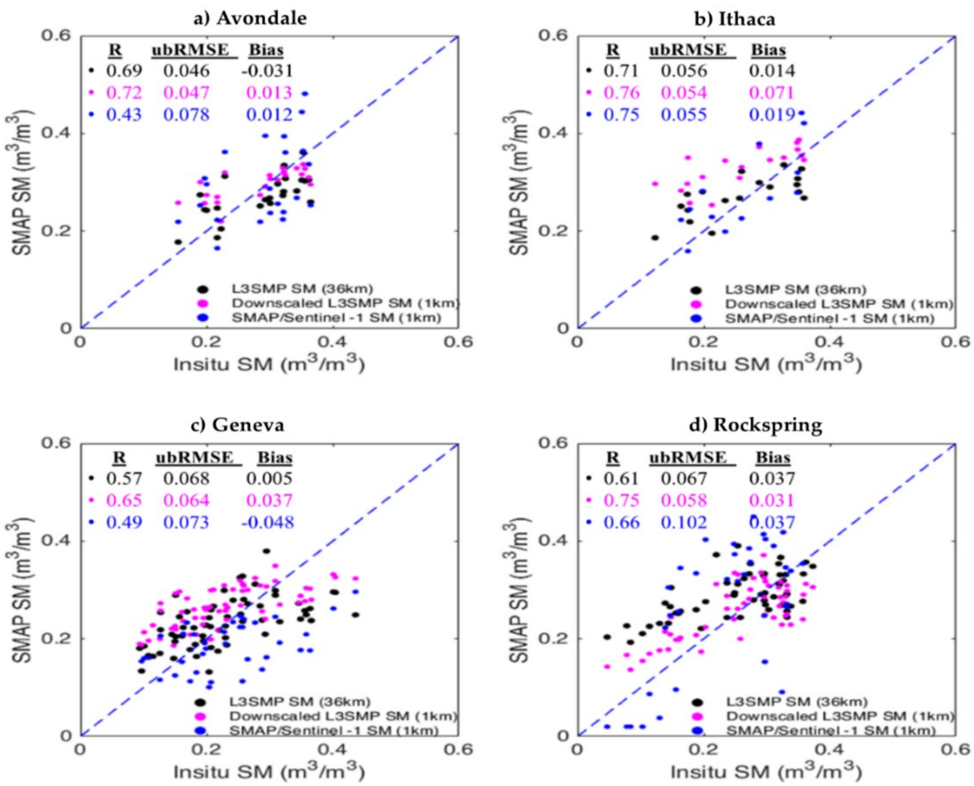

Figure 4.5 Scatter plot of the downscaled L3SMP, the original L3SMP and L2_SM_SP SM versus in-situ SM observation for (a) Avondale, (b) Ithaca, (c) Geneva and (d) Rockspring sites.

The downscaled L3SMP soil moisture had a slightly higher correlation when compared to in-situ soil moisture observations than the original L3SMP soil moisture and L2_SM_SP did, with R values in the range of 0,65 (i.e., at the Geneva site) to 0,76 (i.e., at the Ithaca site).

Similarly, the ubRMSE values of the downscaled L3SMP soil moisture were somewhat lower than that of L3SMP and L2_SM_SP soil moisture for all the four sites, ranging from 0,047 to 0,064 m^3/m^3. In terms of bias, on the other hand, the downscaled L3SMP soil moisture had a slightly higher bias particularly for the Ithaca and Geneva stations. Moreover, L2_SM_SP had less overall accuracy compared to both downscaled L3SMP and its corresponding original product (Table 4.3).

Table 4.3 Performance metrics of downscaled, the original and SMAP/Sentinel-1 products

Stations	Metrics	SMAP L3 (L3SMP)		SMAP/Sentinel-1	SMAP enhanced (L3SMP_E)		SMAP/Sentinel-1
		L3SMP (36km)	Down_ L3SMP (1-km)	L2_SM_SP (1-km)	L3SMP_E (9-km)	Down_ L3SMP_E (1-km)	L2_SM_SP (1-km)
Avondale	R	0,690	0,720	0,430	0,670	0,730	0,430
	ubRMSE	0,046	0,047	0,078	0,047	0,044	0,078
	Bias	-0,031	0,013	0,012	-0,031	-0,006	0,012
Ithaca	R	0,710	0,760	0,750	0,690	0,800	0,750
	ubRMSE	0,056	0,054	0,055	0,056	0,050	0,055
	Bias	0,014	0,071	0,019	0,038	0,048	0,019
Geneva	R	0,570	0,650	0,490	0,590	0,640	0,490
	ubRMSE	0,068	0,064	0,073	0,063	0,059	0,073
	Bias	0,005	0,037	-0,048	0,0004	0,021	-0,048
Rocksprings	R	0,610	0,750	0,660	0,630	0,730	0,660
	ubRMSE	0,067	0,058	0,102	0,066	0,058	0,102
	Bias	0,037	0,031	0,037	0,0123	-0,017	0,037

Similarly, the scatter plots of the downscaled L3SMP_E, the original L3SMP_E soil moisture and L2_SM_SP versus in-situ soil moisture observations are presented in Appendix Figure A.0.2. The downscaled L3SMP_E soil moisture had a slightly better accuracy with R, ubRMSE and bias values in ranges of 0,64 to 0,80, 0,044 to 0,059 m^3/m^3 and -0,006 to 0,048 m^3/m^3, respectively, compared to its original version with R, ubRMSE and bias values in ranges of 0,59 to 0,69, 0,047 to 0,066 m^3/m^3 and -0,031 to 0,038 m^3/m^3, respectively. On the other hand, L2_SM_SP soil moisture had less overall accuracy compared to both downscaled L3SMP_E soil moisture and its corresponding original product (see Table 4.3). As can be observed in Table 4.3 , both downscaled L3SMP and L3SMP_E had comparable performance metrics with only a slight outperformance for the downscaled L3SMP_E.

4.4.5 Evaluation of spatial pattern

A successfully downscaled soil moisture should reproduce the spatial pattern of its corresponding coarse-resolution product, and its evaluation generally relies on a visual comparison of the downscaled soil moisture with its respective native resolution. Figure 4.6 a and b display the spatial distribution of L3SMP and L3SMP_E soil moisture, whereas Figure 4.6 c and d shows their corresponding downscaled soil moisture at a 1-km spatial resolution using RF on 17th June 2018. From a visual comparison, it was evident that the downscaled L3SMP and L3SMP_E soil moisture by the RF model reproduced the spatial patterns of their corresponding native resolutions reasonably well.

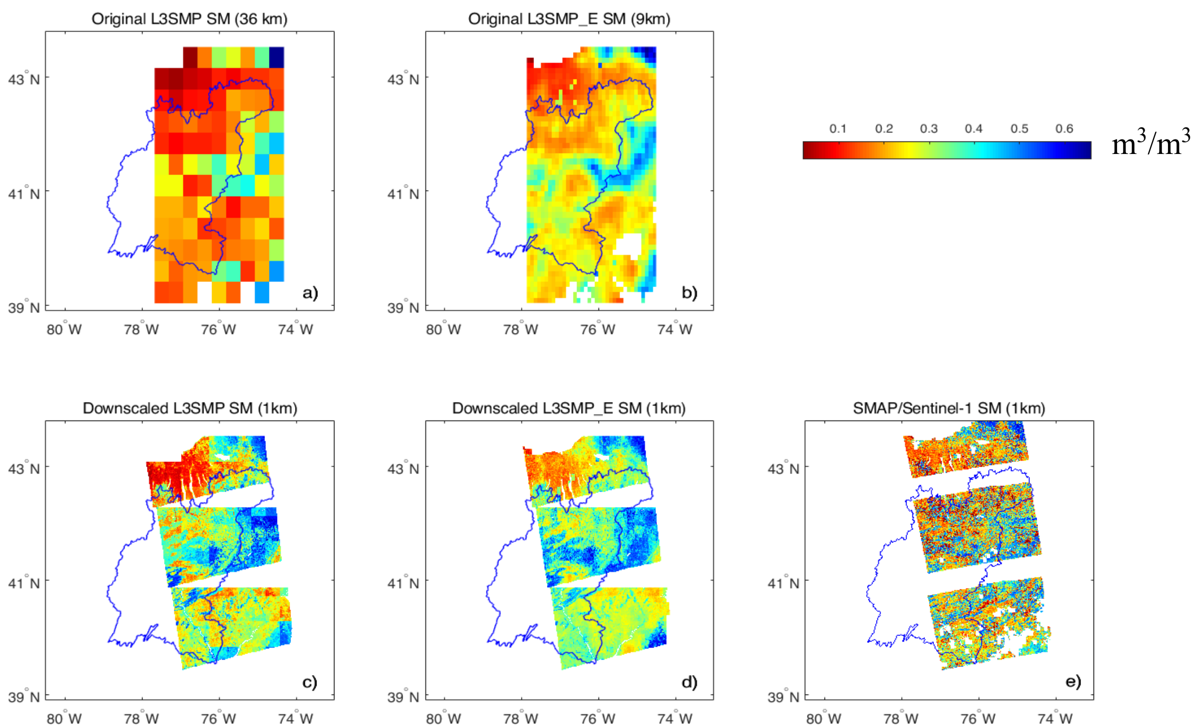

Figure 4.6 Maps of (a) the original L3SMP SM, (b) the original L3SMP_E SM, (c) downscaled L3SMP SM, (d) the downscaled L3SMP_E SM and (e) L2_SM_SP on 17th June 2018.

In addition, they provided spatially more detailed information of soil moisture. The wet and dry regions in the native L3SMP or L3SMP_E soil moisture maps were clearly portrayed in their corresponding downscaled products and L2_SM_SP. As can be seen in the figure, the northern part was characterized by a drier soil moisture, whereas the central–eastern part displayed wetter soil moisture condition. The southern part represented a medium soil moisture range. The

L2_SM_SP also somewhat reproduced the spatial pattern of both L3SMP and L3SMP_E soil moisture, as can be observed in Figure 4.6e, but it was relatively less visually appealing compared to the downscaled L3SMP and L3SMP_E soil moisture.

4.4.6 Evaluation of the value of Sentinel-1 predictors in downscaling

The results presented in previous sections were based on the downscaling of L3SMP and L3SMP_E soil moisture using the Sentinel-1 predictors together with other predictors (i.e., Exp 1, Table 4.2). Based on a visual inspection and the statistical metrics, the downscaled L3SMP and L3SMP_E soil moisture reproduced the temporal dynamics of in-situ observation and the spatial pattern of their original counterparts reasonably well, as indicated in Figure 4.3 to Figure 4.6.

However, in this section, the contribution of Sentinel-1 predictors in improving the predictive performance of RF models for each spatial scale (i.e., 36- and 9-km) was evaluated. Thus, four RF models were trained and tested separately. The first (RF_1) and second (RF_2) RF models were implemented for the downscaling of L3SMP and L3SMP_E soil moisture, respectively, using Sentinel-1 predictors along with other predictors (i.e., Exp 1, Table 4.2), while the third (RF_3) and fourth (RF_4) RF models were implemented using the same predictors as that of RF_1 and RF_2, but without Sentinel-1 predictors for the downscaling of L3SMP and L3SMP_E soil moisture, respectively (i.e., Exp 2, Table 4.2). It is also worth to mention that only both spatially and temporally overlapped Sentinel-1 and other predictors were considered in all four models for the purpose of the comparison. The result of this comparison is presented in Figure 4.7.

Figure 4.7 shows the scatterplots of predicted and observed SMAP soil moisture for both Exp1 and 2 at both spatial scales for the four RF models mentioned above. The visual inspection clearly showed a good performance of the RF models, but with a slight tendency of overestimation and underestimation of minimum and maximum soil moisture values, respectively, at both spatial scales. Likewise, the models attained higher statistical metrics for both experiments (see Table 4.4). Accordingly, at the spatial scale of 36-km, the R values of RF_1 and RF_3 were 0,84 and 0,83, respectively, while the ubRMSE and bias values were the same for both RF_1 and RF_3, which were 0,04 m^3/m^3 and -0,001 m^3/m^3, respectively.

Similarly, at the spatial scale of 9-km, the R values of RF_2 and RF_4 were 0,91 and 0,86, respectively, while the same values of bias were obtained for both, which was -0,001 m^3/m^3. Meanwhile, at the spatial scale of 9-km, the values of ubRMSE for RF_2 and RF_4 were 0,03 m^3/m^3 and 0,04 m^3/m^3, respectively.

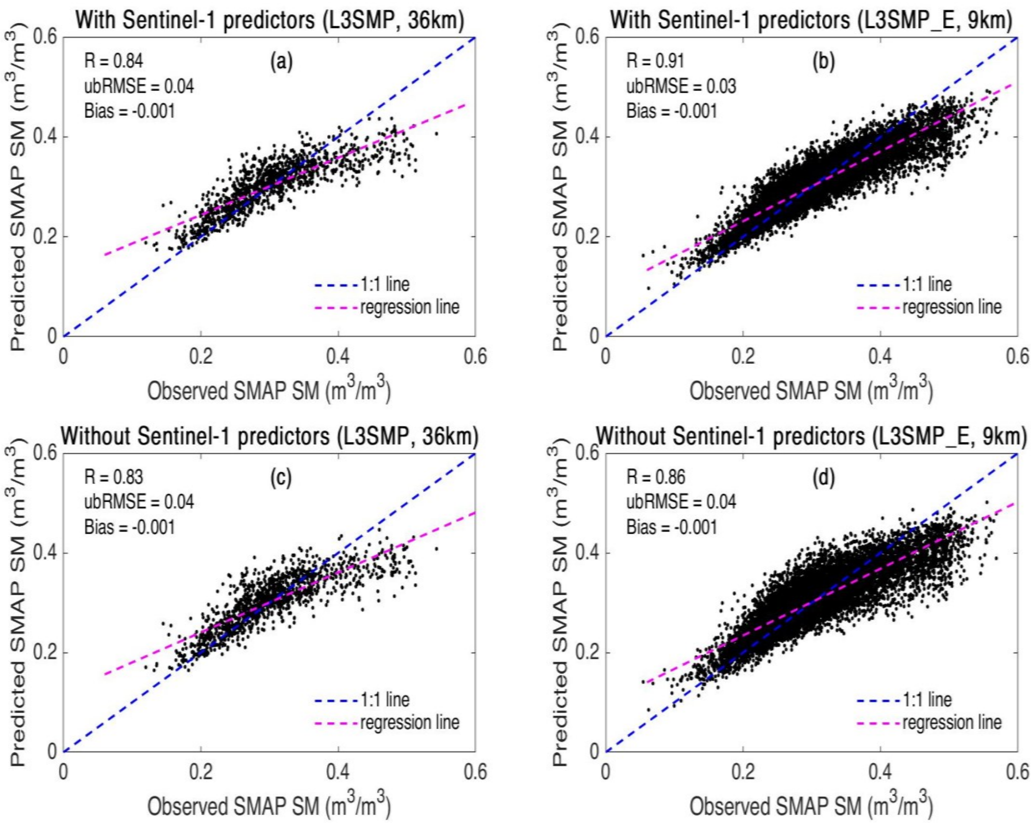

Figure 4.7 Comparison of RF model performances (a) with Sentinel-1 predictors at 36-km, (b) with Sentinel-1 predictors at 9-km, (c) without Sentinel-1 predictors at 36-km and (d) without Sentinel-1 predictors at 9-km.

The comparison results of Exp1 and 2 showed that leaving out Sentinel-1 predictors slightly decreased the performance of RF models at both spatial scales (Figure 4.7 and Table 4.4). For instance, at the spatial scale of 36-km, the R value declined only by 0,01, while at the spatial scale of 9-km, it reduced by 0,05. Apart from R values, the ubRMSE and bias values remained the same for both experiments at 36-km. On the other hand, at the spatial scale of 9-km, the ubRMSE reduced by 0,01, while the bias remained the same. Overall, the models with Sentinel-1 (RF_1 and RF_2) slightly outperformed those without Sentinel-1 predictors (RF_3 and RF_4) at both spatial scales, but their difference in the overall accuracy was minimal. This indicated that the contribution of the Sentinel-1 predictors was marginal in improving the predictive accuracy RF models for the underlying surface and climatic conditions of our study area.

Table 4.4 Performance metrics of RF at spatial scale of 36- and 9-km with and without Sentinel-1 predictors.

	SMAP L3 (L3SMP, 36-km)		SMAP enhanced (L3SMP_E, 9-km)	
Metrics	With Sentinel-1 predictors	Without Sentinel-1 predictors	With Sentinel-1 predictors	Without Sentinel-1 predictors
R	0,84	0,83	0,91	0,86
ubRMSE	0,04	0,04	0,03	0,04
Bias	-0,001	-0,001	-0,001	-0,001

4.4.7 Downscaling of SMAP soil moisture products without Sentinel-1 predictors

By taking into account the lesser importance of the Sentinel-1 predictors for the condition of our study area as shown in Section 4.4.2, in the remaining section of the results, we primarily focused on the downscaling of L3SMP and L3SMP_E soil moisture using predictors of Exp3 (Table 4.2), namely, the TBv, TBh, Albedo, NDVI, API, slope, LST, DEM and aspect. It is important to note that, although Exp2 and 3 had the same predictors, the length of their data record was different. In the case of Exp3, the availability of data depended on the revisit time of the SMAP soil moisture, which was every 2 to 3 days, whereas for Exp2, the length of the data record was limited by the availability of Sentinel-1 images (i.e., every 12 days) for the purpose of the comparison with Exp1. In other words, the RF models developed for Exp3 had substantially more data for training compared to Exp2, and, therefore, should have resulted in more robust models. In addition to having more data for training the RF models, other advantages of Exp3 relative to Exp2 were the increased temporal availability of the downscaled soil moisture and a better spatial completeness. Therefore, based on this premise, new RF models were trained and tested based on Exp3 at both spatial scales.

Figure 4.8a and b display the scatter plots of the observed versus predicted SMAP soil moisture for two separately trained RF model at 36- and 9-km spatial scales, respectively, for the case of Exp3. A total of 325 images of SMAP soil moisture, which were spatio-temporally collocated with other 325 images of each predictor listed in Table 4.2, for Exp3, was used. About two-thirds of these images were used to train the RF models, whereas the remaining third was used for testing. The RF model at the 36-km spatial scale showed good performance with R, ubRMSE and bias values of 0,91, 0,03 m^3/m^3 and -0,0002 m^3/m^3, respectively (Table 4.5). Similarly, the RF model at the spatial scale of 9-km had R, ubRMSE and bias values of 0,87, 0,04 m^3/m^3 and 0,0001 m^3/m^3, respectively (Table 4.5).

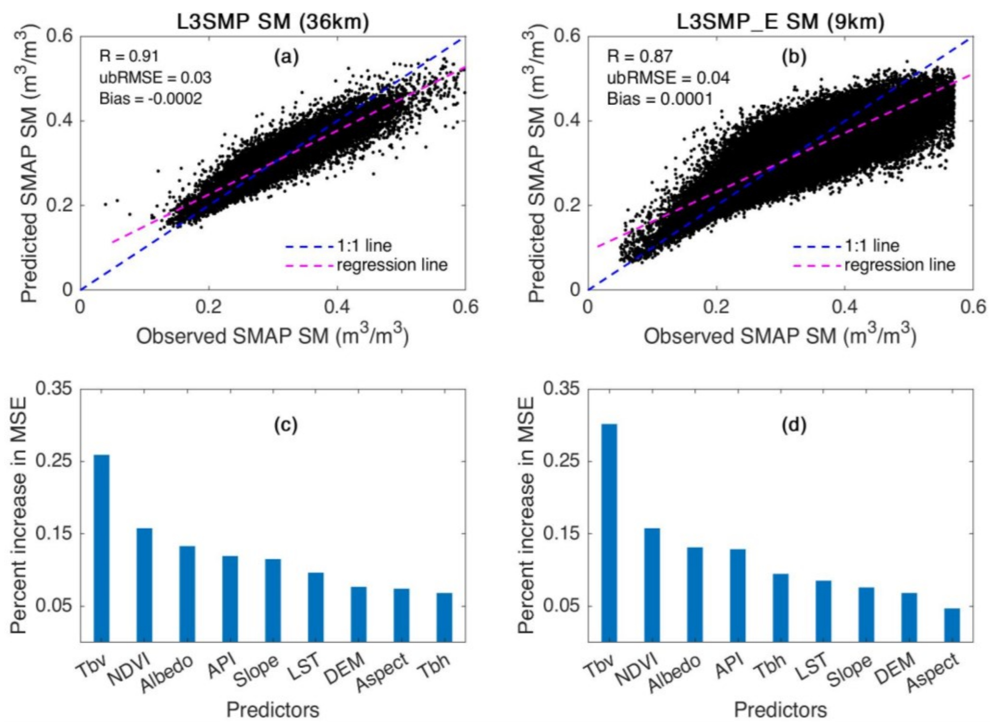

Figure 4.8 RF models verification at (a) 36-km, (b) 9-km and their predictors importance scores at (c) 36-km and (d) 9-km.

Table 4.5 Performance of RF models for Exp3 at spatial scale 36- and 9-km.

Metrics	L3SMP (36-km)	L3SMP_E (9-km)
R	0,91	0,87
ubRMSE	0,03	0,04
Bias	-0,0002	0,0001

In addition, Figure 4.8 c and d presents the relative importance score of the predictors. Among the predictors, the TBv, NDVI, Albedo slope and API had the most influential roles at both spatial scales, with a high percentage of increase in MSE during their permutation. Comparing Table 4.5 to Table 4.4, the performance of RF models of Exp3 increased compared to Exp2. In addition, the bias in extreme soil moisture values was less pronounced in Exp3 compared to Exp1 and 2 (comparing Figure 4.8 and Figure 4.7).

Validation of downscaled SM with in-situ observations

To evaluate the temporal behavior of downscaled L3SMP and the original L3SMP soil moisture, a time series comparison was performed with in-situ soil moisture and precipitation collected from four sites as shown in Figure 4.9.

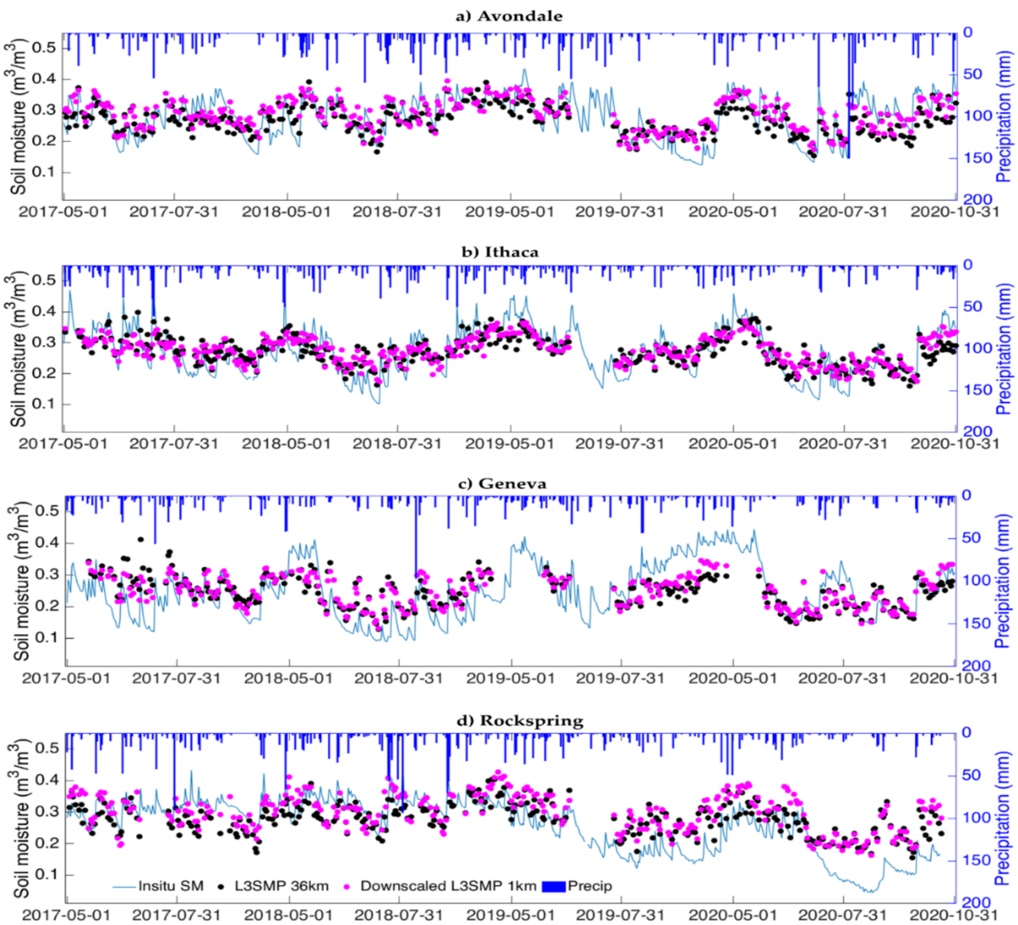

Figure 4.9 Time series of downscaled SM, the original SMAP (L3SMP), daily rainfall and in-situ SM for (a) Avondale, (b) Ithaca, (c) Geneva and (d) Rockspring sites

The downscaled soil moisture showed a good temporal consistency with in-situ soil moisture observations across all sites. It also preserved the temporal dynamics of its corresponding original product (i.e., L3SMP). In addition, the downscaled L3SMP soil moisture and its original counterpart exhibited similar wetting and drying patterns of in-situ soil moisture in response to the rainfall events. However, the lower soil moisture values were not sufficiently captured at all sites, and, in a similar way, some underestimations of higher soil moisture values were observed.

Similarly, a time series comparison of the downscaled L3SMP_E and the original L3SMP_E against in-situ soil moisture observations along with the precipitation was depicted in Appendix Figure A.0.3. The downscaled L3SMP_E soil moisture showed a quite good agreement with the in-situ soil moisture observations across all sites. It also responded well to the temporal change of rainfall events. Comparing Appendix Figure A.0.3 and Figure 4.9, it could be clearly seen that both the downscaled L3SMP and L3SMP_E and their corresponding original products generally followed a similar temporal pattern, and also reacted similarly to the rainfall events.

Qualitative evaluation of downscaled soil moisture

Figure 4.10 depicts the scatter plots of the downscaled L3SMP and their original counterpart against in-situ soil moisture observations collected from the four sites in the study area. The statistical performance metrics computed for each site are also displayed in Table 4.6. The downscaled L3SMP soil moisture agreed quite well with the in-situ measurements, with R values greater than 0,68 for all the sites. However, in terms of the ubRMSE, the Avondale and Ithaca sites had values which were quite closer to the SMAP mission accuracy requirement, which was 0,04 m^3/m^3, as can be observed in Table 4.6. In comparison to the original L3SMP soil moisture, the downscaled L3SMP soil moisture had a slightly higher R and lower ubRMSE values across all the sites, while its bias was relatively higher. Moreover, from a visual inspection of the scatter plots, the model slightly overestimated lower soil moisture values, while the higher soil moisture values were somewhat underestimated.

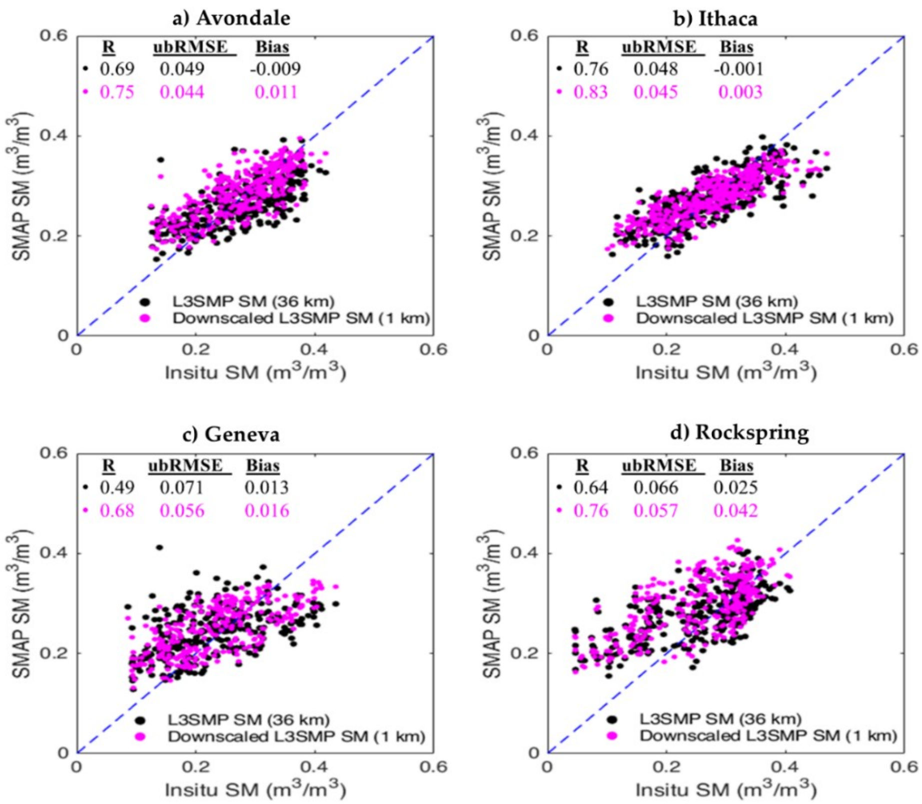

Figure 4.10 Scatter plot of L3SMP and downscaled SM versus in-situ SM observation for (a) Avondale, (b) Ithaca, (c) Geneva and (d) Rockspring sites.

Analogous to Figure 4.10, the scatter plots of the downscaled L3SMP_E and the original L3SMP_E soil moisture against the in-situ soil moisture observations are presented in Appendix Figure A.0.4. The downscaled L3SMP_E soil moisture showed a better performance with R, ubRMSE and bias values in the ranges of 0,74 to 0,85, 0,04 to 0,056 m^3/m^3 and -0,0002 to 0,019 m^3/m^3, respectively, compared to its original counterpart (i.e., L3SMP_E), with R, ubRMSE and bias values in the ranges of 0,58 to 0,83, 0.066 to 0,043 m^3/m^3 and -0,0007 to 0,0392 m^3/m^3, respectively (Table 4.6). Among all sites, the downscaled L3SMP_E and original L3SMP_E soil moisture performed better at the Avondale and Ithaca sites. The comparison of Figure 4.10 with Appendix Figure A.0.4 indicated that the overall performance metrics of the downscaled L3SMP_E soil moisture were modestly higher than that of the downscaled L3SMP soil moisture across all the four sites, although it was not significant (see Table 4.6). In addition, a similar tendency of a slight overestimation and underestimation of lower and higher soil moisture values, respectively, was noticed at both spatial scales.

Table 4.6 Performance metrics of original SMAP and downscaled soil moisture.

Stations	Metrics	SMAP L3 (L3SMP, 36-km)		SMAP Enhanced (L3SMP_E, 9-km)	
		L3SMP	Down_L3SMP	L3SMP_E	Down_L3SMP_E
Avondale	R	0,69	0,75	0,71	0,81
	ubRMSE	0,049	0,044	0,048	0,043
	Bias	-0,009	0,011	-0,02	-0,001
Ithaca	R	0,76	0,83	0.71	0,82
	ubRMSE	0,048	0,045	0,054	0,044
	Bias	-0,001	0,003	0,005	-0,0045
Geneva	R	0,49	0,68	0,51	0,70
	ubRMSE	0,071	0,056	0,068	0,053
	Bias	0,013	0,016	0,006	0,017
Rocksprings	R	0,64	0,76	0,65	0,71
	ubRMSE	0,066	0,057	0,066	0,06
	Bias	0.025	0.042	0.018	0.043

<u>Visual evaluations of spatial pattern of downscaled soil moisture</u>

The visual comparison between the downscaled soil moisture and its corresponding original product allows one to comprehend whether the proposed model is reproducing the spatial pattern of the original product well. Figure 4.11 a and b shows the maps of spatial variations of the downscaled L3SMP and L3SMP_E soil moisture, while their corresponding original equivalents are presented in Figure 4.11c and d, respectively. It can be inferred from the figure that the proposed downscaling method considerably improved the spatial details of the original soil moisture, while maintaining their spatial pattern across the entire basin. It can be observed that a lower soil moisture appeared in the northwest, whereas a higher soil moisture occurred in the middle-eastern part of the basin. The southern part of the basin was mainly characterized by medium range soil moisture values.

When comparing Figure 4.11 c and d, similar spatial patterns of the downscaled soil moisture maps were generated from the two background soil moisture fields (i.e., L3SMP and L3SMP_E). Both of them were characterized by higher and lower soil moisture values in the southern and northwestern parts of the basin, respectively. However, a closer look at the maps of their spatial distribution revealed that the downscaled L3SMP_E more closely matched its corresponding original version than the downscaled L3SMP soil moisture did with its original counterpart. Finally, compared to the maps of soil moisture produced from exp1 (Figure 4.6) on June 17th of 2018, Exp3 provided a map of soil moisture which a is spatially complete with less overestimations and underestimations of the lower and higher soil moisture values, respectively (Figure 4.11).

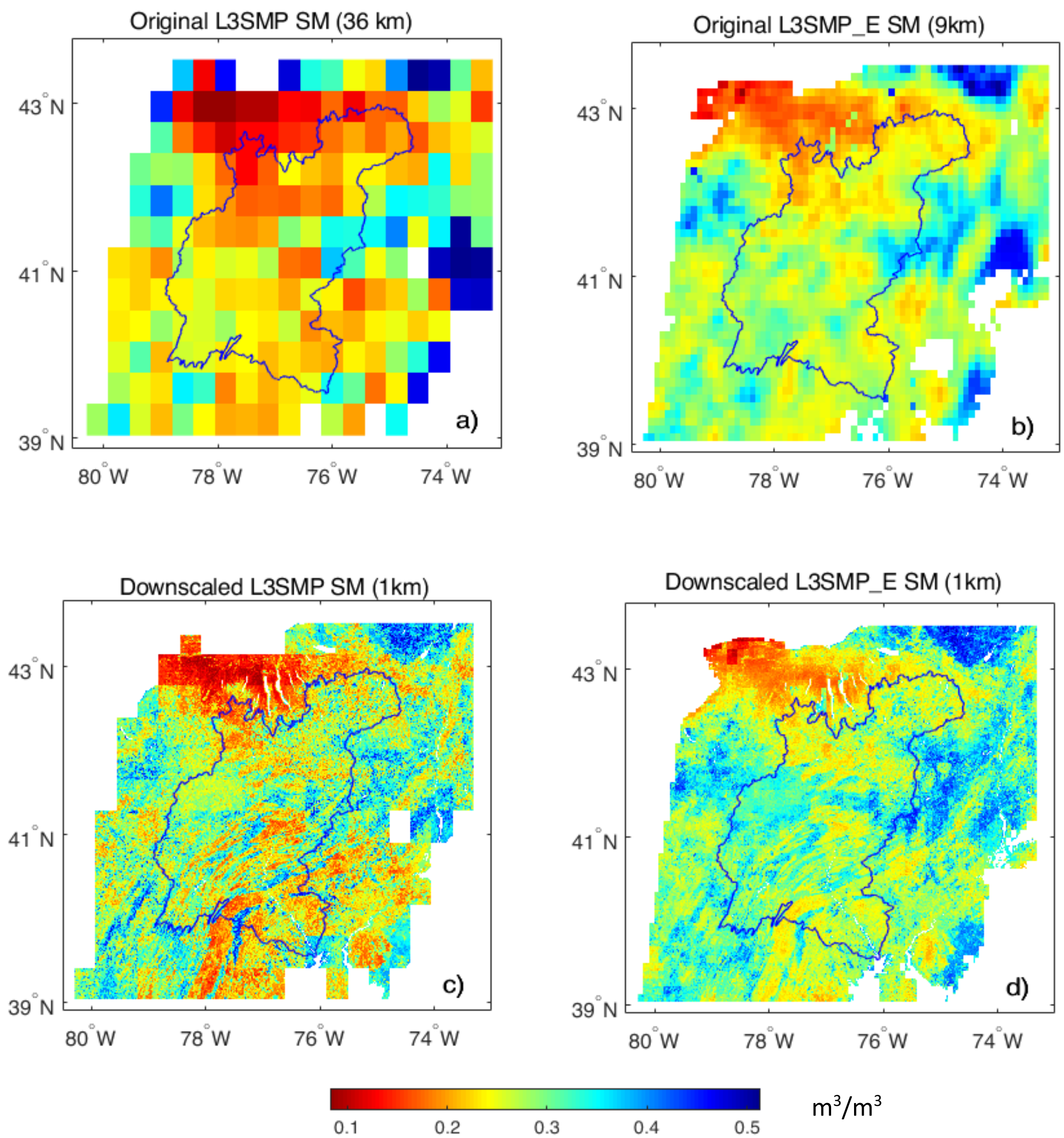

Figure 4.11**.** Maps of (a) the L3SMP, (b) L3SMP_E, (c) downscaled L3SMP and (d) downscaled L3SMP_E on 17th June 2018.

4.5 Discussion

This study demonstrated the downscaling of two SMAP soil moisture products, which were L3SMP (36-km) and L3SMP_E (9-km), to 1-km spatial resolution using an RF model by considering three different experiments. In the first experiment, the RF models were separately implemented at a resolution of L3SMP (i.e.,36-km) and L3SMP_E (i.e., 9-km) using predictors derived from Sentinel-1 and MODIS, together with others, such as the brightness temperature, elevation, slope, aspect and antecedent precipitation index (Exp1, Table 4.2), while in the second experiment, the RF models were employed at both spatial scales by leaving out the Sentinel-1 predictors and focusing only on MODIS products, along with other predictors mentioned above (Exp2, Table 4.2). It is also important to note here that, for the purpose of the comparison of the two experiments (i.e., Exp1 and 2), only dates when the Sentinel-1 images were available were considered. Finally, the third experiment (Exp3, Table 4.2) was proposed based on the limitations of the comparison results between Exp1 and 2.

4.5.1 RF model performance evaluation

An adequate performance of the RF model at a coarse-spatial resolution was required before its use for predicting soil moisture at a high-spatial resolution. Thus, in this study, the evaluation of RF models was carried out at two spatial scales: 1) 36-km resolution for the SMAP L3SMP product and 2) 9-km resolution for SMAP L3SMP_E. In both contexts, the results showed the RF models to have a good capability in predicting soil moisture at both spatial scales for both Exp1 and 2. In addition, the comparative evaluation of the performance measures between the RF models of Exp1 and 2 at both spatial scales indicated that the benefit of the Sentinel-1 predictors in enhancing the predictive accuracy of the RF models was minimal (Figure 4.7 and Table 4.4). This disagreed with the study conducted by Bai et al. [54], who found that the use of a Sentinel-1 predictor, namely, the backscatter in VV polarization (Sigma VV), markedly increased the predictive accuracy of the RF model. One reason for this inconsistency with results herein could be due to differences in the underlying land surface and climatic conditions of the study area. Their study area was mainly located in an arid area of northern China with low vegetation cover as opposed to the Susquehanna watershed, which is dominated by forests and agriculture, with a humid climate condition. The predominance of vegetation may lower and, in some cases, cancel out the sensitivity of the backscatter signal to soil moisture (Li et al., 2018).

As stated earlier, the results of the comparative evaluation indicated the lesser importance of Sentinel-1 predictors in improving the predictive accuracy RF models. For this reason, these predictors were further left out in order to 1) reduce the computational burden associated with the processing of their images, 2) reduce the computational cost associated with the RF model because of the decrease in the dimension of predictors after the removal of the Sentinel-1 predictors and 3) increase the training sample size (as the Sentinel images had a limited temporal and spatial coverage compared to SMAP and MODIS), resulting in more robust RF models. Therefore, the evaluation of the RF model was continued with the remaining predictors, which was denoted as Exp3 (Table 4.2). The evaluation of the RF models based on Exp3 at both spatial scales showed higher performances (Figure 4.7). In addition, the underestimation and overestimation of higher and lower soil moisture values were reduced compared to Exp1 and 2. This could have been due to the large sample size used for the training of the RF model, which resulted in a more robust model and, subsequently, led to better performances.

In general, the performance evaluation results indicated that the RF models developed at both spatial scales for all the experiments showed higher performance measures for the condition of our study area, which was a warm summer humid climate with a predominance of vegetation. This agreed quite well with the results reported by previous studies conducted in different regions of the world under varying climatic and land surface conditions using the RF model, e.g., the subtropical continental monsoon (Abowarda et al., 2021), semi-humid temperate monsoon (Hu et al., 2020), South Asian monsoon (Chen et al., 2019; Zhao et al., 2019), and Mediterranean climate zone (Liu et al., 2020).

4.5.2 RF model predictors importance

In this study, predictors derived from Sentinel-1 and MODIS, together with others, such as the brightness temperature, elevation, aspect, slope and antecedent precipitation index were used to downscale L3SMP and L3SMP_E soil moisture. Many studies on the spatial downscaling of satellite soil moisture products have indicated that the use of numerous predictors with a strong connection to soil moisture generally increased the predictive power of the RF model (e.g., Abbaszadeh et al., 2019; Long et al., 2019; Zhou, et al., 2019)), which was also confirmed in this study. Nevertheless, the level of the contribution of each predictor in improving the model performance was not the same and varied from place to place and from spatial scale to spatial scale, at which the RF model was trained and verified.

The result of the analysis of the relative importance of predictors from Exp 1, at both spatial scales provided by the RF models, showed that the Sentinel-1 predictors, including Sigma VV, Sigma VH, the correlation, contrast, variance and entropy, had lower values of the percent of increase in MSE when each of them was randomly permuted (Figure 4.3 c and d). This implied that the Sentinel-1 predictors played a less significant role in the predictive accuracy of the RF models over the Susquehanna watershed. This agreed with the study by Attarzadeh et al., (2018), which reported a lesser importance of backscatter and its textural features, including the contrast, dissimilarity, energy and entropy, among others, for the retrieval of Sentinel-1 soil moisture over vegetated areas in Kenya, Africa. However, in contrast to our results, Bai at el., (2019) found that a Sentinel-1 predictor (i.e., Sigma VV) was the most important predictor among nine predictors (i.e., Sigma VV, LST, NDVI, EVI, NDWI, LAI, ALB, elevation and slope) used for the downscaling of L3SMP_E soil moisture over a semi-arid area of northern China, with minimal vegetation cover. As previously stated, the main reason for this inconsistency could have been due to a difference in the vegetative cover and climatic conditions of the study area. Besides the difference in land and climatic conditions of the study areas, the difference in predictors used for downscaling could also be the reason for the difference in accuracy reported. For example, Bai et al., (2019) used predictors derived from MODIS (e.g., LST, albedo, NDVI) and Sentinel-1 (e.g., Sigma VV), but they did not include brightness temperature derived from the SMAP and textural information derived from the Sentinel-1 images. However, in our study a combination of predictors derived from MODIS, Sentinel-1(e.g., backscatter and textural information) and SMAP (e.g., Brightness temperature both in V and H polarization) was used for downscaling SMAP soil moisture. Moreover, the difference in methodologies (e.g., different machine learning techniques, radar or optical based downscaling) used for downscaling satellite soil moisture products may also result in contrasting results.

On the other hand, among the predictors used, the vertically polarized SMAP brightness temperature (TBv) was identified as the most important predictor in all the experiments. This was because the brightness temperature carried the integrated information of the factors responsible for the variation of soil moisture, such as vegetation opacity, soil temperature, soil texture and surface roughness, among others (e.g., Choi & Hur, 2012; Liu et al., 2018; Neelam & Mohanty, 2020). This corresponded with the results of Hu et al., (2020), who found the brightness temperature in V polarization to be the most important among the predictors they used in the downscaling of SMAP soil moisture over Inner Mongolia, northern China.

Likewise, the NDVI and surface albedo appeared to exhibit a higher importance next to the brightness temperatures for both RF models developed for L3SMP and L3SMP_E. This was because our study area was predominated by vegetation, which caused the NDVI to be more sensitive to vegetation dynamics and played an influential role in controlling the evapotranspiration, thereby regulating soil moisture (Soliman et al., 2013). Similarly, the albedo also plays an important role in downscaling through its ability to control surface energy fluxes, thereby modulating surface soil moisture (Soliman et al., 2013). In addition, the NDVI and albedo were used in the parameterization of the radiative transfer model (RTM) during SMAP soil moisture retrieval, which, subsequently, impacted the downscaling process. For example, NDVI plays an important role through its control on vegetation optical depth, whereas albedo regulates the surface emissivity in the SMAP soil moisture retrieval algorithm.

In addition, the API was among the most important predictors identified in this study, which was not surprising, because the spatio-temporal variation of soil moisture was strongly influenced by precipitation from which API was calculated. It is worth nothing that, in this study, we considered API as a predictor instead of a precipitation in contrast to the studies by Abbaszadeh et al. (2019) and Long et al., (2019). This was because the soil moisture observation on a given day could be influenced by the precipitation of the past few days.

Lastly, satellite remote sensing, such as MODIS, provides a daily LST temperature at a 1 km spatial resolution, yet its continuous spatio-temporal availability is highly affected by cloud contamination, which, subsequently, resulted in some missing or abnormal values. Indeed, this constrains its usability for several applications. For example, in the downscaling of satellite soil moisture using ML techniques, such as RF as in this study, the deterioration of LST data quality due cloud cover reduced the training sample size. This was because of the exclusion of LST data, which were affected by the presence of clouds during the training of the RF model. However, in recent years, a number of studies have reported the successful reconstruction of cloud-affected LST pixels which, interestingly, would improve its usability in the downscaling of satellite soil moisture, as well as many other applications (Zeng et al., 2018; W. Zhao & Duan, 2020). In the current study, because of the weak link between soil moisture and land surface temperature in a humid climate dominated by vegetation, such as the Susquehanna watershed, the LST showed less importance. However, other SMAP soil moisture downscaling studies (e.g., Sun & Cui, 2021;Zhao et al., 2018)) in arid and semi-arid climates identified LST as an important variable in the downscaling process, because of the strong correlation between the LST and soil moisture for such climates. Similarly, the DEM, aspect and slope were also identified as relatively less important in the downscaling process of the current study.

4.5.3 Validation of downscaled soil moisture

The results of the validation of the downscaled L3SMP and L3SMP_E soil moisture showed a range performance for all the experiments depending on the spatial scale of soil moisture used as background fields and the underlying surface and climate conditions of the in-situ measurement sites. The performance measure of the downscaled soil moisture indicated a slightly higher ubRMSE than the SMAP accuracy requirement of 0,04 m^3/m^3 suggested in Entekhabi et al., (2008) across all four sites. This deviation could be partly explained by the disparity in scale between the downscaled soil moisture and in-situ soil moisture observation, and partly by the predominance of vegetation in the study area as opposed to the conditions for which SMAP's accuracy of 0,04 m^3/m^3 is asserted, which is in less vegetated areas with a vegetation water content (VWC) below 5 kg/m^2 (Entekhabi et al., 2008). It is also important to note that, although the in-situ sites were specifically established in less vegetated areas such as crops and grasses, they were located in mixed pixels of SMAP soil moisture, with a substantial fraction of forests, which in turn might have affected the validation accuracy.

Compared to Exp1 and 2, Exp3 provided relatively more temporally continuous downscaled soil moisture at both spatial scales. This was because of the high temporal availability (i.e., 1 to 3 days) of predictors used for downscaling, which in turn presented an opportunity to have longer time series during validation with the in-situ soil moisture. In addition, a better spatial coverage was obtained when using Exp3 predictors, unless there was an effect of cloud cover and a scanning gap due to the SMAP radiometer. However, for Exp1, the temporal availability of downscaled soil moisture was scarce because of the temporal/spatial constraints imposed by Sentinel-1 data, which might have affected the performance metrics calculation during validation with in-situ soil moisture. Besides metrics, the clear visualization of the temporal dynamics of the downscaled soil moisture was problematic because of the discontinuity of time series of downscaled soil moisture for the same reason mentioned above.

The study also compared the downscaled L3SMP and L3SMP_E SM against the most recent SMAP/ Sentinel-1 (L2_SM_SP) 1-km soil moisture. Accordingly, the L2_SM_SP more or less exhibited a similar temporal variation as the downscaled soil moisture and in-situ soil moisture observations. However, the performance metrics were relatively lower for L2_SM_SP compared to the downscaled soil moisture (see Table 4.3). This agreed with the L2_SM_SP validation study by Das et al.(2019) using an in-situ soil moisture observation from core validation sites. They found an inferior performance of L2_SM_SP, notably for the sites with crop cover, which were similar to the surface conditions of the Ithaca and Rocksprings sites in this study. This was partly explained by the strong effect of vegetation cover on the backscatter signal of Sentinel-1 SAR, which in turn affected the soil moisture retrieval during the downscaling process.

This study also demonstrated that the RF models tended to provide biased downscaled soil moisture extremes. Such a behavior was consistently exhibited in the models developed at both spatial scales for all the experiments. This agrees with previous studies. For instance, Long et al. (2019) found an overestimation and underestimation of low and high values, respectively, of the downscaled CCI and CLDAS soil moisture. Similarly, Hutengs and Vohland [60] found a tendency of an over and under estimation of low and high values of the downscaled LST. On the other hand, Zhao et al. (2018) found a systematic dry bias in the downscaled SMAP soil moisture. The potential reasons for these discrepancies include (1) an inadequate representation of low and high soil moisture values in the data used for the training of the RF model, (2) the smoothing effect due to the aggregation of predictors used for the training of the RF model at a coarse-spatial resolution and (3) the averaging of the predictions of the ensemble of RF trees, which was the main feature of RF to reduce bias and variance, and this in turn smoothed the final predicted soil moisture.

The study also demonstrated that both the downscaled L3SMP and L3SMP_E soil moisture showed comparable results with respect to the temporal evolution and spatial patterns. This was because the L3SMP_E soil moisture was basically derived from the oversampled L3SMP brightness temperature by the Backus–Gilbert interpolation method, indicating that they had similar characteristics. One possible advantage of L3SMP_E over L3SMP was that L3SMP_E had a resolution of 9-km, which needed a large number of pixels to cover the entire study area. This, in turn, helped to generate a large training sample size, which is important for RF modelling (and for ML models in general). For instance, about 2080 pixels of L3SMP_E were required to cover the entire study area, while it was only 130 pixels for L3SMP. Thus, the training sample size of L3SMP_E was about 16 times that of L3SMP, which helped to develop a more robust RF model, which made it more trustworthy for downscaling. In addition, this might also be the reason why the downscaled L3SMP_E showed a slightly better performance than the downscaled L3SMP.

4.5.4 Advantages and limitations of this study

A key advantage of this study was its exploitation of combined optical (i.e., MODIS) and C-band SAR (i.e., Sentinel-1) data in the downscaling of L-band passive microwave soil moisture (i.e., SMAP soil moisture) using the RF model as opposed to most of the previous studies which solely based their research on optical data for the downscaling of satellite soil moisture products. Another advantage was that even though the Sentinel-1-derived predictors did not add substantial information in the downscaling process of the present study because of the predominance of vegetation over the Susquehanna watershed, this study added noteworthy value to the growing interest in the use of SAR data. This is true in light of the increasing availability of SAR sensors with low frequencies, such as the existing C-band SAR, e.g., the Sentiniel-1 and RADARSAT constellation mission (RCM), and the upcoming SAR missions, e.g., the Earth Observing System Synthetic Aperture Radar (EOS SAR) with an S-band radar and the NASA-ISRO Synthetic Aperture Radar (NISAR) mission with an L-band satellite.

In addition, while most of the previous studies focused on the downscaling of either L3SMP soil moisture (Abbaszadeh et al., 2019; Hu et al., 2020; Zhao et al., 2018) or L3SMP_E soil moisture (Bai et al., 2019) using the RF model, this study contributed to the limited study on the use of two SMAP soil moisture products as background fields in downscaling, which notably helped to identify the most suitable SMAP soil moisture products for the conditions of our study area. Moreover, this study attempted to discover whether the recent high-resolution SMAP/Sentinel-1 soil moisture product could be potentially suitable compared to the downscaled soil moisture. Furthermore, the downscaled soil moisture had a great potential for local environmental applications such as flood forecasting and irrigation planning.

This study also had some limitations that are noteworthy to mention, which possibly offer the opportunity for further efforts to improve the spatial downscaling of coarse-resolution satellite soil moisture products. First, our study area was predominantly forested, which appeared to affect the quality of the SMAP soil moisture retrieval (i.e., L3SMP and L3SMP_E) for some of the pixels that covered the study area. Thus, the use of these products as background fields for downscaling may possibly lead to the propagation of errors through the downscaling algorithm. Recently, the NASA SMAP mission started the validation of SMAP soil moisture over temperate forests (i.e., SMAP Validation Experiment 2019–2021 (SMAPVEX19-21)), which might provide the opportunity to revise the SMAP algorithm through a proper parameterization forest canopy in the future (Colliander et al., 2020), which in turn would help in the retrieval of good-quality SMAP soil moisture over vegetated pixels.

Second, the in-situ soil moisture measurements were sparse across the study area (i.e., only four stations were available, including Avondale, Ithaca, Geneva and Rockspring), which led to the validation of downscaled soil moisture with a single in-situ soil moisture observation point within a given downscaled pixel at those four stations. However, it is known that in-situ observations provide only a point measurement of soil moisture in the immediate vicinity of the stations, which is in the order of a few centimeters, whereas the downscaled soil moisture represents the average soil moisture over an area of 1-km. This indicated the scale disparity between the in-situ soil moisture observation and downscaled soil moisture, implying that a single in-situ observation within a downscaled pixel might not be able to sufficiently reflect the downscaled soil moisture, which partly explains the observed systematic bias. Considering the spatial heterogeneity of soil moisture within a pixel, many satellite soil moisture validation studies recommended to have multiple in-situ observations within a pixels, which could then be upscaled during the validation of downscaled soil moisture (Colliander et al., 2017; Yee et al., 2016), but the present study lacked this.

Third, the spatial aggregation of high-resolution predictors, e.g., NDVI, Albedo, LST, had a smoothing effect on the extreme values, which led the training of the RF models with few extremes, as also stated by Zhao et al., (2018). Similarly, the resampling of coarse-resolution predictors, e.g., brightness temperature and precipitation, also had some smoothing effects on their extreme values, which in turn affected the prediction of the extreme soil moisture values during downscaling. Yet, this was actually not special to our study, since the currently existing downscaling methods rely on the calibration of the downscaling methods at a coarse-spatial resolution as the first step (Hu et al., 2020; Zhao et al., 2018), which caused the aggregation of high-spatial-resolution predictors to be inevitable. In addition, given that some of the predictor variables were not available at a fine resolution, e.g., precipitation, resampling to a fine resolution during the prediction of soil moisture at 1-km was also unavoidable.

Fourth, although it was found that the RF is a powerful tool for the handling of the high-dimensional datasets used in this study, it tended to overestimate lower soil moisture values and underestimate higher soil moisture values. Several studies have also reported similar behavior of RF in the spatial downscaling of soil moisture (Long et al., 2019), land surface temperature (Hutengs & Vohland, 2016) and precipitation (Shi et al., 2015). It seems likely that producing biased extreme values is an inherited behavior of the RF model. This is because the final prediction value is obtained by the averaging of the prediction of each tree, which could reduce the value range due to the smoothing effect of averaging (Wolfensberger et al., 2020; Zhang & Lu, 2012). Therefore, post processing, which entails the rescaling of the downscaled soil moisture to in-situ soil moisture observations using approaches such as linear regression or cumulative distribution function (CDF) matching, could correct the observed bias of extreme soil moisture values if a in-situ measurements network was available.

Finally, despite the fact that the downscaling process presented spatially detailed soil moisture information, their usability was recommended after the correction of bias in extreme values using the aforementioned approaches. The correction of bias in extreme values, notably higher values, is important for hydrological applications such as flood forecasting.

4.6 Conclusion

This study demonstrated the effectiveness of the RF models in the downscaling of two SMAP soil moisture products (i.e., L3SMP (36-km) and L3SMP_E (9-km)) over the Susquehanna watershed using a suite of predictors derived from different sources, including Sentinel-1 and MODIS. It appeared that the RF model developed at both spatial scales performed quite well for the underlying surface and climate condition of our study area. Among the predictors used in RF, the brightness temperature played the most important role in reinforcing the link between SMAP soil moisture and predictors, followed by the NDVI, albedo and API, whereas the Sentinel-1 predictors only had a marginal impact on the predictive accuracy of the RF model.

The validation results indicated that the time series of the downscaled L3SMP and L3SMP_E soil moisture closely agreed with in-situ soil moisture observations. Similarly, the statistical performance measures indicated that the downscaled soil moisture had a better overall performance than their original counterparts. The results also suggested that the downscaled L3SMP and L3SMP_E soil moisture reproduced the spatial pattern of their original counterparts sufficiently well, while providing detailed spatial information. Nevertheless, a slightly higher ubRMSE was obtained compared to the retrieval accuracy requirement of the SMAP mission (ubRMSE = 0,040 m^3/m^3). In addition, the downscaled extreme soil moisture values were slightly biased for both spatial scales.

The downscaled L3SMP and L3SMP_E soil moisture using Sentinel-1 predictors, along with others (i.e., Exp1, Table 4.2), only had a low temporal availability because of the long revisit time of Sentinel-1 satellites. In addition, due to the narrow swath width of Sentinel-1 compared to that of SMAP, the downscaled soil moisture only partially covered the study area, thereby reducing its spatial completeness. In contrast, the downscaled L3SMP and L3SMP_E soil moisture using other predictors rather than that of Sentinel-1 (i.e., Exp3, Table 4.2) improved the temporal availability of the downscaled soil moisture and also expanded the spatial completeness of the downscaled soil moisture apart from during the presence of clouds and the SMAP scanning gap.

The potential of L3SMP and L3SMP_E soil moisture was also explored to identify the best performing product for the condition of the study area. A comparison between the downscaled L3SMP and L3SMP_E soil moisture and between their corresponding original version showed comparable results, but with a slightly better performance by L3SMP_E soil moisture.

In future studies, one could consider the use of recently launched RADARSAT constellation mission (RCM) data, which have the temporal resolution of a few days (i.e., 4 days) compared to Sentinel-1 SAR for the downscaling of satellite soil moisture over varieties of land surfaces and climatic conditions. In addition, the upcoming Earth Observing System Synthetic Aperture Radar (EOS SAR) with an S-band radar and NASA-ISRO Synthetic Aperture Radar (NISAR) mission with an L-band satellite would offer a new opportunity, especially for the better representation of soil moisture in forested regions, since they have the capability to penetrate through the forest canopy in contrast to C-band SAR. Another possible new perspective can be improving the RF algorithm and/or bias correction after downscaling. As noted previously, as long as RF is used for downscaling soil moisture, there is a possibility of predicting biased soil moisture extremes due to its inherent behavior. Thus, improving the algorithm of the RF model in order to accommodate all ranges of soil moisture may be important, otherwise, bias correction before the use of the downscaled soil moisture for subsequent application such as flood forecasting is crucial. Moreover, one could also explore a suite of satellite soil moisture products, e.g., SMAP, SMOS, and AMSR-E (to name a few), as background fields for downscaling in order to identify the best-performing downscaled soil moisture product for the condition of the selected study area. Furthermore, future studies should also consider the potential of cascade downscaling (e.g., from 36-km to 9-km and then from 9-km to 1-km) compared to direct downscaling (i.e., directly from 36-km to 1-km). Finally, revising the soil moisture retrieval algorithms as well requires further investigation, notably over forested regions, in order to hamper the propagation of poor-quality retrieval to the downscaled soil moisture.

Chapter 5 ASSIMILATION OF THE DOWNSCALED SMAP SOIL MOISTURE

Avant-propos

Auteurs et affiliation

- Shimelis Asfaw Wakigari : étudiant au doctorat, Département de Génie Civil, Faculté de Génie, Université de Sherbrooke.
- Robert Leconte : professeur, Département de Génie Civil, Faculté de Génie, Université de Sherbrooke.

Date de soumission : 11 November 2022.

Revue : Journal of Hydrology: Regional Studies

Titre français : Evaluation de l'utilisation de différentes résolutions spatiales d'humidité du sol issue d'une désagrégation d'échelle des produits SMAP pour améliorer la modélisation hydrologique

Contribution au document : Cet article vise à évaluer l'assimilation de données d'humidité du sol à différentes résolutions spatiales après désagrégation d'échelle (1-, 3- et 9-km) dans un modèle hydrologique distribué à base physique. Deux bassins versants avec des aires de drainage très différentes sont pris comme sites d'étude : le bassin versant de la rivière Susquehanna (grande aire de drainage) et le sous-bassin issu de sa branche principale, appelé bassin versant de la Upper Susquehanna (en comparaison de petite taille). La qualité des modélisations hydrologiques d'ensemble a été évaluée selon trois aspects : l'assimilation d'humidité du sol de différentes résolutions spatiales, la fréquence de l'assimilation des données et la taille du bassin versant. Les résultats de cet article se trouvent dans le chapitre 5 de la thèse.

Résumé français

Sites d'étude

Bassins versants des rivières Susquehanna et Upper Susquehanna, situés dans le nord-est des États-Unis.

Objectif de l'étude

Cette étude explore l'utilisation de l'humidité du sol issue du satellite SMAP selon différentes résolutions spatiales afin d'améliorer la modélisation hydrologique. Une désagrégation d'échelle de l'humidité du sol issue du produit de base de SMAP d'une résolution d'environ 40-km a été faite à 1-, 3- et 9-km sur les bassins versants des rivières Susquehanna et Upper Susquehanna. Les expériences menées visent à faire une insertion direction de ces produits issus de la désagrégation d'échelle de l'humidité du sol de SMAP dans un modèle hydrologique distribué à base physique.

Perspectives hydrologiques innovantes pour les sites d'étude

Comparé au modèle sans mise à jour de l'humidité du sol, la mise à jour du modèle à partir des données de base de SMAP ou après désagrégation d'écehlle améliorent considérablement la qualité des simulations hydrologiques d'ensemble sur les deux bassins versants. De plus, la dispersion de l'ensemble est réduite et la moyenne de l'ensemble proche de l'observation. Pour le bassin de la Upper Susquehanna, le débit est le mieux simulé avec l'utilisation de l'humidité du sol à plus fine échelle, d'une résolution de 1-km, alors que pour le bassin versant de la Susquehanna, les sorties de débit du modèle hydrogogique sont de qualité équivalente avec l'utilisation du produit à 9-km de résolution ou ceux à plus fines résolutions. En plus de la mise à jour de l'humidité du sol dans la couche de surface, l'assimilation dans la seconde couche de sol de l'humidité du sol issue de l'extrapolation verticale du produit SMAP améliore légèrement la précision du modèle.

Mots clés : Humidité du sol ; SMAP ; désagrégation d'échelle ; mise à jour ; amélioration ; débit.

EXPLORING THE UTILITY OF THE DOWNSCALED SMAP SOIL MOISTURE PRODUCTS IN IMPROVING STREAMFLOW SIMULATION

5.1 Abstract

Study region

The Susquehanna and upper Susquehanna watersheds in the Northeastern of the United States of America (USA)

Study focus

This study explored the utility of the SMAP soil moisture with a range of spatial resolutions for improving the ensemble streamflow simulations. The SMAP level 3 soil moisture product with spatial resolution of roughly 36-km was downscaled to a range of spatial resolutions including 1-, 3- and 9-km over the Susquehanna and Upper Susquehanna watersheds. A set of experiments was conducted through direct insertion of the downscaled SMAP soil moisture into a physically-based distributed hydrological model.

New hydrological insights for the region

The updating of the model with the original and downscaled SMAP surface soil moisture markedly improved the accuracy of the ensemble streamflow simulations when compared to the non-updated model for both watersheds. In addition, the ensemble spread was reduced, and the ensemble mean compares well with the observed streamflow. The 1-km downscaled SMAP soil moisture showed the highest accuracy in improving streamflow simulation for the Upper Susquehanna watershed, but for the Susquehanna watershed the downscaled SMAP at 9-km adequately improved the ensemble streamflow simulations without the need to go to higher spatial resolutions. Besides the top layer of the model, updating the second layer of the model with the vertically extrapolated SMAP soil moisture slightly further improved the accuracy of the model.

Keywords: Soil moisture; SMAP; Downscaling; updating; improving; streamflow

5.2 Introduction

Soil moisture plays an essential role in the runoff generation as it controls the partitioning of incoming precipitation into surface runoff and infiltration. This is particularly important in the case of the saturation excess runoff generation mechanism (Dunne et al., 1975), where the magnitude of the flood is determined by the pre-storm soil moisture state of the watershed. Soil moisture also plays an important role in the partitioning of incoming net radiative energy into latent and sensible heat fluxes, which in turn control near-surface variables such as temperature and humidity (Entekhabi, 1995). Moreover, a host of hydrometeorological and agricultural applications including hydrological forecasting, weather forecasting and irrigation planning is heavily dependent on soil moisture state.

Conventionally, in-situ measurements are used to characterize the spatiotemporal variability of soil moisture. They provide the most accurate and reliable soil moisture information, but with low spatial representativeness (Dobriyal et al., 2012). Besides, they are time consuming and costly to establish their network over a larger area. Alternatively, satellite passive microwave remote sensing, notably at L-band frequency (1,4GHz), offers the potential to map near-surface soil moisture (approximately at 5 cm depth) over a large area with decent temporal resolution (i.e., 2 to 3 days) (Entekhabi et al., 2008; Njoku & Entekhabi, 1996). However, despite of this advantage, its spatial resolution is coarse, which hampers its use for local and regional scale hydrometeorological applications.

Many studies have attempted to downscale the coarse-spatial resolution satellite passive microwave soil moisture products using methods such as those based on the universal-triangle concept or ML among many others (Peng et al., 2017; Sabaghy et al., 2018). For example, Zhao et al. (2018), Abbaszadeh et al. (2019) and Wakigari et al. (2022) downscaled the Soil Moisture Active Passive (SMAP) soil moisture from 36- to 1-km over Iberian Peninsula, Continental United States (CONUS), and Susquehanna River Basin, respectively. Likewise, Djamai et al. (2015), Piles et al. (2016) and Zheng et al. (2021) downscaled the Soil Moisture and Ocean Salinity (SMOS) soil moisture to 1-km over the Canadian Prairies, Murrumbidgee watershed (Australia) and Huai River Basin (China), respectively.

Apart from downscaling, there has been a growing interest over the last decades in the incorporation of remotely sensed soil moisture measurements into hydrological models for improving their prediction accuracy (Brocca et al., 2011; Peng et al., 2021; Reichle, 2008). This is mainly because of the increased availability of spatially distributed soil moisture observations with decent temporal resolution, notably from satellite L-band passive microwave remote sensing, which is a well-established method for mapping of near-surface soil moisture at global scale. In addition, the key role that soil moisture plays as the main state variable of hydrological model, thereby controlling streamflow estimates, could be considered as another reason.

However, there is an ongoing argument regarding the usefulness of incorporating remotely sensed soil moisture into hydrological models to improve streamflow simulation and forecasting. While some studies have found significant improvement in simulation and forecasting of the streamflow (Brocca et al., 2012; Massari et al., 2015; Wanders et al., 2014), a considerable number of other studies have reported no to little benefit of satellite soil moisture in improving the streamflow simulation and forecasting (Alvarez-Garreton et al., 2014; Kumar et al., 2014; Parajka et al., 2006).

For example, Jadidoleslam et al. (2021) found significant improvement in streamflow simulation after assimilation of the coarse-spatial resolution SMOS and SMAP soil moisture into a distributed hydrological model. Similarly, Azimi et al. (2020) reported the marked improvement in streamflow simulation following the assimilation of Sentinel-1 (a high resolution synthetic aperture radar) and SMAP soil moisture into a hydrological model. Conversely, Kumar et al. (2014) reported slight improvement in streamflow simulation after assimilation of soil moisture retrievals from AMSR-E. Lievens et al. (2015) found improvement only in peak flows after assimilation of SMOS soil moisture, whereas Parajka et al. (2006) reported no improvement at all in streamflow simulation when ingesting ERS scatterometer soil moisture.

Overall, the lack of consensus among these studies on whether the incorporation of remotely sensed soil moisture into models improves the accuracy of streamflow simulation and flood forecasting or not might be explained by 1) the difference in models used (e.g., lumped, semi-distributed or physically-based fully distributed models), 2) methods of assimilation used (i.e., direct insertion, ensemble Kalman filter, or particle filter), 3) watershed characteristics (e.g., location, size, land cover, soil type, and dominant hydrological processes), and 4) characteristics of the assimilated remotely sensed soil moisture products (e.g., retrieval algorithm, accuracy and quality, penetration depth, spatial support, and temporal availability) (Mao et al., 2019). Moreover, as remotely sensed soil moisture represents only the upper few centimeters of the soil, the degree of its coupling with subsurface soil moisture could also determine the accuracy of hydrological model prediction (Capehart & Carlson, 1997; Chen et al., 2011b).

The contrasting results among the different studies encourage to further explore the benefit of passive microwave satellite soil moisture in improving the accuracy streamflow simulation and forecasting. Interestingly, most of the abovementioned studies have assimilated remotely sensed soil moisture at its original resolutions (i.e., 25-50-km). This resolution is deemed sufficient for continental and global scale hydrometeorological applications, yet a host of hydrometeorological and agricultural applications at local and regional scales require soil moisture at high-spatial resolution (i.e., a kilometer or less).

As mentioned above, considerable efforts have been made to downscale passive microwave satellite soil moisture products. However, the main focus of these efforts was to bridge the scale gap between the original coarse-resolution passive microwave soil moisture observations and ground measurements for the purpose of validation. Nevertheless, only few studies have explored the benefit of downscaled passive microwave soil moisture for hydrological applications, e.g. Abbaszadeh et al. (2020); López et al. (2016).

Considering the above insights, the present study focuses on assimilation of the original (36-km) and downscaled SMAP soil moisture to a range of spatial resolutions including 1-, 3- and 9-km into a physically based distributed hydrological model. Previous studies have considered only downscaled SMAP soil moisture at 1-km resolution (Abbaszadeh et al., 2020; Le et al., 2022), but in the present study the downscaled SMAP to a range of spatial resolutions are assimilated . In addition, distinctively from most of the previous studies which focused on the use of computationally demanding techniques such as ensemble Kalman filter for the integration of remotely sensed soil moisture products into hydrological models, here we used the simplest data assimilation scheme known as the direct insertion approach, as the intended purpose is to assess the added value of incorporating soil moisture for streamflow prediction.

Overall, this study aims to evaluate the utility of the insertion of the SMAP soil moisture products with a range of spatial resolutions (i.e., downscaled and original resolutions) into a physically based distributed hydrological model for enhancing of streamflow prediction.

5.3 Materials and Methods

5.3.1 Study area

The Susquehanna watershed and its head sub-watershed, the Upper Susquehanna, located in the Northeastern part of the United States with an approximate area of 71,000 and 21,000 km^2, respectively, were considered in this study, see Figure 5.1. The average annual precipitation of the Susquehanna watershed varies between 800 and 1250-mm, whereas the mean annual temperature for the northern and the southern portion of the watershed is 6 and 12 ^{o}C, respectively. The dominant land cover is forest (55 %) followed by agriculture (33 %). Peak streamflow occurs during the spring season because of the collective effect of snowmelt and rainfall, whereas low flow is usually experienced in late summer and early fall.

On the other hand, the Upper Susquehanna River Basin is characterized by a mixture of forest and farmland. Its average annual precipitation varies between 86 and 110-mm, while its minimum and maximum temperature varies between -11 and -9, and 25 and 29 ^{o}C, respectively. Its peak and low streamflow occur during spring and late summer, respectively.

5.3.2 Description of data

The SMAP soil moisture products

The SMAP level 3 soil moisture product (L3SMP) with spatial resolution of 36-km was downloaded from the NASA National Snow and Ice Data Center Distributed Active Archive Center (NSIDC DAAC) (https://nsidc.org/data/smap/ accessed date 30 May 2019). Besides the L3SMP, we also utilized the downscaled L3SMP soil moisture data. For downscaling, we trained random forest machine learning (RF ML) technique using a range of predictors such as land surface temperature (LST), normalized difference vegetation index (NDVI), albedo and elevation among many others. The L3SMP was used as response variable and the training was carried out at resolution of L3SMP (i.e., 36-km) (Wakigari & Leconte, 2022).

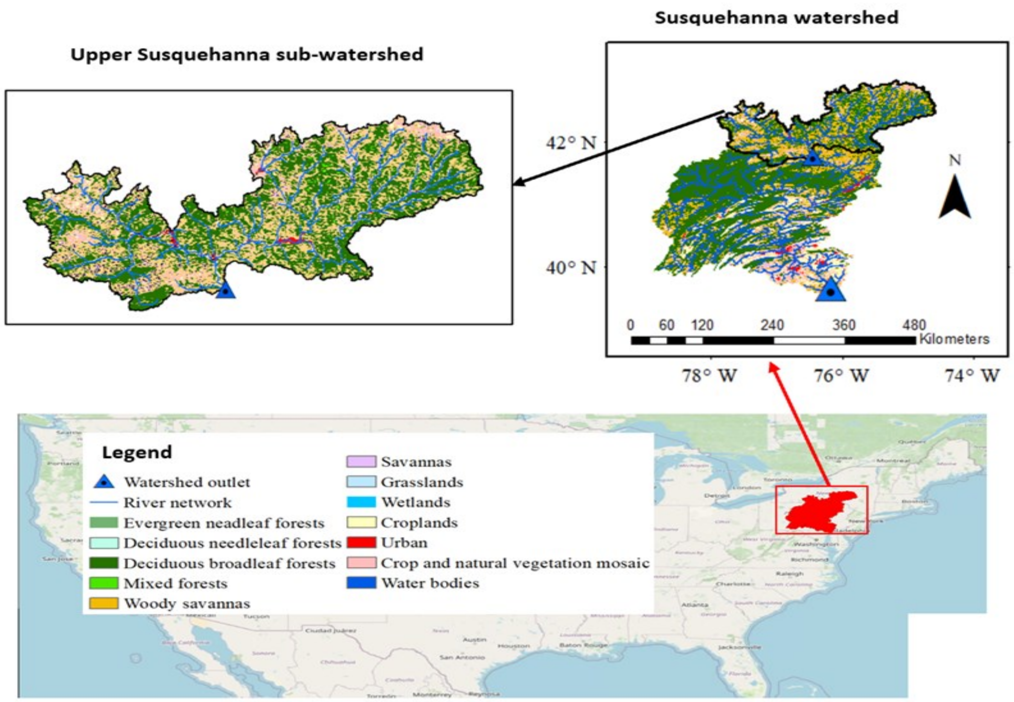

Figure 5.1 Location of the study area: Upper-Susquehanna (left) and Susquehanna watersheds (right)

The high-resolution predictors were aggregated to this resolution during the training. Finally, after verification of the model at coarse-spatial resolution, it was applied to the high-resolution predictors to predict soil moisture at higher spatial resolutions. Hence, for this study, we downscaled the L3SMP soil moisture to 1-, 3- and 9-km. For further details, the reader can refer to our recent publication (Wakigari & Leconte, 2022).

In a nutshell, the soil moisture active passive (SMAP) is the NASA's Earth satellite mission launched on January 31, 2015 for the estimation of near-surface soil moisture across the globe. It is a sun-synchronous, near-polar orbit satellite with a revisit time of 2-3 days. Its equator crossing time for ascending and descending overpasses is 6:00 PM and 6:00 AM local time, respectively. It was originally formed from combination of an L-band radar and an L-band radiometer, but the radar component ceased shortly after launch. Nevertheless, the SMAP radiometer component alone continued the generation of near-surface soil moisture roughly at 36-km resolution.

Hydrometeorological data

Daily surface precipitation and temperature (i.e., maximum and minimum) were collected from PRISM (Parameter-Elevation Regressions on Independent Slopes Model) (Daly et al., 2008), which is a gridded dataset with a 4 × 4 km grid resolution produced from a set of in-situ measurements across CONUS using terrain-aware interpolation techniques. It was developed by the PRISM Climate Group at Oregon State University (http://www.prism.oregonstate.edu)

The streamflow observations with daily temporal resolution were obtained from the United States Geological Survey (USGS) at gauge stations located at Towanda (1531500), Harrisburg (1570500), and Conowingo (1578310).

Ensemble meteorological data

The ensemble of daily precipitation and temperature data was obtained from the Ensemble Meteorological Dataset for Planet Earth (EM-Earth), which is downloaded from Federated research data respiratory (FRDR) of Canada (https://www.frdr-dfdr.ca/repo/). It is a recently developed global dataset consisting of deterministic and probabilistic estimates of daily precipitation, mean and range of daily temperature for historical period of 1950 to 2019 (Tang et al., 2022). It has a spatial resolution of roughly 10-km.

EM-Earth was produced following two sequential steps. First, the EM-Earth deterministic estimate was produced by merging station-based serially complete dataset with ERA-5 reanalysis and second, the ensemble members were generated from merged data using spatiotemporally correlated random fields. For further technical details about this data is available in Tang et al. (2022).

In this study, precipitation and maximum/minimum temperature with an ensemble size of 25 members were used to drive a distributed hydrological model. As EM-Earth did not provide the maximum and minimum temperature, they were estimated from the mean and range of daily temperature. Characteristics of all the data used in this study are summarized in Table 5.1.

Table 5.1 Summary of the data used in this study

Data	unit	Data source	Spatial resolution (km)	Temporal resolution (Days)	Length of data record (years)
Data for model setup					
Precipitation	mm	PRISM	4	1	1981-2016
Temperature	°C	PRISM	4	1	1981-2016
Streamflow	m^3/s	USGS	-	1	1981-2016
Data for ensemble streamflow simulations					
Precipitation	mm	EM-Earth	10	1	2017-2019
Temperature	°C	EM-Earth	10	1	2017-2019
SMAP soil moisture	m^3/m^3	NSIDAC	36/9/3/1	2-3	2017-2019

5.3.3 Hydrological model description

A continuous and physically based distributed hydrological model called HYDROTEL (Fortin et al., 2001) was employed. It is developed at Institut National de la Recherche Scientifique (INRS) in Quebec (Canada) with aim of making good use of GIS and to capitalize on remotely sensed data for simulation of hydrological processes (Fortin et al., 2001) in watersheds with a variety of physiographic and hydroclimatic characteristics. It is operationally in use at the Direction principale des prévisions hydriques et de la cartographie (DPPHcC) (a government organization which produces river forecasts).

In HYDROTEL, the watershed is spatially discretized into relatively homogeneous hydrological units (RHHUs) using PHYSITEL, which is a GIS framework specifically developed for HYDROTEL to spatially discretize the watershed based on river networks, elevation, land cover and soil types. Accordingly, in this study, the Susquehanna watershed was discretized into 1025 RHHUs.

This model has six major modules and each one of them has several options of simulation with a range of complexity and data requirement (Table 5.2).

Table 5.2 HYDROTEL modules and their simulation alternatives

Modules	Available options
Interpolation of meteorological data	Thiessen polygons*, Weighted mean of nearest three stations
Snow accumulation and melt	Mixed (degree-day) energy budget model, Multilayer model*
Soil temperature and soil freezing	Rankinen*, Thorsen
Potential evapotranspiration	Thornthwaite, Linacre, Penman, Priestley-Taylor, Hydro-Quebec*, Penman-Monteith
Vertical water budget Overland water routing	BV3C*, CEQUEAU Kinematic wave equation*
Channel water routing	Modified Kinematic wave equation*, Diffusive wave equation

*refers to selected sub-modules for simulation in this study

As this study focused on the assimilation of SMAP soil moisture the soil moisture module (i.e., the vertical water balance module) is briefly described herein. An extensive description of the remaining modules can be found in Fortin et al. (2001).

The vertical water balance module simulates vertical distribution of soil moisture in the soil profile. There are two possible sub-modules: BV3C and CEQUEAU. In this study, BV3C was employed. The BV3C is a French abbreviation which stands for "Bilan Vertical à 3 Couches", which can be translated as 'a three-layer vertical water budget' (Fortin et al., 2001; Jougla & Leconte, 2022). In BV3C, the soil column is vertically discretized into three layers. The first layer represents the top few centimeters (typically 5 to 10 cm) which is commensurate with the penetration depth of L-band passive microwave remote sensing satellites such as SMAP. This layer is used for partitioning of the rainfall and/or snow melt into surface runoff and infiltration. The second layer has a depth of 70 to 120 cm and is responsible for producing interflow, whereas the third layer is often close to saturation with typical depth of 200 to 400 cm, and it is responsible for producing baseflow. The water exchange between these layers is controlled by the one-dimensional Richards equation.

Calibration and validation of HYDROTEL

The dynamically dimensioned search (DDS), an automatic calibration technique, was used to calibrate model simulation against observed streamflow for the historical time period of 1981-2006 with the objective of maximizing the Nash Sutcliffe efficiency (NSE) criterion. The model was validated for the 2007-2016 time period. Both calibration and validation were performed at a daily time step. Distinct models were calibrated/validated for the Susquehanna and Upper Susquehanna watersheds. The entire Susquehanna watershed is first calibrated at the outlet and then the Upper Susquehanna watershed is separately calibrated.

It is worth noting that the calibrated model is used in this study. This is because HYDROTEL model has some conceptual representation requiring calibration to obtain better results. However, if the model was fully physically-based (e.g., WRF-HYDRO), calibration would not be that important. In addition, assimilation of brightness temperatures could be possible for fully physically based models, whereas for HYDROTEL it is not handy compared to assimilation of satellite soil moisture.

The performance of the model was evaluated using NSE and the Kling Gupta Efficiency (KGE). The NSE and KGE value ranges between $-\infty$ and 1. NSE or KGE value of 1 indicates a perfect model agreement between the simulated and observed streamflow (i.e., ideal condition). In general, the higher the NSE and KGE values, the better the performance of the model and vice versa.

5.3.4 Experimental setup

A simple assimilation technique known as direct insertion technique (Walker et al., 2001) was adopted for updating the HYDROTEL model with the SMAP soil moisture. This approach involves directly replacing the model simulated top-layer soil moisture with the SMAP soil moisture. In this study, we used assimilation interchangeably with the direct insertion. Prior to replacing the simulated soil moisture, the SMAP soil moisture was bias corrected to model simulated soil moisture using the cumulative distribution function matching approach (CDF) so that they could have similar climatology. As the main objective of this paper is to scrutinize the added value of SMAP-derived soil moisture in a hydrological model for flow simulation, the direct insertion technique was deemed appropriate compared to more complex assimilation techniques such as Ensemble Kalman Filter.

The introduction of the SMAP soil moisture into HYDROTEL model involves replacing the simulated soil moisture at each RHHUs by the model with corresponding SMAP grid cell values. If there are more than grid pixels of SMAP in given RHHUs, then their average was taken to replace RHHUs.

A suite of experiments was implemented by alternately updating the HYDROTEL model with different spatial resolutions of SMAP soil moisture including the original SMAP at 36-km and downscaled SMAP at 9-, 3- and 1-km resolutions. Model updating was carried out whenever the SMAP soil moisture with full watershed coverage is available (every 3-5 days), followed by model propagation of soil moisture (from surface soil to deeper layers) until the next SMAP soil moisture observation becomes available. Putting differently, the frequency of updating depends on the availability of SMAP soil moisture observations with full watershed coverage.

Alternatively, the model was also updated depending on important rainfall events that could generate high streamflow events, this is to investigate whether varying the frequency of updating, hence the duration of model propagation, will impact on streamflow simulation (which is relevant in an operational flow forecasting environment). In the end, a non-updated HYDROTEL model, also known as open loop, was forced with the ensemble of precipitation and temperature time series to generate ensemble streamflow and used as benchmark for comparison to other experiments (Figure 5.2).

In another experiment, in addition to the model's top-layer, its second layer was updated by using vertically extrapolated SMAP soil moisture. The adopted vertical extrapolation method was based on a simple assumption that the model simulated surface soil moisture has strong relationship with deep layer soil moisture. Thus, polynomial curve fitting was performed in order to establish an empirical relationship between model simulated surface and second layers soil moisture, which then used to vertically extrapolate the SMAP surface soil moisture. This experiment is meant to bypass the time the model takes to self adjust to new surface soil moisture observations, i.e., model propagation of soil moisture from top to bottom layer.

Here it is worthwhile to be aware that this assumption is not the best assumption, as the surface and deep layer soil moisture response to precipitation varies. In other words, the lagging of root zone soil moisture behind surface soil moisture both in wetting and drying could affect the assumption. In particular, this assumption neglects potential surface/subsurface soil moisture decoupling effects, which can be significant in fine textured soils with low vegetated surfaces (Capehart and Carlson, 1997). However, our intent is to give a general picture because of the paucity of in-situ soil moisture over the Susquehanna watershed to implement more robust method such as exponential filtering or ML techniques.

Finally, to investigate the relationship between the spatial resolution of downscaled SMAP soil moisture and size of the watershed, the model was run at the outlets of the Susquehanna watershed and its upper sub-watershed (i.e., known as the Upper Susquehanna watershed) (Figure 5.1).

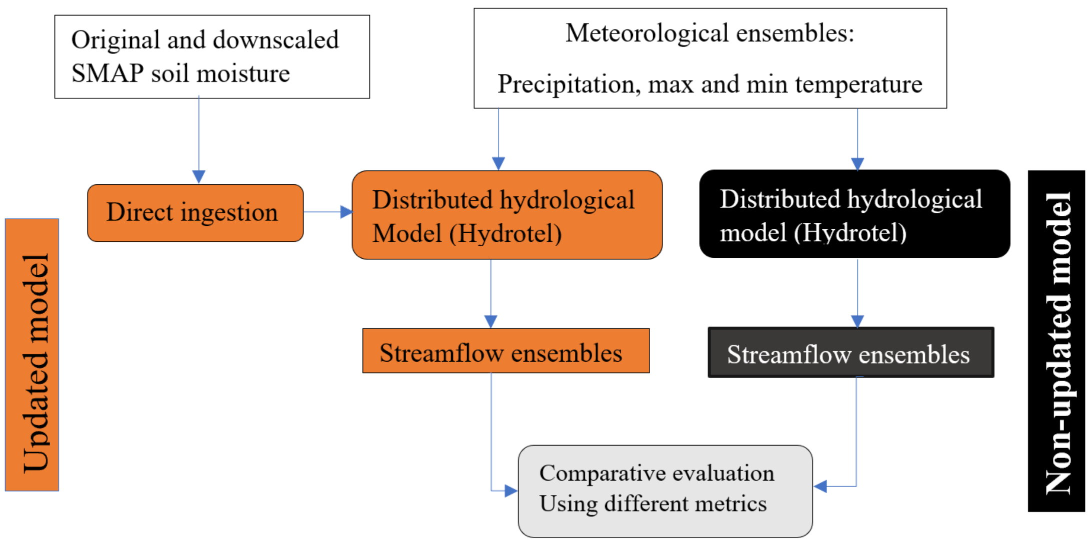

Figure 5.2 Framework of the methodology implemented

5.3.5 Evaluation of the benefit of updating the model with SMAP soil moisture

The added value of direct insertion of the SMAP soil moisture in improving model performance for prediction of the ensemble streamflow simulation was estimated by means of the efficiency index (Eff), the Normalized root mean square error (NRMSE) and the Continuous Ranked Probability Skill Score (CRPSS). The Eff (equation (5.1)) measures the effect of assimilation with respect to the open loop run. A positive Eff value indicates improvement of ensemble streamflow simulation due to assimilation of the SMAP soil moisture, whereas the negative value indicates deterioration.

$$Eff = \left(1 - \frac{\sum_{t=1}^{T}\left(Q_{\overline{up},t} - Q_{obs,t}\right)^2}{\sum_{t=1}^{T}\left(Q_{\overline{ol},t} - Q_{obs,t}\right)^2}\right) \tag{5.1}$$

where Q_{obs} is observed streamflow, $Q_{\overline{up}}$ and $Q_{\overline{ol}}$ mean of ensemble of simulated streamflow by the updated and non-updated (ol) model, respectively.

The NRMSE (equation (5.2)) represents the ratio of the root mean square error (RMSE) between predicted ensemble streamflow after soil moisture assimilation and observed streamflow to the RMSE of between the open loop ensemble and observed streamflow. It provides information about the ensemble spread and the ensemble mean performance. Its value greater than 1 indicates the deterioration of the results, whereas a value lower than 1 shows improvement.

$$NRMSE = \frac{\frac{1}{n}\sum_{i=1}^{n}\sqrt{\sum_{t=1}^{T}\left(Q_{up,t} - Q_{obs,t}\right)^2}}{\frac{1}{n}\sum_{i=1}^{n}\sqrt{\sum_{t=1}^{T}\left(Q_{ol,t} - Q_{obs,t}\right)^2}} \tag{5.2}$$

where n and T are ensemble size and length of time series considered.

In addition, the CRPSS (equation (5.3)) is employed to assess the skill of the ensemble streamflow simulation. Its score varies between $-\infty$ and 1. A score greater than zero indicates an improvement in the accuracy of the ensemble streamflow simulations due to updating of the model with the SMAP soil moisture.

$$CRPSS = 1 - \frac{CRPS_{updated}}{CRPS_{non-updated}} \tag{5.3}$$

$$CRPS = \frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{\infty}\left[\, F_i(x) \; - F_i^0(x)\,\right]^2 dx \tag{5.4}$$

$$F^0(x) = \begin{cases} 0, x \leq \; observed\ streamflow\ value \\ 1, x > \; observed\ streamflow\ value \end{cases}$$

where the CRPS is the Continuous Ranked Probability Score, $F_i(x)$ and $F^o_i(x)$ are probability density function of the member of the ensemble simulation and observation, respectively.

The summer season of 2018 (May to October) was selected for evaluation.

5.4 Results

5.4.1 Deterministic calibration and validation of HYDROTEL

Figure 5.3 displays a comparison of daily time series of observed and simulated streamflow at the outlet of the Susquehanna watershed during calibration and validation periods. The visual comparison indicates that both the timing and magnitude of simulated streamflow is in reasonable agreement with the observed streamflow, apart from the underestimation or overestimation of some of the peak events. Besides, the low flows are well reproduced by the model. Furthermore, the statistical metrics confirmed good model performance in both calibration (NSE=0,81 and KGE=0,76) and validation (NSE =0,78 and KGE=0,74). Similarly, for the Upper Susquehanna watershed the NSE (KGE) values of 0,68 (0,64) and 0,63 (0,60) were obtained during calibration and validation, respectively.

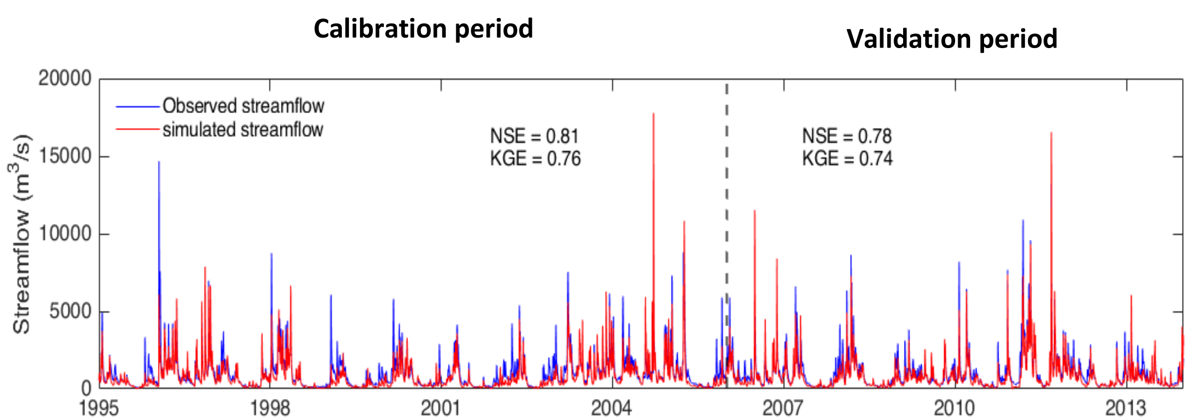

Figure 5.3 Comparison of time series of simulated and observed streamflow at the outlet the Susquehanna watershed

5.4.2 Downscaling of the SMAP soil moisture

Figure 5.4 shows an example of maps of the spatial distribution of the original SMAP soil moisture with resolution of 36-km and its downscaled version at 1-, 3- and 9-km, for 17th of June 2018. The downscaled SMAP soil moisture at all spatial resolutions well reproduced the spatial patterns of their original counterpart. For example, the eastern and western parts of the watershed are characterized by wetter soil moisture conditions, whereas the northern and southern parts are characterized by lower and medium range soil moisture, respectively both for the original and downscaled versions. In addition, at higher-spatial resolutions, e.g., at 1-

and 3-km, the effect of land covers on the downscaled soil moisture is easily discernable. This is a direct effect of including land cover related predictors in the downscaling algorithm, such as the NDVI and LST. In addition, it can be inferred from the figure that the spatial details of the SMAP soil moisture is increased after downscaling, notably the downscaled SMAP at 1-km presents more spatially detailed soil moisture information followed by downscaled SMAP at 3-and 9-km.

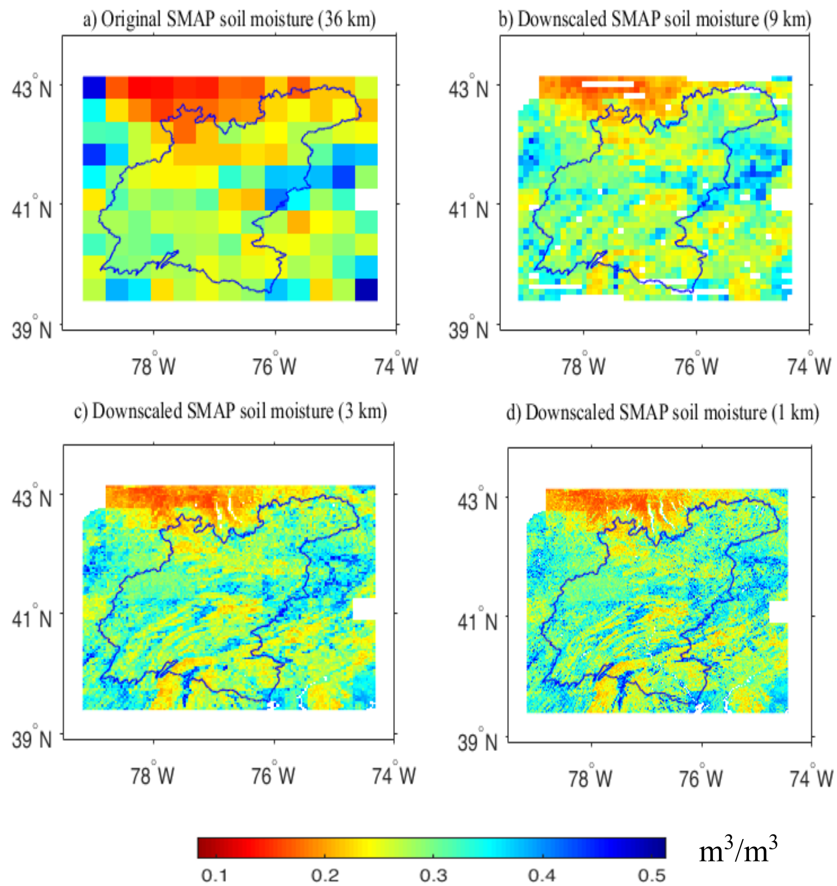

Figure 5.4 The SMAP soil moisture maps at a) 36-, b) 9-, c) 3- and d) 1-km spatial resolutions on June 17,2018 (Wakigari & Leconte, 2022)

Figure 5.5. shows maps of spatial distribution of PRISM precipitation across the Susquehanna watershed from June 13 to 16 inclusively. As can be seen in the figure the lower part of the watershed is characterized by low precipitation especially on June 13th of 2018, whereas the northeastern and northwestern parts are characterized relatively by higher precipitation. The next day, June 14th the lower part of the watershed received high precipitation, whereas the

middle-western part received precipitation with medium to high ranges. During the remaining consecutive days, the watershed almost did not receive precipitation.

Overall, the SMAP soil moisture displayed in Figure 5.4 well responded to precipitation showed in Figure 5.5. For example, soil moisture in the western and lower northeastern parts are wet due to precipitation on June 13 and 14 of 2018.

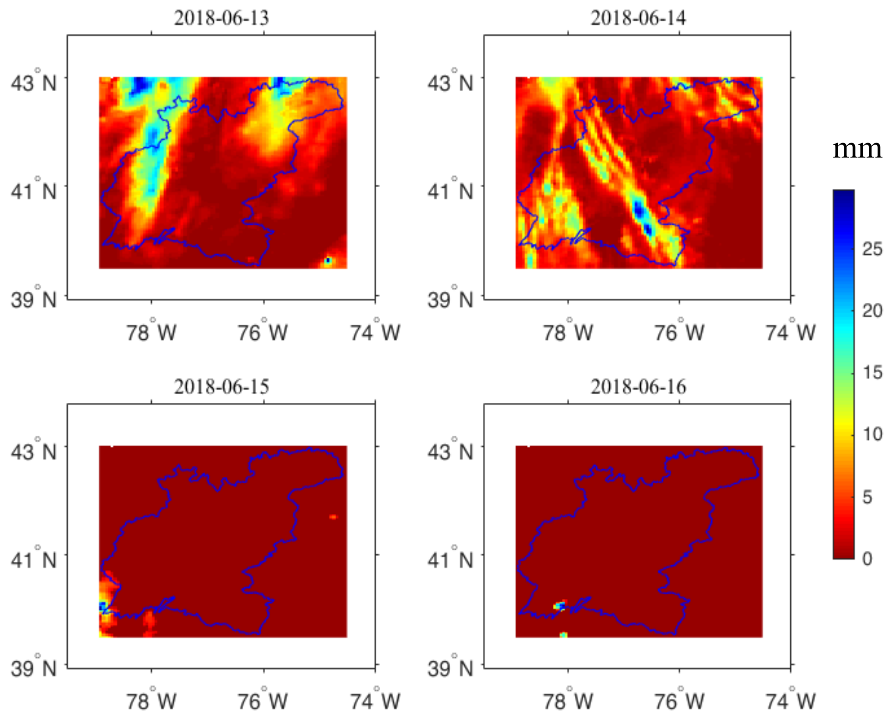

Figure 5.5 Precipitation maps over the Susquehanna watershed for four consecutive days preceding the date of SMAP soil moisture map shown in Figure 5.4

5.4.3 Updating model with SMAP soil moisture of different spatial resolutions

This section presents the results of assimilation of the SMAP soil moisture with different spatial resolutions into the HYDROTEL model to update its topsoil layer which has optimized depth of 5 cm to commensurate with depth of the SMAP soil moisture. However, only the results of the model updated by 36- and 1-km soil moisture products were presented in the figures to avoid over crowdedness, but the performance metrics for all the resolutions are presented in Table 5.4 and Table 5.5.

Figure 5.6 displays the comparison of the ensemble streamflow simulations between the non-updated (i.e., open loop) and updated model with the original (36-km) and downscaled SMAP soil moisture (1-km). As can be seen from the figure, compared to the non-updated model (Figure 5.6a), updating of the model both with the original (Figure 5.6b) and downscaled SMAP (Figure 5.6c) soil moisture reduced the ensemble spread while keeping the observed streamflow within the simulated ensemble spread. For example, during the end of July and beginning of August the non-updated model produced a large ensemble spread, but after assimilation of the original and downscaled SMAP at 1-km, the simulated ensemble peaks were reduced and matched well with the observation. Indeed, the ensemble mean also shifted towards the observation and adequately captured its magnitude and timing.

When comparing among different spatial resolutions, assimilation of the 1-km downscaled SMAP soil moisture reduced the ensemble spread and improved the simulation of ensemble mean followed by the 3- and 9-km downscaled SMAP products, and lastly, by the original 36-km version of SMAP. Corresponding statistical metrics are presented in Table 5.4.

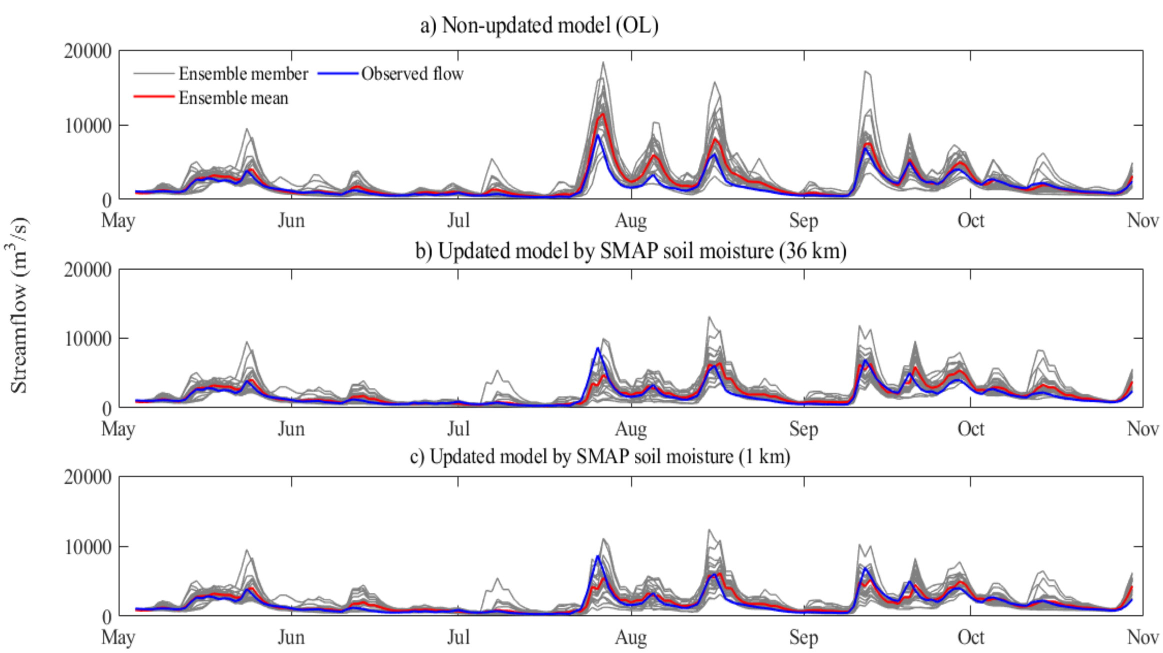

Figure 5.6 The ensemble streamflow simulations by the a) non-updated model, and updated model with SMAP soil moisture at b) 36- and c) 1-km.

5.4.4 Updating the model intermediate soil layer

The soil moisture in the intermediate (i.e., the second layer) layer plays an important role in the runoff generation. In this experiment besides the top layer of the model, the vertically extrapolated SMAP soil moisture was used to update the second layer of the HYDROTEL model, which has thickness of about 70 cm.

Figure 5.7 shows the comparison of the ensemble streamflow simulations between the non-updated and the updated model with both surface and vertically extrapolated SMAP soil moisture. Compared to the non-updated model (Figure 5.7a), the assimilation of both surface and vertically extrapolated SMAP soil moisture into model's top and second layers, respectively, reduced the ensemble spread and moved the ensemble mean closely to observation both for the cases of the original (Figure 5.7b) and downscaled SMAP at 1-km (Figure 5.7c).

When comparing among the SMAP spatial resolutions, the assimilation of the 1-km downscaled SMAP soil moisture showed better results than the original SMAP. For example, the ensemble mean of the updated model with the downscaled SMAP at 1-km closely matches the observed streamflow than when updating the model with the original SMAP (Figure 5.7). Similarly, the spread of the ensemble is slightly lower for the downscaled SMAP at 1-km. Generally, the 1-km downscaled SMAP is more effective in improving the accuracy of the ensemble streamflow simulations followed by the 3-, 9- and 36-km SMAP resolutions, respectively.

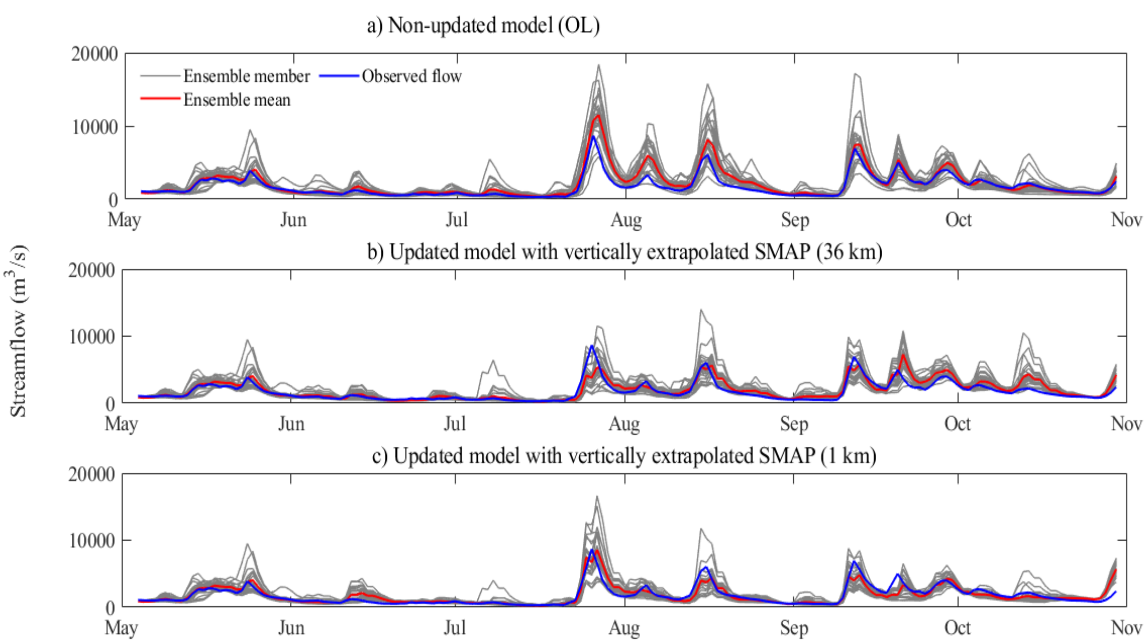

Figure 5.7 The ensemble streamflow simulations by the a) non-updated, and updated model with SMAP surface and vertically extrapolated soil moisture at b) 36- and c) 1-km

Comparing the results of updating only the top layer of the model (Figure 5.6) with the updating both the top and second layers (Figure 5.7), the latter agrees well with the observation, notably when the 1-km downscaled SMAP soil moisture is assimilated. For example, in Figure 5.6c the ensemble underestimated the observed streamflow peak near the end of July, but after updating both top and second layers, the ensemble well captured the observed peak as can be seen in Figure 5.7c. Corresponding statistical metrics are presented in Table 5.4

5.4.5 Updating the model at different frequencies

This section presents the results of the impact frequency of updating of the model on its prediction. The model was updated based on the watershed-averaged high rainfall events (Figure 5.8) prior to a runoff event, in contrast to the previous sections where the updating of the model was based on the availability of SMAP soil moisture with full watershed coverage. As can be seen in the Figure 5.8, about 10 rainfall events which produced high streamflow events were selected, of which three are very high (i.e., end of August, mid-August and mid-September). The plotted histogram represents time series of watershed averaged precipitation, whereas the blue line indicates the observed streamflow at the outlet of the Susquehanna watershed. The selection of high rainfall events was based on a simple graphical visualization approach, where the watershed averaged rainfall events are plotted along with the corresponding magnitude of streamflow. In this context, only those rainfall events which generated high streamflow events are selected visually. It is worth noting that this approach reduces the frequency of updating compared to the updating when the SMAP soil moisture with full watershed coverage is available. In other words, the model was left to propagate the soil moisture information over longer periods of time between successive updating times. The model updating dates are presented in Table 5.3 and also indicated in Figure 5.8 by vertical black dashed lines.

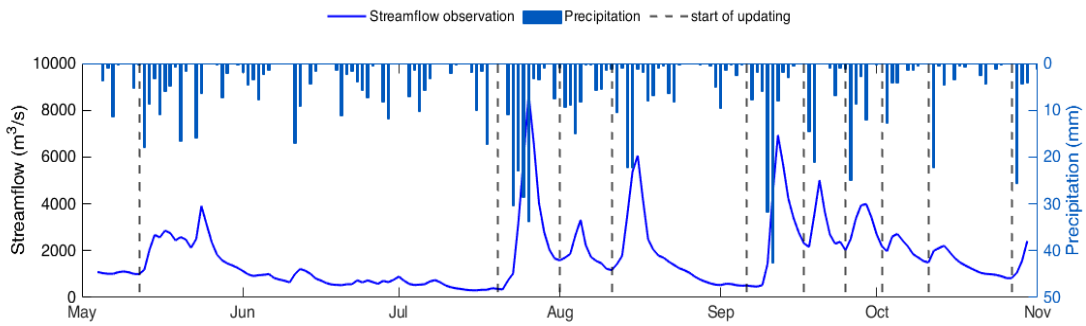

Figure 5.8 The observed streamflow at the outlet of the Susquehanna watershed and watershed averaged rainfall time series

Table 5.3 Model updating dates for the rainfall events-based updating

S.No.	Date	S.No.	Date	S.No.	Date
1	2018-05-12	5	2018-09-06	9	2018-10-11
2	2018-07-20	6	2018-09-17	10	2018-10-25
3	2018-08-01	7	2018-09-25		
4	2018-08-11	8	2018-10-02		

Figure 5.9 shows comparison of the ensemble streamflow simulations of the non-updated and updated model with the 1-km downscaled and original SMAP surface soil moisture prior to the selected rainfall events (Figure 5.8). The updating was made for the top layer of the model. After updating the model, the spread of the ensemble is reduced, and mean of the ensemble members closely matches the observation. A visual comparison between the results of updating the model based on high rainfall events (Figure 5.9) and updating of the model based on the availability of SMAP soil moisture (Figure 5.6) shows no apparent differences.

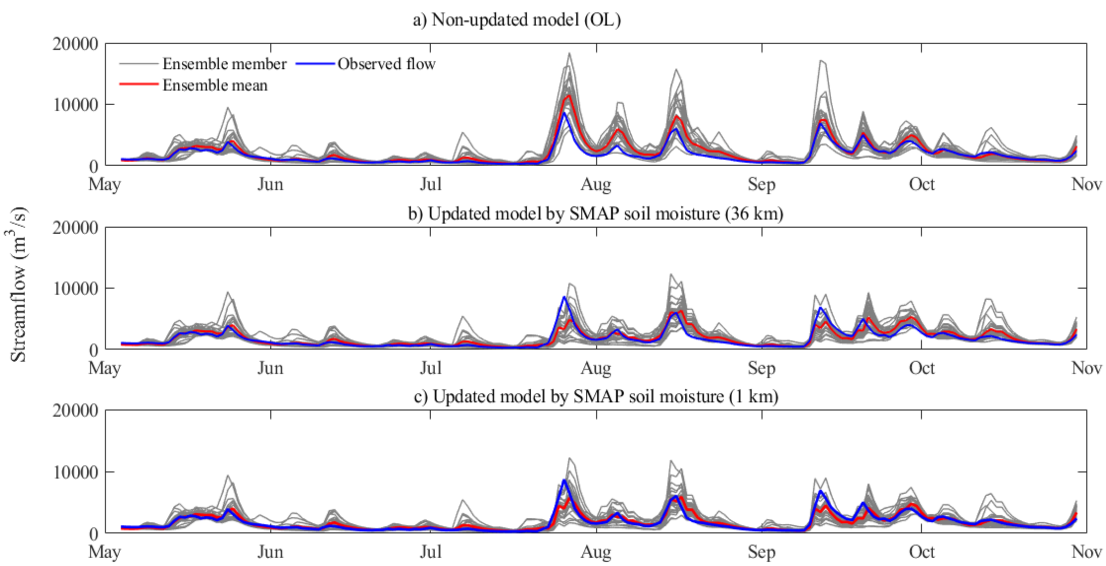

Figure 5.9 The ensemble streamflow simulations by the a) non-updated, and b) updated model with SMAP soil moisture at b) 36- and c) 1-km based on high rainfall events

Besides the visual interpretations, Table 5.4 summarizes the statistical performance of all the experiments implemented in this study for the Susquehanna watershed. The statistical metrics generally indicated the outperformance of the updated model for all the SMAP resolutions compared to the non-updated model. However, the degree of accuracy varies with the spatial resolution assimilated into the model. For example, updating with the 1-km downscaled SMAP surface soil moisture showed better performance (Eff=87,30 m^3/s, NRMSE=0,81 and CRPSS=0,17) compared to updating with the original version of SMAP surface soil moisture with resolution of 36-km (Eff=66,03 m^3/s, NRMSE=0,85 and CRPSS=0,10). Similarly, the 3- and 9-km spatial resolution generally have performance better than the original resolution of SMAP. However, downscaling to higher-spatial resolution such as 1- and 3-km did not bring significant further improvements compared to the 9-km downscaled SMAP (Table 5.4).

Table 5.4 Statistical metrics results of assimilation experiments

SMAP resolution (km)	Updated by surface soil moisture					
	Updated when SMAP available			Updated based on rainfall events		
	Eff	NRMSE	CRPSS	Eff	NRMSE	CRPSS
36	66,03	0,85	0,10	89,50	0,84	0,11
9	92,89	0,80	0,17	98,73	0,78	0,21
3	91,55	0,795	0,16	98,97	0,77	0,23
1	87,30	0,81	0,17	99,50	0,77	0,24
	Updated by surface and vertically extrapolated soil moisture					
SMAP resolution (km)	Updated when SMAP available			Updated based on rainfall events		
	Eff	NRMSE	CRPSS	Eff	NRMSE	CRPSS
36	62,20	0,87	0,03	76,86	0,84	0,10
9	91,55	0,79	0,06	99,86	0,74	0,23
3	99,58	0,80	0,04	100	0,76	0,20
1	99,70	0,80	0,04	99,72	0,76	0,17

When comparing the model updated with the SMAP surface soil moisture with the one updated with both surface and vertically extrapolated SMAP, the latter demonstrated a slightly better performance than the former in terms of Eff for the 1- and 3-km resolution. However, the NRMSE almost remain the same for all the resolutions and CRPSS is reduced for all the resolutions as shown in Table 5.4.

Similarly, when considering SMAP surface soil moisture only, the high rainfall events based updating generally improved the model performance for all the resolutions of SMAP soil moisture than when updating based on the availability of SMAP surface soil moisture. For example, updating the model with the 1-km downscaled SMAP surface soil moisture based on high rainfall events increased Eff from 87,30 m^3/s to 99,50 m^3/s, reduced NRMSE from 0,81 to 0,77 and increased the CRPSS from 0,17 to 0,24.

5.4.6 Updating the model at sub-watershed scale

The assimilation of the SMAP surface soil moisture downscaled to different spatial resolutions was also carried out at sub-watershed scale (i.e., the Upper Susquehanna watershed (Figure 5.1a)) and compared with the results obtained in the previous sections for the Susquehanna watershed. The objective here is to evaluate the impact of assimilation of different spatial resolutions of downscaled SMAP on model performance in relation to watershed size.

Figure 5.10 shows the ensemble streamflow simulations by the non-updated and updated model with the original and downscaled SMAP surface soil moisture at 1-km for the Upper Susquehanna watershed. Compared to the non-updated model (Figure 5.10a), the updated model with the SMAP soil moisture at both spatial resolutions reduced the ensemble spread while keeping the ensemble mean closer to the observation. When comparing between the 1-km and original SMAP soil moisture, for example during mid-August and end of September, the spread of the ensemble further lowered when assimilating the 1-km downscaled SMAP soil moisture (Figure 5.10b) than the original SMAP with resolution of 36 km (Figure 5.10c). Likewise, the ensemble mean agrees well with observed streamflow for the 1-km downscaled SMAP than the original SMAP. These results are comparable to those obtained over the entire Susquehanna watershed.

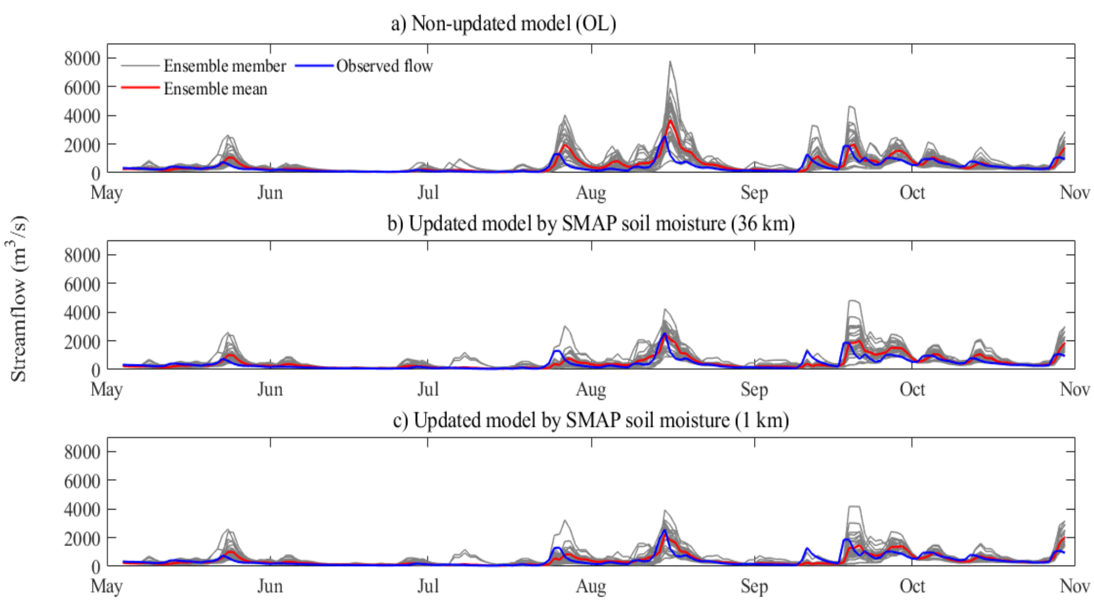

Figure 5.10 Comparison of the ensemble streamflow simulations between the a) non-updated, and b) updated model with SMAP surface soil moisture at b) 36-km and c) 1-km for the Upper Susquehanna watershed

Table 5.5 presents the comparison of the performances of assimilation of the SMAP soil moisture at watershed and sub-watershed scales. At both scales, the assimilation of the SMAP soil moisture at all spatial resolutions improved the performance of the model, as indicated by the NRMSE value less than one, the Eff and CRPSS values greater than zero (Table 5.5). Interestingly, the assimilation of the 1-km downscaled SMAP soil moisture shows more improvement at sub-watershed scale with Eff, NRMSE and CRPSS values of 97,90, 0,72 and

0,21, respectively, than at watershed scale with Eff, NRMSE and CRPSS values of 87,30, 0,81 and 0,17, respectively. Model improvement is also noted at 3-km resolution on the Upper Susquehanna, but to a lesser extent than at 1-km. Likewise, the assimilation of the 1-km vertically extrapolated downscaled SMAP soil moisture shows significant improvement at the sub-watershed scale than at the watershed scale as can be seen in Table 5.5.

Table 5.5 Statistical performance of the experiments at watershed and sub-watershed scale

SMAP spatial resolution (km)	Watershed			Sub-watershed		
	Eff	NRMSE	CRPSS	Eff	NRMSE	CRPSS
Top layer of the model updated						
36	66,03	0,85	0,10	62,76	0,77	0,14
9	92,89	0,80	0,17	90,30	0,74	0,16
3	91,55	0,795	0,16	96,80	0,73	0,19
1	87,30	0,81	0,17	97,90	0,72	0,21
Top and deep layer of the model updated						
36	62,20	0,87	0,03	63,27	0,85	0,07
9	91,55	0,79	0,06	94,83	0,81	0,08
3	94,58	0,80	0,04	99,98	0,75	0,07
1	96,70	0,80	0,04	99,42	0,76	0,06

5.5 Discussion

5.5.1 Assimilation of the SMAP soil moisture

Updating of the model through direct insertion of the SMAP soil moisture at all the spatial resolutions resulted in a considerable improvement in the accuracy of the ensemble streamflow simulations compared to the non-updated model (i.e., open loop), see Table 5.4. However, the degree of improvement varies depending on the spatial resolution of the inserted SMAP soil moisture. For example, the direct insertion of the 9-km downscaled SMAP soil moisture markedly improved the ensemble streamflow simulations and performed better compared to the original SMAP soil moisture with resolution of 36-km, whereas the insertion of the 1- and 3-km downscaled SMAP soil moisture did not bring additional improvement compared to the downscaled SMAP at 9-km.

The downscaling of the SMAP soil moisture generally improved the representation of spatial heterogeneity of soil moisture and this heterogeneity is heavily dependent on the complex interaction of intertwined geophysical and atmospheric factors such as soil properties, topography, vegetation and meteorological forcing (Crow et al., 2012). However, the effect of these factors varies based on the spatial scale. For example, at larger scale the meteorological forcing plays an important role, whereas at finer scale topography and soil properties such as texture become more crucial (Western et al., 2002). This insight supports the importance of downscaling the coarse resolution SMAP soil moisture products to higher-spatial resolutions in order to capture soil moisture spatial variability at finer scale. Indeed, possibly for this reason, the downscaled SMAP at all resolutions gave better results than the original SMAP soil moisture. However, because of the large size of the Susquehanna watershed, downscaled SMAP at 9-km adequately represented the spatial heterogeneity of soil moisture in the watershed, thereby substantially improving the streamflow simulation and reducing the importance of going to higher spatial resolutions (i.e., 1- and 3-km) as there is no further improvement in the accuracy of ensemble streamflow simulation as shown in Table 5.4.

Differently from previous studies such as that of Abbaszadeh et al. (2020) which focused on assimilation of SMAP at 1- and 36-km for the watershed with area of roughly 2,577 km^2, our study updated the model with the SMAP soil moisture with resolutions of 1-, 3- ,9- and 36-km for a large watershed with area of 71,000 km^2. Nevertheless, their results are in line with our findings showing improvement in the model prediction of streamflow after assimilation of the downscaled SMAP at 1-km compared to its original version with resolution of 36-km. Similarly, Le et al. (2022) found improvement in the accuracy of streamflow simulation after assimilation of the SMAP enhanced soil moisture (9-km) and its downscaled version at 1-km into SWAT model. They reported significant improvement in the accuracy of streamflow simulation when the downscaled SMAP at 1-km assimilated for eight watersheds with area ranging from 267 to 6430 km^2. In general, our study area is roughly 10 to 20 times larger than the study areas of Abbaszadeh et al. (2020) and Le et al.(2022), which suggested that for our case the downscaled SMAP at coarser resolution (i.e., 9-km) is adequate enough for simulation of accurate streamflow as mentioned previously.

In the context of the ongoing controversy regarding the usefulness satellite soil moisture, our study supports those studies that found the effectiveness of satellite soil moisture in improving the accuracy of model prediction (Abbaszadeh et al., 2020; Azimi et al., 2020; Luca Brocca et al., 2012; Le et al., 2022; Loizu et al., 2018b). This is because, in our study, the insertion of the SMAP soil moisture both at its original and downscaled resolutions showed considerable improvement in the ensemble streamflow simulation when compared to the non-updated model.

This study also demonstrated the effectiveness of a simple direct insertion technique for the assimilation of the SMAP soil moisture. This technique is simple and could be computationally faster than those used in many soil moisture assimilation studies such as the Ensemble Kalman Filter (Azimi et al., 2020; Lievens et al., 2015; Loizu et al., 2018a), particle filtering technique (Abbaszadeh et al., 2020) and a smoothing framework (Crow & Ryu, 2009).

5.5.2 The effect of assimilation rootzone soil moisture

Soil moisture at deeper depths has a strong impact on runoff generation (Brocca et al., 2012) . In this study, the assimilation of the vertically extrapolated SMAP soil moisture further improved the accuracy of the ensemble streamflow prediction when compared to the assimilation SMAP surface soil moisture only. However, the improvement is not that significant. For example, the NRMSE is only reduced by 0,01 for the 9- and 1-km resolutions, whereas for the 3- and 36-km there is no change. The Eff values are moderately increased, notably for the 3- and 1-km resolutions. However, the CRPSS values are reduced for all the SMAP resolutions, see Table 5.4. This inconsistency in improving the model performance among the inserted resolutions of the SMAP soil moisture could be attributed to the simple empirical polynomial fitting approach adopted for the vertical extrapolation of the SMAP surface soil moisture. This approach has no strong physical basis, as opposed to those used in the previous studies such as the exponential filter (Albergel et al., 2008; Brocca et al., 2012). In addition, the paucity of in-situ soil moisture measurements in the study area impedes the validation of the proposed method.

Compared to previous studies such as the one by Brocca et al. (2012) which reported marked improvement in the model performance after assimilation of the vertically extrapolated ASCAT soil moisture, our study found only slight improvement after updating the second layer of the model with the vertically extrapolated SMAP soil moisture. This is because besides the difference in the extrapolation method used as stated above, part of this inconsistency could be attributed to the difference in physiographic characteristics between the study watersheds. Our study watershed is large and dominated by forest where there could be strong vertical coupling between the top and second layers compared that of Brocca et al. (2012) which is small watershed with area of 137 km^2 and mainly covered with woods and croplands. Thus, in our study, updating the second layer of the model might add water to the system, thereby upsetting the water balance and resulting in only slight improvement in the streamflow prediction compared to updating the top layer only.

It is worth noting that Hydrotel has no means of compensating for the violation of water balance due to assimilation. However, it appears that the impact of assimilation propagates to the output (e.g., streamflow simulation or forecasting). Thus, correction of the model output through post-processing could be a way forward. Another possible way is improving the subsurface physics of the model. Many hydrological models including Hydrotel were not designed for data assimilation purposes. Therefore, improving the algorithm of soil moisture module is very important in order to take into account the effect of assimilation (e.g., adding or removing water to soil)

5.5.3 The effect of frequency of updating

In the present study two approaches were adopted to update the model. In the first one, the model was updated based on the availability of the SMAP soil moisture with the full watershed coverage, whereas in the second approach the model was updated based on the expected high rainfall events. In the second approach, the frequency of updating was reduced since there are only a few high rainfall events.

The decrease in the frequency of updating in the second approach further improved the accuracy of the ensemble streamflow simulation compared to the first approach. This improvement could be explained by the fact that the reduction in the frequency of updating provides more time for the assimilated SMAP soil moisture to adjust itself to the ambient conditions during the time between two successive events. Besides, the frequency of updating could be affected by the SMAP soil moisture retrieval quality, which in turn could affect the accuracy of the ensemble streamflow simulations. This simply to mean that the model performance possibly benefited from reducing the frequency of updating because the quality of the SMAP soil moisture might not be as optimal as it should be. This is primarily because of the presence of vegetation over the Susquehanna watershed which affects the retrieval quality of the SMAP (Wakigari & Leconte, 2022).

When it comes to updating based on the availability of SMAP soil moisture, this study is in line with those of Abbaszadeh et al. (2020), Jadidoleslam et al. (2021) and Le et al. (2022), who also reported improvement in model performance. However, with regard to updating based on high rainfall events, our finding disagree with the findings of Azimi et al. (2020) who reported decrease in model performance after reducing the model updating frequency with the ASCAT soil moisture for two small watersheds (i.e., Niccone (137 km^2) and Chiascio (165 km^2)) of the upper Tiber river basin (Italy).

5.5.4 The impact of watershed size

This study demonstrated that for the larger watershed (i.e., the Susquehanna watershed) the assimilation of the original resolution SMAP (i.e., 36-km) or downscaled SMAP at coarser resolutions (e.g., 9-km) per se improved the ensemble streamflow simulations. This is because the Susquehanna watershed is large in size and could be covered by a large number of pixels of the original or downscaled SMAP at coarser spatial resolutions. This in turn allows to adequately characterize the spatial variability of soil moisture in relation to runoff generation. On the other hand, for the Upper Susquehanna watershed, the 1- and 3-km downscaled SMAP soil moisture markedly improved the accuracy of the ensemble streamflow simulation. This is because these resolutions better represent the spatial heterogeneity of soil moisture for this watershed than the downscaled SMAP at 9-km or the original SMAP with resolution of 36-km. In other words, such spatial resolutions are hydrologically more commensurable with the size of the Upper Susquehanna watershed. Expressed differently, since only few pixels of coarse resolutions SMAP (e.g., 15 pixels for the 36-km product) can cover the Upper Susquehanna watershed, they are not sufficient enough to describe the spatial heterogeneity of soil moisture and its impact on watershed behavior. This is why downscaling to higher-spatial resolution is important for the Upper Susquehanna watershed.

Similar results were reported by Le et al. (2022) through the assimilation of the SMAP enhanced soil moisture (9-km) and its downscaled version at 1-km resolution for streamflow simulation in eight catchments with a range of areas from 267 to 6430 km^2 across tropical Vietnam. They found that the benefit of downscaled SMAP at 1-km is more pronounced for streamflow simulation in the smaller watersheds. This is consistent with our study which revealed that the 1-km downscaled SMAP relative to the original SMAP improved the ensemble streamflow simulation, notably for the Upper Susquehanna watershed. In general, for the smaller watershed, going to higher-spatial resolution is important to adequately characterize the spatial heterogeneity of soil moisture and its impact on hydrological processes, yet for the larger watershed going to higher-spatial resolutions does not seem to offer significant benefits, while increasing computational workload. This because for the larger watershed spatial heterogeneity tend to decrease with the watershed size (Asano & Uchida, 2010).

5.6 Conclusions

Over the last decades, the availability of spatially distributed remotely sensed observations of near-surface soil moisture has been growing. This has sparked interest in the incorporation of these observations into distributed hydrological models. Hence, this study was designed to ingest the SMAP soil moisture with a range of spatial resolutions from 1- to 36-km into a distributed hydrological model called HYDROTEL using a direct insertion technique. A set of experiments was carried out to evaluate the impact of vertical extrapolation, frequency of updating and spatial resolution on simulation of the ensemble streamflow for two contrasting watersheds (i.e., the Susquehanna and one of its head sub-watersheds (the Upper Susquehanna watershed)), mainly in terms of size.

The results showed that the assimilation of the SMAP soil moisture, at all the spatial resolutions (i.e., 1-, 3-, 9- and 36-km), improved the accuracy of the ensemble streamflow simulations in the Susquehanna watershed. When comparing the impact of spatial resolution of the SMAP soil moisture on the accuracy of ensemble streamflow simulations, the direct insertion of the 9-km downscaled SMAP soil moisture showed better accuracy and going to higher-resolution did not result in further improvement.

As the root zone soil moisture plays an essential role in the runoff generation, the present study vertically extrapolated the SMAP surface soil moisture to deep layer and assimilated into the HYDROTEL model. The results demonstrated that the insertion of the vertically extrapolated SMAP soil moisture only slightly further improved the accuracy of the ensemble streamflow simulations compared to the insertion of SMAP surface soil moisture only. This could be partly attributed to the simplicity of the assumption adopted, physiographic characteristics of the watershed (e.g., soil type and land cover) and lack of the in-situ soil moisture data for validation of the adopted extrapolation approach.

The findings of this study also demonstrated that the direct insertion of the SMAP soil moisture based on the expected high rainfall events slightly increased the accuracy of the ensemble streamflow simulations and lowered the computational time compared to the updating based on the availability SMAP soil moisture. This is because reducing the frequency of updating offers more time for soil moisture to propagate to the rootzone between two subsequent updating times.

With regards to the impact of the watershed size, the ensemble streamflow simulation at sub-watershed scale was more benefited from the 1- and 3-km downscaled SMAP soil moisture than the original (36-km) or downscaled SMAP at coarse-spatial resolution (9-km). For the Susquehanna watershed, which is larger in size (i.e., 71,000 km^2), the downscaled SMAP soil moisture at coarse-spatial resolutions could be sufficient for the optimum streamflow simulation, yet for the upper Susquehanna watershed with area of 20,000 km^2 the downscaled SMAP at higher-spatial resolutions (i.e., 1- or 3-km) is important to properly describe the spatial heterogeneity of soil moisture.

It is worth to mention that above half our study area is forested, which in turn affected the quality of the SMAP soil moisture retrieval. Indeed, the assimilation of the sub-optimal quality SMAP soil moisture retrieval over such an area can have impact on the accuracy of the simulated streamflow. Our study did not investigate this issue, which should be addressed particularly in the context of operational streamflow forecasting.

In future studies, one could also compare the simple direct insertion technique adopted in the present study with computationally demanding techniques such as the ensemble Kalman filter and particle filter. The studies which use the ensemble of the actual weather forecast data in conjunction with assimilation of satellite soil moisture products for the operational streamflow forecast is also encouraged. Finally, studies on the comparison of the assimilation of different satellite soil moisture products (passive and active microwave remote sensing) could be interesting for identifying the most suitable product for a given watershed.

Chapter 6 MERGING SMAP AND IN-SITU SOIL MOISTURE FOR IMPROVING STREAMFLOW FORECAST

Avant-propos

Auteurs et affiliation

- Shimelis Asfaw Wakigari : étudiant au doctorat, Département de Génie Civil, Faculté de Génie, Université de Sherbrooke.
- Robert Leconte : professeur, Département de Génie Civil, Faculté de Génie, Université de Sherbrooke.

Date d'acceptation : 16 Janvier 2023.

État de l'acceptation : Version finale publiée

Revue : Hydrology (MDPI)

Référence : Wakigari, S. A., & Leconte, R. (2023). Assessing the Potential of Combined SMAP and In-situ Soil Moisture for Improving Streamflow Forecast. *Hydrology*, *10*(2), 31; **https://doi.org/10.3390/hydrology10020031**

Titre français : Utilisation conjointe des données d'humidité du sol SMAP et *in-situ* pour améliorer la prévision hydrologique

Contribution au document: Cet article vise à utiliser conjointement le meilleur de l'humidité du sol issue du satellite SMAP et le meilleur de l'humidité du sol venant de mesures in-situ sur le bassin versant de la rivière Au Saumon. Le produit combiné d'humidité du sol SMAP/in-situ est assimilé dans un modèle hydrologique distribué à base physique. Une comparaison est faite en assimilant séparément l'humidité du sol issue de SMAP seul et des mesures in-situ seules. Le chapitre 6 de la thèse rend compte des résultats de cet article.

Résumé français

En hydrologie, l'humidité du sol est une variable clef pour de nombreuses applications. La variation spatio-temporelle de l'humidité du sol peut être estimée à partir des données de télédétection satellitaire (SMOS et SMAP par exemple) ou de mesures au sol in-situ, chacune ayant ses propres forces et faiblesses. Par exemple, la télédétection permet une bonne captation de l'hétérogénité spatiale en surface, alors que les mesures in-situ offrent une haute qualité de mesure aux points d'observation et permette ainsi une bonne représentation de la dynamique verticale entre données de surface et en profondeur. Par conséquent, cette étude a pour objectifs 1) de combiner les forces des produits SMAP avec les mesures in-situ et 2) d'explorer l'utilisation de cette combinaison pour améliorer les prévisions hydrologiques d'ensemble. Le site d'étude est le bassin versant de la rivière Au Saumon, drainant une surface de 1025 km^2 dans l'est du Canada. Les mesures in-situ d'humidité du sol collectées sur le bassin versant Au Saumon ont été utilisées conjointement soit avec le produit d'humidité du sol SMAP de base (9-km de résolution) soit avec sa version à fine résolution spatiale obtenue par désagrégation d'échelle (1-km). Le produit d'humidité du sol à 1-km de résolution spatiale est issu d'un modèle de forêt aléatoire basé sur le produit de base SMAP à 9-km. Un filtre exponentiel a permis de faire l'extrapolation verticale de l'humidité du sol de surface issue de SMAP. La méthode par insertion directe a été utilisée comme méthode simple d'assimilation de données pour faire la mise à jour l'humidité du sol en surface dans un modèle hydrologique distribué à base physique. Quatre produits d'humidité du sol ont alors été testées comme intrants : 1) les produits combinés de mesures in-situ avec données SMAP à 9- et 1-km de résolution, 2) le produit SMAP de base à 9-km de résolution seul, 3) celui à fine échelle à 1-km de résolution seul et 4) les mesures in-situ seules. De plus, la mise à jour des données d'humidité du sol en surface et en profondeur du modèle hydrologique utilisé est faite à partir des produits combinés SMAP/ in-situ extrapolés verticalement et des mesures d'humidité du sol en profondeur in-situ. Les résultats montrent que la désagrégation d'échelle de l'humidité du sol de SMAP à 1-km de résolution améliore la représentation de l'hétérogénéité spatiale de l'humidité du sol dans le modèle tout en conservant le schéma spatial d'ensemble du produit de base. De même, le produit combiné de l'humidité du sol de SMAP avec les mesures in-situ permet de conserver à la fois la dynamique issue des mesures in-situ et l'hétérogénéité spatiale de SMAP. La mise à jour de l'humidité du sol en surface du modèle avec le modèle combiné de l'humidité du sol SMAP à 1-km de résolution avec les mesures in-situ améliore les prévisions d'ensemble. Les résultats de prévision sont meilleurs qu'avec les autres produits d'humidité du sol considérés, que ce soit le produit de base de SMAP ou les mesures in-situ seules. Par ailleurs, la mise à jour des couches de surface et profonde avec respectivement les produits combinés SMAP/ in-situ de surface et issus de l'extrapolation verticale ne permettent pas d'améliorer la qualité des prévisions hydrologique d'ensemble. Finalement, cette étude démontre le potentiel pour la prévision hydrologique de l'utilisation conjointe des données d'humidité SMAP avec les mesures in-situ.

Mots clefs : humidité du sol ; SMAP ; in-situ ; combinaison ; assimilation de données ; désagrégation d'échelle

ASSESSING THE POTENTIAL OF COMBINED SMAP AND IN-SITU SOIL MOISTURE FOR IMPROVING STREAMFLOW FORECAST

6.1 Abstract

Soil moisture is an essential hydrological variable for a suite of hydrological applications. Its spatio-temporal variability can be estimated using satellite remote sensing (e.g., SMOS and SMAP) and in-situ measurements. However, they both have their own strengths and limitations. For example, remote sensing has the strength of maintaining the spatial variability of near-surface soil moisture, while in-situ measurements are accurate and preserve the dynamics range of soil moisture at both surface and larger depths. Hence, this study is aimed at 1) merging the strength of SMAP with in-situ measurements and 2) exploring the effectiveness of merged SMAP/in-situ soil moisture in improving ensemble streamflow forecasts. The conditional merging technique was adopted to merge the SMAP enhanced soil moisture (9-km) and its downscaled version (1-km) separately with the in-situ soil moisture collected over the au Saumon watershed, a 1025 km^2 watershed located in Eastern Canada. The random forest ML technique was used for downscaling of the near-surface SMAP enhanced soil moisture to 1-km resolution, whereas the exponential filter was used for the vertical extrapolation of the SMAP near-surface soil moisture. A simple data assimilation technique known as direct insertion was used to update the topsoil layer of a physically-based distributed hydrological model with four soil moisture products: 1) the merged SMAP/in-situ soil moisture at 9- and 1-km resolutions; 2) the original SMAP enhanced (9-km) , 3) downscaled SMAP enhanced (1-km), and 4) in-situ surface soil moisture only. In addition, the vertically extrapolated merged SMAP/in-situ soil moisture and subsurface (rootzone) in-situ soil moisture were used to update the intermediate layer of the model. Results indicate that downscaling of the SMAP enhanced soil moisture to 1-km resolution improved the spatial variability of soil moisture while maintaining the spatial pattern of its original counterpart. Similarly, merging of the SMAP with in-situ soil moisture preserved the dynamic range of in-situ soil moisture and maintained the spatial heterogeneity of the SMAP soil moisture. Updating of the top layer of the model with the 1-km merged SMAP/in-situ soil moisture improved the ensemble streamflow forecast compared to the model updated with either the SMAP enhanced or in-situ soil moisture alone. On the other hand, updating the

top and intermediate layers of the model with surface and vertically extrapolated SMAP/in-situ soil moisture, respectively, did not further improved the accuracy of the ensemble streamflow forecast. Overall, this study demonstrated the potential of merging the SMAP and in-situ soil moisture for streamflow forecast.

Keywords: soil moisture; SMAP; in-situ; merging; assimilation; downscaling

6.2 Introduction

Soil moisture plays a paramount role in the land-atmosphere interactions by controlling the exchange of water and energy between the land surface and the atmosphere (Entekhabi, 1995; Seneviratne et al., 2010a) and this in turn controls a number of hydrological processes including infiltration, evapotranspiration and runoff generation. It also plays an essential role in many hydrometeorological and agricultural applications including flood forecasting, weather forecasting and irrigation water management (Brocca et al., 2016).

Knowledge of soil moisture spatial and temporal variability is important to improve our understanding of its role in hydrological processess and applications. Traditionally, in-situ measurements are used to characterize the spatio-temporal variability of soil moisture, but they are scarce and costly to implement over a large area (Dobriyal et al., 2012). On the other hand, remote sensing has become an invaluable alternative for global mapping of near-surface soil moisture in the recent decades (Entekhabi et al., 2008; Kerr et al., 2001).

Passive microwave remote sensing at lower frequency bands (e.g., L-band) has become an established technique for mapping of near surface soil moisture (i.e. up to 5-10 cm) because of its high sensitivity to soil moisture and high capability to penetrate through cloud and vegetation canopy notably at low to moderate vegetation density (Jackson & Schmugge, 1989; Schmugge et al., 1986). Currently, there are two L-band passive microwave satellite missions which are fully dedicated to soil moisture measurements: Soil Moisture and Ocean Salinity (SMOS) and Soil Moisture Active Passive (SMAP). Besides the L-band, there are other sensors with X- and C-bands such as the Advanced Microwave Scanning Radiometer (AMSR-E/2), the Special Sensor microwave/Imager (SSM/I) and the Special Sensor Microwave Imager/Sounder (SSMIS) passive microwave sensor have been used to infer soil moisture.

Nevertheless, the coarseness of the resolution of soil moisture products derived from passive microwave satellites remains one of the major challenges for its use at local and regional scale hydrological applications. To address this, many studies have attempted to downscale such soil moisture products, e.g., Abbaszadeh et al. (2019), Djamai et al. (2015), and Wakigari & Leconte, (2022). These studies used a range of downscaling approaches, from simple polynomial fitting based on the universal triangle/trapezoid techniques (e.g., Piles et al. (2011), Djamai et al. (2015)); to more advanced ML techniques such as random forest (e.g., Bai et al. (2019), Wakigari & Leconte, (2022)), neural network (e.g., Alemohammad et al. (2018)) and support vector machines (e.g., Srivastava et al. (2013)).

Given the increasing availability of remotely sensed soil moisture products, their assimilation into hydrological models for improving streamflow simulation and forecast has received much attention in recent decades. A number of studies have assimilated soil moisture derived from different satellites (e.g., SMOS, SMAP, AMSR-E, and Sentinel-1) into different hydrological models to improve the accuracy streamflow simulation and forecast (Abbaszadeh et al., 2020; Azimi et al., 2020; Loizu et al., 2018a; Massari et al., 2018a). However, the degree of improvement varies from no (Han et al., 2012) or minor improvement (Lievens et al., 2015) to marked improvement (Le et al., 2022; Patil & Ramsankaran, 2017). Indeed, it is difficult to compare and generalize the results of these studies as they are based on different soil moisture assimilation techniques (e.g., ensemble Kalman filter, direct insertion or particle filter), model structures (fully/semi-distributed and lumped) and physiographic characteristics of the watershed among many others (Loizu et al., 2018b; Mao et al., 2019).

When looking specifically at the effect of physiographic characteristics among the stated factors, the success of assimilation of remotely-sensed soil moisture critically depends on the quality of the assimilated soil moisture product (Vereecken et al., 2015), which in turn depends on the study area's physiographic characteristics. For example, in low to moderately vegetated areas the effect of vegetation on soil moisture retrieval is minimal (Entekhabi et al., 2008) and this allowed a number of soil moisture assimilation studies to be carried out in such areas to take advantage of good quality soil moisture retrieval (Abbaszadeh et al., 2020; Azimi et al., 2020; Cenci et al., 2016; Loizu et al., 2018b).

On the other hand, over densely vegetated areas, it is difficult to obtain satellite soil moisture with the desired accuracy e.g., the unbiased root-mean-square error (ubRMSE) of 0,04 m^3/m^3 set by some of the satellite missions such as SMAP and SMOS (Entekhabi et al., 2008; Kerr et al., 2001). Hence, we deem that the use of remotely sensed soil moisture products together with in-situ soil moisture measurements would be a potential way forward for such areas. Put differently, the merging of the strengths of SMAP and in-situ soil moisture has the potential to generate a better-quality soil moisture product than any single one of them.

Therefore, the present study aims to explore the utility of the combination of in-situ soil moisture with the SMAP enhanced soil moisture in improving streamflow forecast skills for a small heavily-forested watershed located in Eastern Canada

6.3 Materials and methods

6.3.1 Study area

The au Saumon Watershed situated in Eastern Canada was selected for this study (Figure 6.1). It has a drainage area of 1025 km^2. It receives annual precipitation (rain and snow) of roughly up to 1250 mm with an average annual temperature of 4.5 ºC, whereas the average summer precipitation and temperature are 760 mm and 19.5 ºC, respectively.

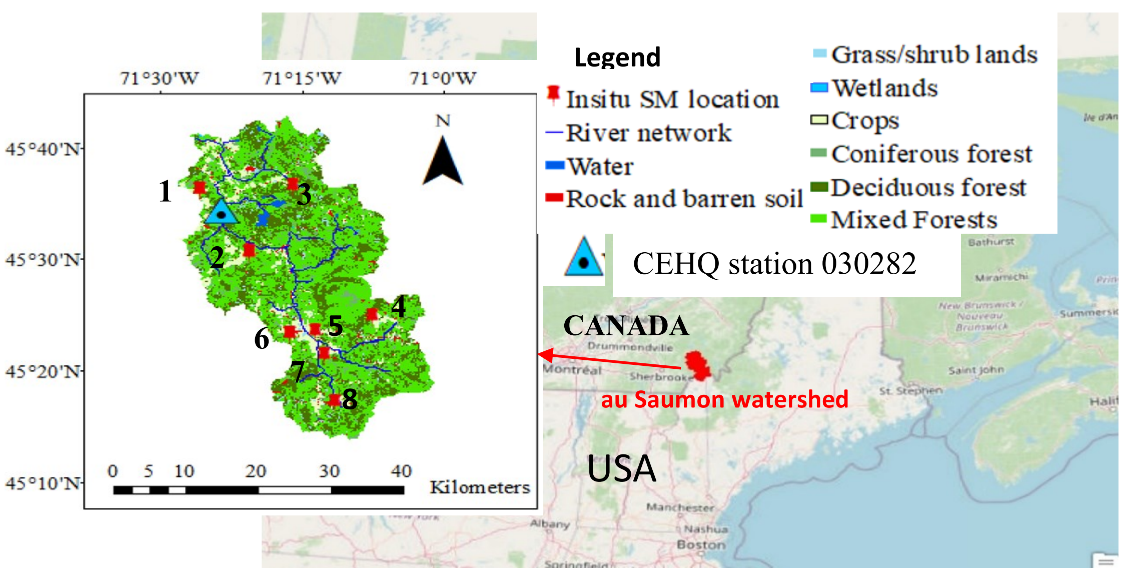

Figure 6.1 Study are location and location of in-situ soil moisture

The watershed often experiences high flows in spring and fall from snow melt and rainfall, respectively. Its elevation varies between 277 and 1092 m. Forest is the main land cover type of this watershed.

6.3.2 Data

SMAP enhanced soil moisture product

The SMAP enhanced level 3 passive microwave soil moisture product (SPL3SMP_E) with a daily global coverage was selected for this study. This product is derived from a daily composite of SMAP enhanced L2 half-orbit products, which in turn is generated from the SMAP Enhanced L1 Gridded Brightness Temperature Product (L1CTB_E) using the Backus-Gilbert (BG) optimal interpolation technique (Chaubell et al., 2016). SPL3SMP_E has a spatial resolution of 9-km displayed on Equal-Area Scalable Earth (EASE) Grid 2.0. Its descending (06:00 local time) and ascending (18:00 local time) orbits soil moisture products are retrieved separately, yet for the present study the SPL3SMP_E descending product was selected because at this time there is a better thermal equilibrium between soil surface and vegetation layer. These data can be accessed freely from the NASA Snow and Ice Data Center (NSIDC) (https://nsidc.org/data/smap/).

In addition, a downscaled SMAP enhanced soil moisture to 1-km was used. For downscaling, the random forest (RF) machine learning technique was employed. Its implementation involves training of RF with predictors derived from MODIS such as land surface temperature, NDVI SMAP TB and albedo (Wakigari & Leconte, 2022). In addition, topographic derivatives such as elevation, slope and aspect were used. It is worth mentioning that the Sentinel-1 predictors were not used as predictors for downscaling of SMAP enhanced soil moisture on the au Saumon watershed. This is mainly because the Au Saumon watershed is almost fully forested (> 80 %) which makes penetration of active sensors signal through vegetation canopy difficult by reducing its sensitivity to soil moisture. Training was carried out at resolution of the SMAP enhanced product which is 9-km. After training and testing RF at 9-km spatial resolution, it was used to estimate soil moisture at 1-km from the 1-km resolution predictors assuming the developed model is spatial scale invariant (Wakigari & Leconte, 2022). It is worth noting that for the au Saumon watershed the SMAP enhanced (9 km) is downscaled instead of SMAP L3

(36 km) this mainly because many pixels of SMAP enhanced soil moisture can cover the entire watershed than pixels of SMAP L3. This helps to have more training data for RF model, which is strongly sensitive to the training data size.

In-situ soil moisture

In-situ soil moisture observations collected during the summer season of 2019 over the au Saumon watershed using EC-5 soil moisture probes (Kizito et al., 2008) were used. A total of 8 soil moisture probes were installed at 8 selected representative locations (i.e., open and forested sites) to collect hourly volumetric soil moisture at depth of 5 and 20 cm (Sites numbered 1 to 8, see Figure 6.1).

Hydrometeorological data

Daily deterministic precipitation and temperature (i.e., maximum and minimum temperature) were derived from MSWEP (Multi-Source Weighted-Ensemble Precipitation) (Beck et al., 2017) and ERA-5 land (Hersbach et al., 2020), respectively, for the period 1980 to 2018. Similarly, the daily streamflow observations at station 030282 (draining an area of 769 km^2, Figure 6.1) were obtained from the Centre d'expertise hydrique du Québec (now the Direction de l'expertise hydrique du Québec, a provincial agency whose mandate is to manage Québec's water regime) for the same period.

Apart from the deterministic data, an ensemble of daily precipitation and maximum and minimum temperature were forced to the model to produce ensemble streamflow forecasts for the summer season 2019. The forcing data was extracted from the EM-Earth (The Ensemble Meteorological Dataset for Planet Earth) data (Tang et al., 2022). This data has a spatial resolution of roughly 10-km, and it is available from 1950 to 2019. It has 25 ensemble members which can be used for ensemble hydrological simulations.

6.3.3 Hydrological model

The HYDROTEL model was selected for experimenting on the effect of assimilation of soil moisture measurements on the accuracy of model prediction. HYDROTEL is a physically based, semi-distributed and continuous time hydrological model developed at the Institut National de la Recherche Scientifique Eau Terre Environnement (INRS-ETE), Québec (Canada)

(Fortin et al., 2001). It has attractive features such as the easiness in the incorporation of spatially distributed GIS and remotely sensed data such as soil moisture and snow water equivalent and minimal meteorological data requirement (i.e., only precipitation and maximum and minimum temperature).

PHYSITEL, a GIS tool accompanied with HYDROTEL, was used to spatially discretize the watershed into relatively homogeneous hydrological response units (RHHUs) based on elevation, land cover, soil type and river networks. Accordingly, the au Saumon watershed was discretized to 205 RHHUs with average area of 5,0 km^2.

HYDROTEL is composed of six modules: 1) interpolation of precipitation, 2) accumulation and melt of snowpack, 3) potential evapotranspiration estimation, 4) vertical water budget (i.e., soil moisture module), 5) surface and subsurface flow generation, and 6) river flow routing. Among these modules, the vertical water budget module (i.e., soil module) was used for the incorporation of remotely sensed and in-situ soil moisture observations into the model. This module has two sub-modules: BV3C and CEQUEAU. The BV3C was selected for this study. BV3C vertically discretizes a soil column into three layers. The first layer has a depth which is normally 5 to 10 cm, and it controls the partitioning of rainfall into surface runoff and infiltration. The second and third layers have typical depth of 60 to 80 cm and 120 to 200 cm, respectively and they are used to control the generation of interflow and baseflow, respectively. The water exchange between the layers is controlled by the Richards-1D equation (Fortin et al., 2006).

Model calibration and validation

The HYDROTEL model was calibrated and validated against the observed daily streamflow at station 030282 (see Figure 6.1) for the period 1980 to 2008 and 2009 to 2018, respectively. The dynamically dimensioned search-uncertainty analysis (DDS-UA) algorithm (Tolson & Shoemaker, 2007) was applied for calibration using the Nash Sutcliffe efficiency (NSE) as the objective function. The Kling Gupta efficiency (KGE) was also used for evaluation of the model performance.

6.3.4 Merging of SMAP and in-situ soil moisture

The conditional merging (Sinclair & Pegram, 2005) is a spatial interpolation technique initially developed to merge radar and rain gauge rainfall measurements. Here, it was used to merge gridded SMAP soil moisture with in-situ measurements. The major advantage of this method is that it preserves the spatial covariance structure of the grid-based measurement (i.e., SMAP soil moisture) while maintaining the accuracy of in-situ measurements. Its implementation involves six successive steps: a) extraction of SMAP grid points that covers the study area, b) interpolation of in-situ soil moisture measurements using ordinary kriging to the regular grid points of SMAP, c) extraction of the SMAP soil moisture corresponding to in-situ soil moisture measurement locations, d) interpolation of extracted SMAP soil moisture to the regular grid points of SMAP, e) estimation of the residual between interpolated SMAP (d) and extracted SMAP soil moisture (a), and f) addition of the estimated residual to interpolated in-situ soil moisture (b) to produce final merged SMAP /in-situ soil moisture. These steps are summarized in Figure 6.2

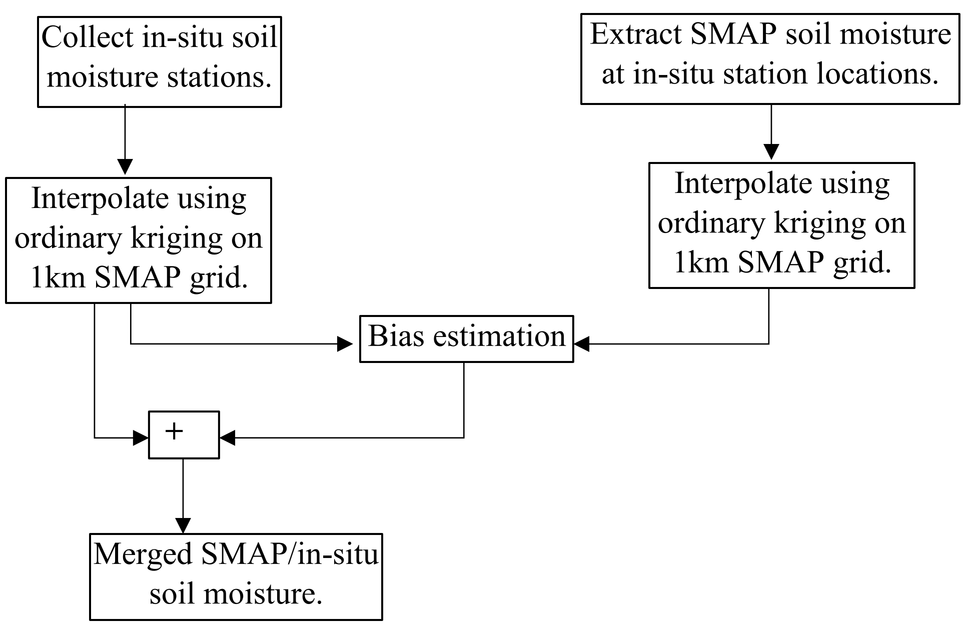

Figure 6.2 Flow chart of conditional merging

Both the SMAP enhanced (9-km) and the 1-km downscaled SMAP enhanced soil moisture are merged with in-situ soil moisture to see the impact of the resolutions of merged SMAP/in-situ soil moisture on the streamflow forecasting. In addition, to see the impact of interpolation of in-situ soil moisture into different grids (i.e., to 9 and 1-km grid). It is worth noting that for the SMAP/in-situ at 9-km the in-situ soil moisture was extrapolated to 9-km grid, whereas for SMAP/in-situ at 1-km the interpolation of in-situ was done over 1-km grid of SMAP.

6.3.5 Model updating with soil moisture

A set of experiments was implemented by updating the HYDROTEL model with the SMAP enhanced (9-km spatial resolution); downscaled SMAP enhanced (1-km); in-situ soil moisture; and merged SMAP/in-situ soil moisture at 9- and 1-km spatial resolutions, for the streamflow forecasts. The updating was based on replacing the model top layer with SMAP or in-situ soil moisture or their merged versions using the simplest data assimilation technique known as direct insertion (Walker et al., 2001). The summary of implemented experiments is displayed in Figure 6.3.

Besides the model top-layer, the intermediate layer (i.e., the second layer) of the model was updated by vertically extrapolated SMAP soil moisture using a semi-empirical approach called exponential filter, see equation (6.1) (Wagner et al., 1999). This approach has a single parameter T (characteristics time length) which is used to indicate the temporal variability of soil moisture in the root zone profile and formulated as in equation (6.2).

$$SMrz_{tn} = SMrz_{tn-1} + K_n(SMsurf_{tn} - SMrz_{tn-1}) \tag{6.1}$$

$$K_n = \frac{K_{n-1}}{K_{n-1} + \exp\left(\frac{-1}{T}\right)} \tag{6.2}$$

where SMsurf and SMrz surface and root zone soil moisture, T is the optimal characteristic decay time, tn is time step and K is gain.

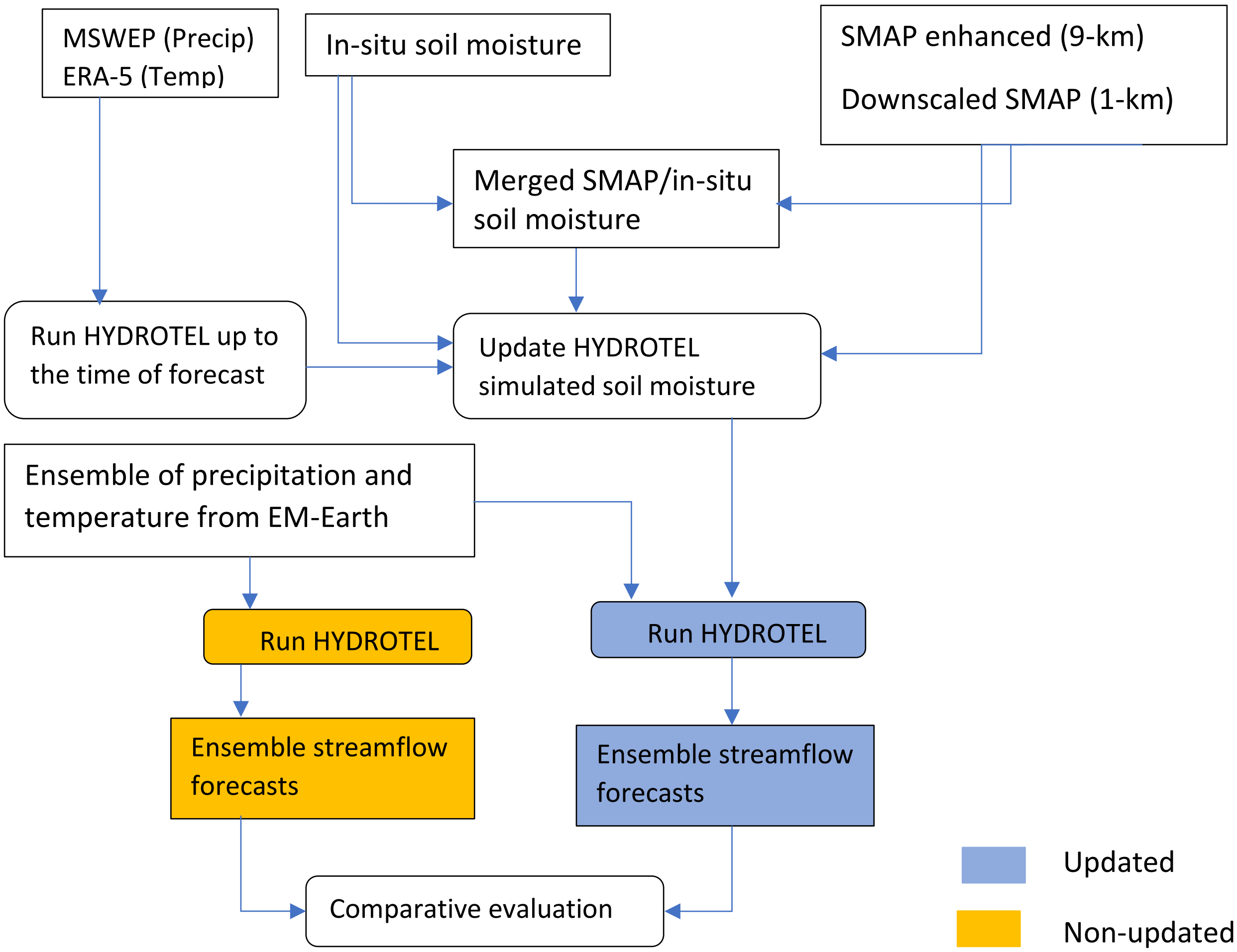

Figure 6.3 Framework of implemented experiments

6.3.6 Evaluation of deterministic and ensemble streamflow forecast skills

The accuracy of deterministic forecast was evaluated using the mean absolute error (MAE), which was calculated using equation (6.3). The lower the MAE, the better the performance.

$$MAE = \frac{\sum_{i=1}^{n} \left| Q_{sim,i} - Q_{obs,i} \right|}{n} \tag{6.3}$$

where Q_{sim} and Q_{obs} are simulated and observed streamflow, respectively. n is the length of time series.

On the other hand, the Continuous Ranked Probability Skill Score (CRPSS) was used to evaluate the skill of the ensemble streamflow forecast (equation (6.4)). This score has a value range between $-\infty$ and 1. A score greater than zero indicates an improvement in ensemble streamflow forecast skill due to updating of the model with the SMAP or in-situ or the merged SMAP / in-situ soil moisture.

$$CRPSS = 1 - \frac{CRPS_{updated}}{CRPS_{non-updated}} \tag{6.4}$$

$$CRPS = \frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{\infty}\left[\,F_i(x)\ - F_i^0(x)\,\right]^2 dx \tag{6.5}$$

$$F^0(x) = \begin{cases} 0, x \leq \ observed\ streamflow\ value \\ 1, x > \ observed\ streamflow\ value \end{cases}$$

where the CRPS is the Continuous Ranked Probability Score, $F_i(x)$ and $F^o_i(x)$ is probability density function of the member of the ensemble simulation and that of observation, respectively.

6.4 Results

6.4.1 Model performance evaluation

Figure 6.4 shows the comparison of observed and simulated hydrographs for a portion of calibration and validation periods to allow for clear presentation. The qualitative inspection of the figure shows a good agreement between the simulated and observed hydrographs, except for overestimation or underestimation of some of the peak flows.

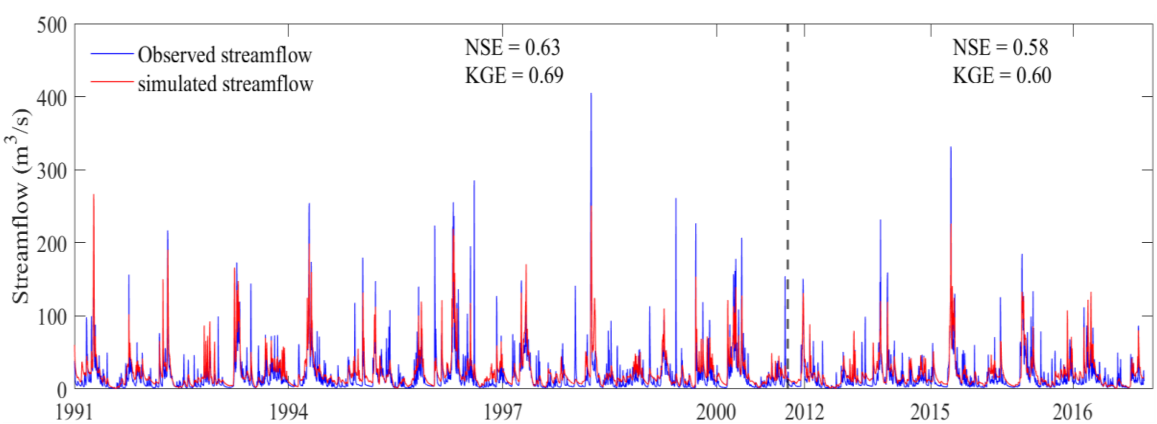

Figure 6.4 Comparison of observed and simulated streamflow hydrographs at station 030282

In addition, the statistical metrics indicate a good model performance with NSE (KGE) values of 0,63 (0,69) and 0,58 (0,60) during calibration and validation periods, respectively.

Figure 6.5 shows an example of maps of spatial distribution of SMAP enhanced (i.e., both the original (9-km) and downscaled version (1-km)), in-situ soil moisture and the merged SMAP/in-situ soil moisture at 9- and 1-km resolutions on the 8th of August 2019. As can be observed in the figure, the 1-km downscaled SMAP enhanced soil moisture (Figure 6.5d) brings out the fine-scale spatial heterogeneity of soil moisture over the au Saumon watershed compared to its original counterpart (Figure 6.5a). It also maintains the spatial pattern of its original counterpart (i.e., the original SMAP enhanced soil moisture (9-km)) For example, the drier soil moisture condition of the original SMAP enhanced soil moisture in the western part of the watershed is clearly reflected in downscaled SMAP enhanced soil moisture.

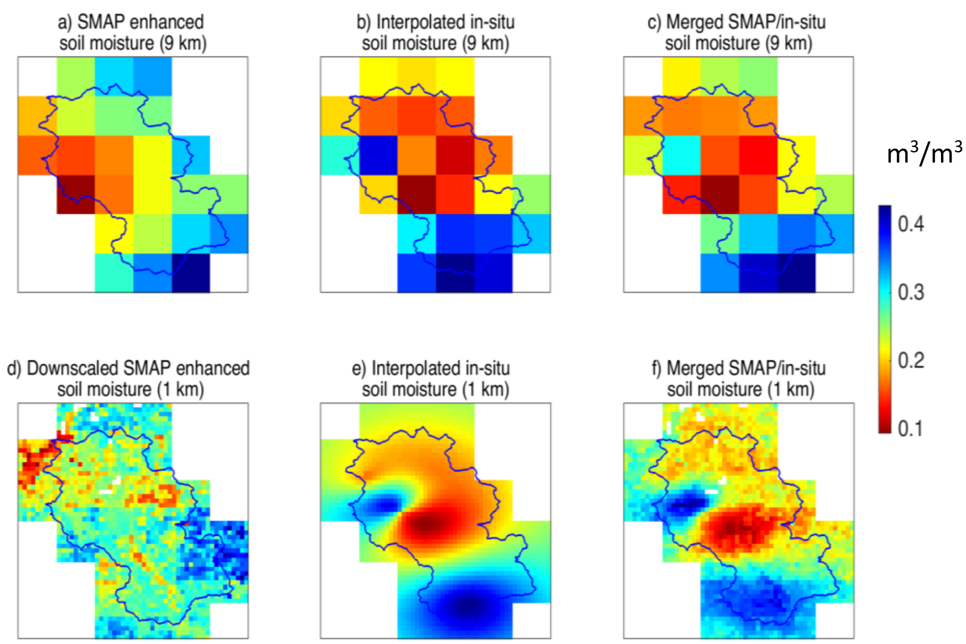

Figure 6.5 Maps of the spatial distribution of a) SMAP enhanced soil moisture, b) interpolated in-situ soil moisture at 9-km, c) merged SMAP/in-situ at 9-km, d) downscaled SMAP enhanced at 1-km, e) interpolated in-situ at 1-km and f) merged SMAP/in-situ soil moisture at 1-km

The interpolated in-situ soil moisture maps over the au Saumon watershed at 9-km and 1-km resolutions are shown in Figure 6.5b and Figure 6.5e, respectively. Despite the small number of soil moisture probes and their uneven distribution in the watershed (see Figure 6.1), these maps somewhat exhibit a similar spatial pattern to that of the SMAP enhanced soil moisture. For example, the maps of both SMAP enhanced and interpolated in-situ measurements display medium range of soil moisture at the top part of the watershed, while the center and lower parts are characterized by lower and higher soil moisture ranges, respectively.

On the other hand, the merged SMAP/in-situ soil moisture reasonably maintained the spatial patterns of SMAP enhanced soil moisture, while keeping the accuracy of the interpolated in-situ soil moisture as can be seen in Figure 6.5c and f for the 9- and 1-km resolution, respectively.

6.4.2 Comparison of in-situ and SMAP soil moisture

Figure 6.6 shows a time series comparison of the 1-km downscaled SMAP enhanced and its original counterpart with the in-situ soil moisture measurements for selected locations in the au Saumon watershed from August to October of 2019. Both the original and downscaled SMAP enhanced soil moisture reasonably well reproduced the temporal dynamics of the in-situ soil moisture during wetter soil moisture condition in late August and October.

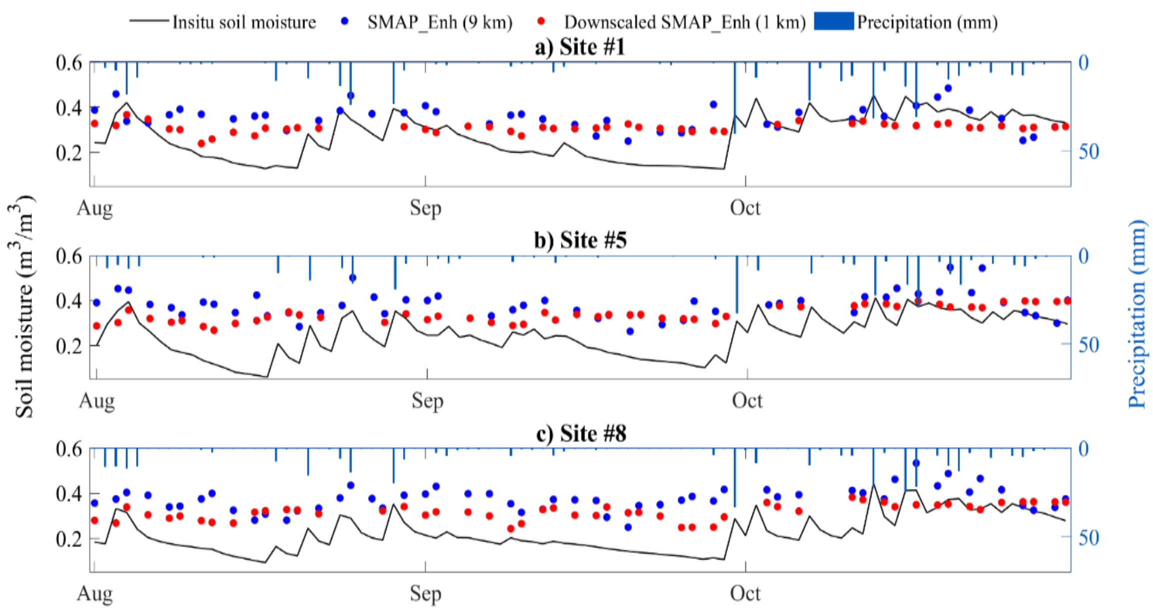

Figure 6.6 Comparison of time series of in-situ, SMAP enhanced and 1-km downscaled SMAP enhanced soil moisture at three selected locations in the au-Saumon watershed (Figure 6.1)

However, during the drier soil moisture conditions, for example mid-August and mid to end of October, both the original and downscaled SMAP overestimated soil moisture. Therefore, it can be inferred from the figure that the SMAP enhanced less reacts to dryness when compared to the in-situ soil moisture. This is because the forest cover in the au Saumon watershed interferes with the passive microwave signal emitted from the underlying soil. However, when compared to its original version, the 1-km downscaled SMAP enhanced soil moisture tends to match better with the in-situ soil moisture.

6.4.3 Updating the model with the SMAP and in-situ soil moisture

Figure 6.7 shows the comparison of the ensemble streamflow forecast with 15-days lead time between the non-updated (open-loop) and updated model with different soil moisture products including the original SMAP enhanced (9-km), downscaled SMAP enhanced (1-km), and in-situ surface soil moisture. Three streamflow events were selected to examine the impact of updating the model with different soil moisture products on the accuracy of the ensemble streamflow forecast. Accordingly, the left panels show the ensemble streamflow forecast issued on 31st of July, whereas the middle and right panels show the forecast on 6th and 14th of August, respectively. Here, the updating was carried out only for the top layer of the model using the direct insertion assimilation technique.

As can be inferred from the figure, the ensemble streamflow forecast is improved by updating the model with the SMAP enhanced (i.e., the original and downscaled) and in-situ surface soil moisture compared to the open loop. The ensemble mean (using EM-Earth product, shown in black) and the deterministic forecast (using MSWEP (precip) and ERA-5 land (temp), shown in red) better agrees with the observed streamflow notably during the first few days of the forecast lead time. On the other hand, the ensemble spread is increased when the model is updated with SMAP enhanced soil moisture, while it decreased when updated with in-situ soil moisture. This is probably because SMAP 'sees' higher soil moisture than in-situ measurements (see Figure 6.6). In all the forecasts, the ensemble generally better encompassed the observed streamflow, yet with the increase of the lead time the ensemble spread becomes wider, as expected, resulting in the deterioration forecast accuracy.

When looking at the impact of the spatial resolution of SMAP enhanced soil moisture on the accuracy of the ensemble streamflow forecast, updating the model with the 1-km downscaled SMAP enhanced soil moisture resulted in a better ensemble streamflow forecast than when the model is updated with its original counterpart (9-km). For example, for all the three forecasts the ensemble members better captured the observed streamflow when the model updated with the 1-km downscaled SMAP ehnanced than its original counterpart. Similarly, the mean of the ensemble members and the deterministic forecast reasonably agree with the observed streamflow for the model updated with the 1-km SMAP enhanced soil moisture product.

Similarly, updating the model with in-situ surface soil moisture improved the ensemble streamflow forecast, yet when compared to the 1-km downscaled SMAP enhanced soil moisture the improvement is less for the forecast on 31st of July and 14th of August 2019. However, for the forecast on 6th of August 2019 the updating with the in-situ surface soil moisture produced better forecast than the 1-km downscaled and original SMAP enhanced soil moisture (9-km).

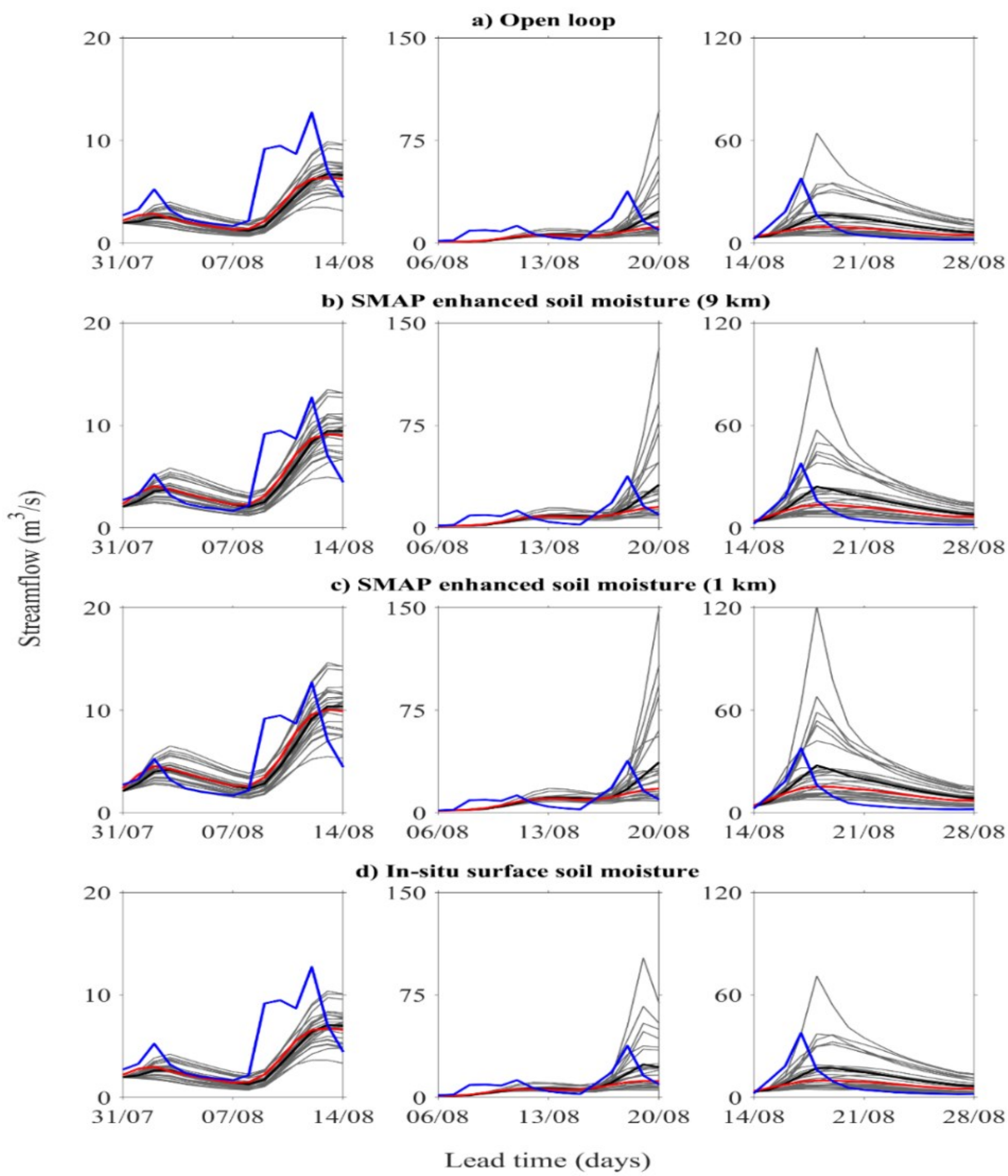

Figure 6.7 Ensemble streamflow forecast at the outlet au Saumon watershed: a) without updating the model (open loop) and after updating with b) SMAP enhanced (9-km), c) downscaled SMAP enhanced (1 -km) and d) in-situ soil moisture

6.4.4 Updating the model with the merged SMAP/in-situ soil moisture

Figure 6.8 shows the comparison of the ensemble streamflow forecast between the model updated with the merged SMAP/in-situ soil moisture at 9- and 1-km spatial resolutions and the open loop. As can be observed from the figure, the ensemble streamflow forecast is improved when the model updated with merged SMAP/in-situ soil moisture both at 9- and 1-km spatial resolutions compared to the open loop. In addition, the ensemble mean and the deterministic forecast closely agree with the observed streamflow notably during the first few days of lead times, yet tends to deteriorate with the increase of the lead time.

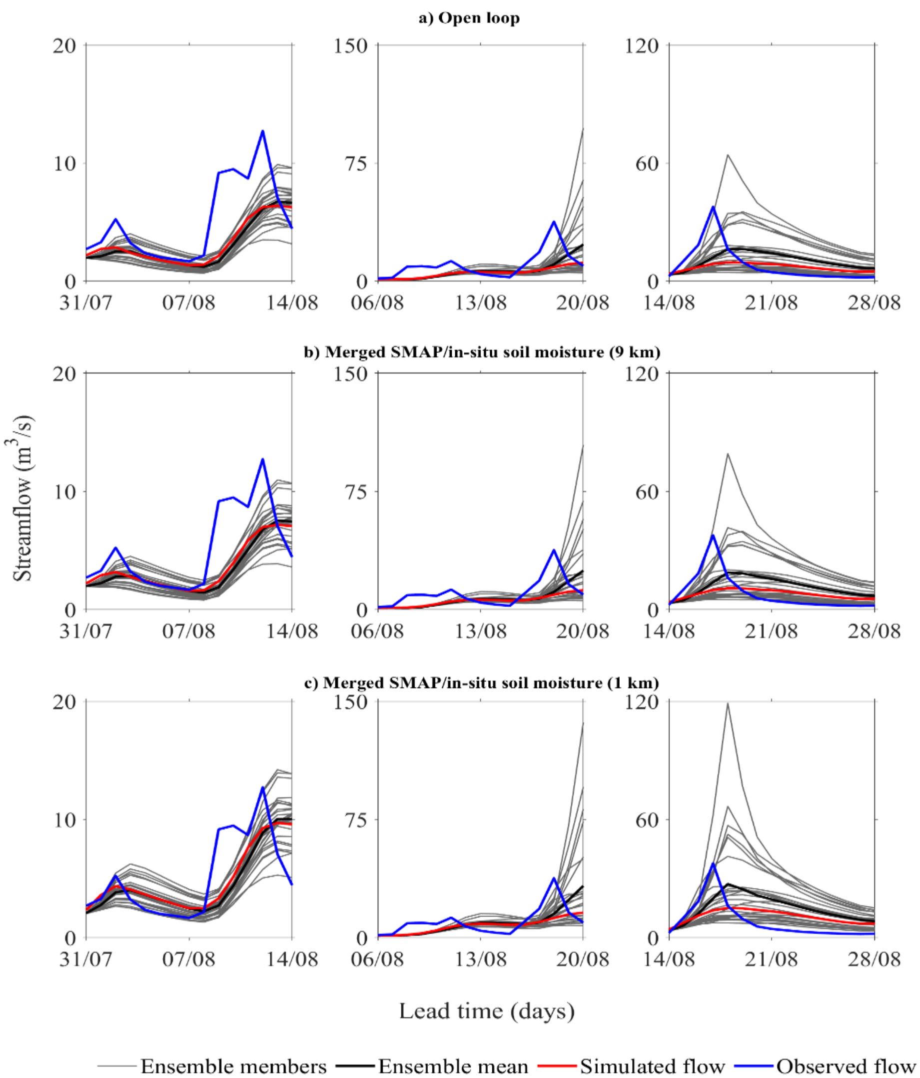

Figure 6.8 Ensemble streamflow forecast for the au Saumon watershed: a) without updating the model (open loop) and after updating with b) the merged SMAP/in-situ (9-km), and c) merged SMAP/in-situ soil moisture (1-km).

The ensemble members captured the observed streamflow better when the model updated with 1-km merged SMAP/in-situ soil moisture than when it was updated with the 9-km merged SMAP/in-situ soil moisture. Likewise, the ensemble mean and the deterministic forecast agreed well with the observed streamflow for the 1-km merged SMAP/in-situ soil moisture. For example, for the forecast on 14th of August the observed streamflow falls outside of the ensemble members when the model updated with the 9-km merged SMAP/in-situ soil moisture, but when the model updated with the 1-km merged SMAP/in-situ soil moisture the ensemble members reasonable shifted towards the observed streamflow.

On the other hand, the 1-km merged SMAP/in-situ soil moisture (Figure 6.8c) improved the ensemble streamflow forecast better than the model updated separately with either the 1-km downscaled SMAP (Figure 6.7c) or in-situ soil moisture (Figure 6.7d). However, the improvement was not that significant. For example, for the forecast on 14th of August the ensemble members are not able to capture the observed streamflow when updating the model with in-situ soil moisture alone, but when updating the model with the 1-km merged SMAP/in-situ soil moisture the ensemble members reasonably shifted towards the observed streamflow.

6.4.5 Updating the model with the vertically extrapolated SMAP soil moisture

Figure 6.9 shows the ensemble streamflow forecast after updating top and intermediate layers of the model with surface and vertically extrapolated SMAP enhanced soil moisture, respectively. In addition, the ensemble streamflow forecast for the model updated with surface and subsurface (rootzone) in-situ soil moisture is shown in Figure 6.9d. Compared to the open loop (Figure 6.9a), updating the top and intermediate layers of the model overall moderately improved the ensemble streamflow forecasts for both the 9- and 1-km resolution SMAP enhanced soil moisture. However, the forecast on 31st of July 2019 overestimated the ensemble streamflow forecast when the model is updated with the vertically extrapolated original SMAP enhanced soil moisture (9-km). On the other hand, the ensemble mean and the deterministic forecast closely match the observed streamflow notably during the first few days of lead time for all forecast. In general, updating the top and second layers increased the forecasted streamflow.

When looking at the impact of spatial resolution, updating the model with the 1-km downscaled SMAP enhanced vertically extrapolated soil moisture produced better ensemble streamflow forecast than when the model update with its coarser counterpart or subsurface in-situ soil moisture. For example, for the forecast on 31st of July the ensemble members were able to capture the observed streamflow when the model updated with the 1-km vertically extrapolated SMAP enhanced soil moisture, yet updating the model with the 9-km vertically extrapolated SMAP enhanced soil moisture overestimated the ensemble streamflow forecast.

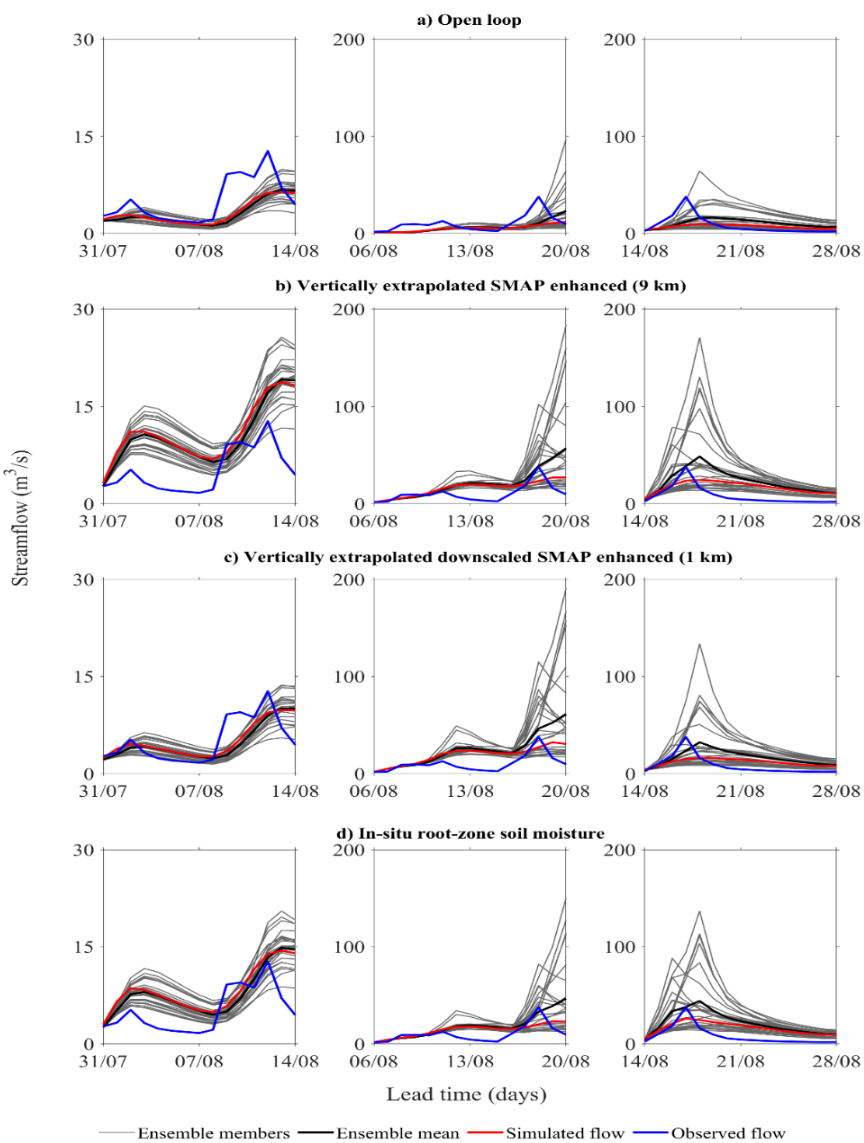

Figure 6.9 Ensemble streamflow forecast for the au Saumon watershed: a) without updating the model (open loop) and after updating with the surface and vertically extrapolated: b) SMAP enhanced (9-km), c) downscaled SMAP enhanced (1-km) and d) in-situ soil moisture

Moreover, updating the top and intermediated layers of model with in-situ surface and subsurface soil moisture (Figure 6.9d), respectively, resulted in overestimation of the ensemble streamflow forecast compared to that of the open loop and updated model with the 1-km vertically extrapolated downscaled SMAP enhanced soil moisture for the forecast on 31st of July 2019.

6.4.6 Updating with the vertically extrapolated SMAP/in-situ soil moisture

Figure 6.10 shows the comparison of ensemble streamflow forecast between the open loop and updated model with the vertically extrapolated merged SMAP/in-situ soil moisture at 9- and 1-km spatial resolutions. Compared to the open loop, updating the model with 9-km vertically extrapolated merged SMAP/in-situ soil moisture slightly improved the streamflow forecast for the forecast on 6th of August 2019, whereas the forecast on 31st of July and 14th of August are overestimated.

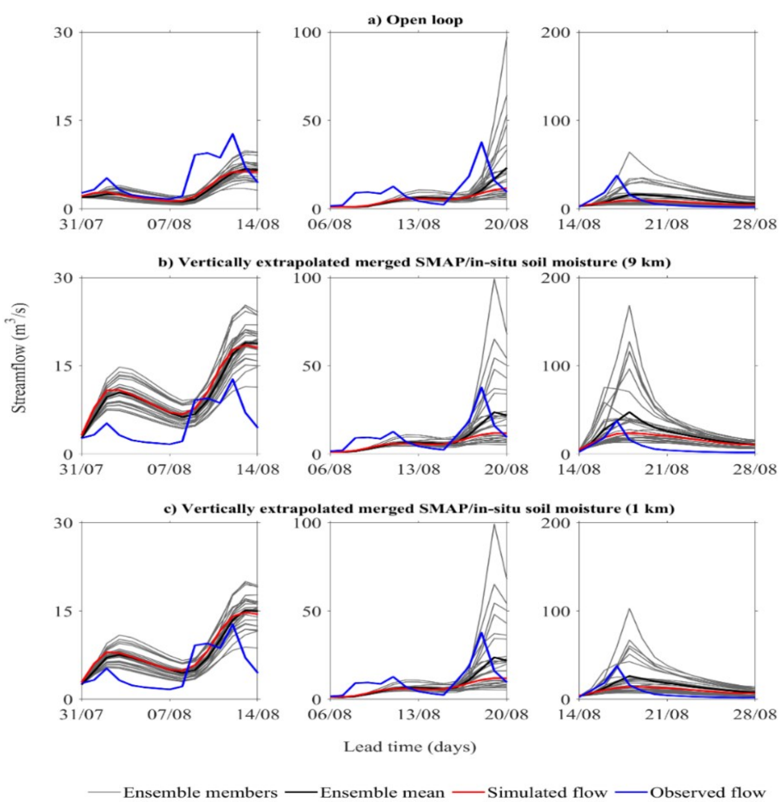

Figure 6.10 Ensemble streamflow forecast at the outlet of au Saumon watershed by the a) open loop and the model updated with the merged SMAP / in-situ vertically extrapolated SMAP / in-situ (9-km) and c) (1-km)

On the other hand, the assimilation of the 1-km vertically extrapolated SMAP/in-situ soil moisture resulted in a better ensemble streamflow forecast for the forecast on 6th and 14th of August, yet the forecast on 31st of July was still overestimated, however to a lesser extent as compared with the assimilation of the 9-km product.

When looking at the impact of spatial resolution, the ensemble streamflow forecast by the model updated with the 1-km vertically extrapolated merged SMAP/in-situ soil moisture generally outperformed than the one updated with the 9-km vertically extrapolated merged SMAP/in-situ soil moisture. For example, for the forecast on 31st of July the model updated with the 9-km vertically extrapolated merged SMAP/in-situ soil moisture considerably overestimated the streamflow forecast than when updated with 1-km vertically extrapolated merged SMAP/in-situ soil moisture.

6.4.7 Comparison between experiments

Besides the graphical comparison, the statistical metrics were used to evaluate the accuracy of the ensemble and deterministic streamflow forecasts. Table 6.1 summarizes the calculated CRPSS values of the ensemble streamflow forecast made by the model updated with different soil moisture products for the three selected events previously. As can be inferred from the table, the CRPSS values are greater than zero for most of the cases indicating that updating the top layer of the model with the SMAP enhanced (i.e., the original and downscaled) and in-situ surface soil moisture improved the ensemble streamflow forecast. However, for a few cases such as the forecast on 6th August updating the model with the 9-km SMAP enhanced soil moisture deteriorated the accuracy of streamflow forecast as indicated by negative CRPSS values.

The merged SMAP/in-situ soil moisture also improved the accuracy of the ensemble streamflow forecast with positive CRPSS values for all the forecasts. The 1-km merged SMAP/in-situ soil moisture further improved the accuracy of the ensemble streamflow forecast than when the model updated either with SMAP enhanced or in-situ soil moisture alone or merge SMAP/in-situ soil moisture at 9-km

On the other hand, updating the model with the vertically extrapolated SMAP enhanced soil moisture (i.e., both the original and downscaled) and subsurface in-situ soil moisture deteriorated the ensemble streamflow forecast as indicated by the negative CRPSS. Similarly, the vertically extrapolated merged SMAP/in-situ soil moisture deteriorated the accuracy of the ensemble streamflow forecast.

Table 6.1 The CRPSS values for different ensemble streamflow forecast

Resolution (km)	SMAP surface			Merged SMAP / in-situ		
	31st of July	6th of August	14th of August	31st of July	6th of August	14th of August
	CRPSS					
9	0,172	-0,04	0,276	0,070	0,030	0,2350
1	0,150	0,09	0,349	0,180	0,120	0,350
In-situ (surface)	0,034	0,230	0,039			
	SMAP vertically extrapolated			Merged SMAP / in-situ vertically extrapolated		
9	-1,250	-0,880	-0,920	-1,200	0,223	-0,880
1	0,145	-1,510	-0,312	1,400	0,780	1,154
In-situ (Root zone)	-0,450	-0,560	-0,743			

Table 6.2 shows the calculated MAE values of the deterministic streamflow forecast by the model updated with different soil moisture products. The MAE was calculated between the observed and the deterministic forecast based on the real observation of meteorological data (MSWEP (precip) and ERA-5 (temp)). Compared to the open loop, updating the model with SMAP enhanced (i.e., both the original and downscaled) and in-situ soil moisture improved streamflow forecast by reducing the MAE, notably for the forecasts on 31st of July and 6th of August 2019. However, the forecast on 14th of August deteriorated.

On the other hand, the merged SMAP/in-situ soil moisture both at 9- and 1-km resulted in a better streamflow forecast compared to when using either SMAP enhanced or in-situ soil moisture independently. For example, for the forecast on 31st of July when the model updated with the 1-km downscaled SMAP enhanced soil moisture the estimated MAE value was 2,01 m^3/s, but it further reduced to 1,90 m^3/s for the 1-km merged SMAP/in-situ soil moisture. Similarly, for the 9-km the MAE reduced from 1,90 to 1,85 m^3/s for the forecast on 31st of July.

On the other hand, updating the model top and intermediate layers with SMAP surface and vertically extrapolated SMAP enhanced soil moisture, respectively, did not bring further improvement. Indeed, the MAE values are increased for all the three forecasts than when updating only the top layer of the model. For example, for the 1-km vertically extrapolated downscaled SMAP enhanced soil moisture the MAE value increased from 2,01 to 2,20 m^3/s. Similarly, updating the model with the vertically extrapolated merged SMAP/in-situ soil moisture did not improve the accuracy of stream forecast.

Table 6.2 The MAE for deterministic forecast

Resolution (km)	SMAP surface			Merged SMAP/in-situ		
	31st of July	6th of August	14th of August	31st of July	6th of August	14th of August
	MAE (m^3/s)					
Open loop	2,10	5,19	5,73			
9	1,90	4,39	6,73	1,85	5,00	6,00
1	2,01	4,31	7,11	1,90	4,28	6,88
In-situ (surface)	1,98	4,52	5,87			
	SMAP surface vertically extrapolated			SMAP-merged vertically extrapolated		
9	5,99	7,77	10,46	5,85	4,56	10,22
1	2,20	9,96	7,50	3,74	4,53	6,87
In-situ (Root zone)	3,84	6,62	9,45			

6.5 Discussion

The Hydrotel model performance is lower for the au Saumon watershed compared to that of the Susquehanna watershed. This mainly because of poor quality of meteorological forcing used in the model. The EM-Earth meteorological data used for the au Saumon watershed is generated from the merging of rain gauge and ERA-5 reanalysis data. However, there are only few rain gauge stations in the au Saumon watershed. In addition, the size and forested nature of the watershed impacts the performance of the model.

The downscaled SMAP enhanced soil moisture (1-km) well reflected the spatial detail of soil moisture over the au Saumon watershed, while maintaining the spatial pattern of the original SMAP enhanced soil moisture (9-km). However, the SMAP enhanced soil moisture (i.e., both the downscaled and original) less reacted to dryness and tends to overestimate soil moisture when compared the in-situ measurements notably during dry condition. This is primarily because of the sub-optimal quality of SMAP soil moisture retrievals over the forested watershed like ours and this is ascribed to the weak penetration of SMAP through a dense vegetation canopy (Entekhabi et al., 2008). Over such an area, both soil surface and vegetation emission contribute to the received signal by the SMAP radiometer and it is complicated to decouple the contribution of both the vegetation and soil surface.

The merging of the SMAP enhanced and in-situ soil moisture resulted in improved maps of soil moisture by maintaining the spatial heterogeneity of SMAP enhanced soil moisture while preserving the dynamic range of in-situ of soil moisture. This agreed with the study by Kim et al. (2016) who merged AMSR2 soil moisture with in-situ measurements over Korean Peninsula.

The updating of the top layer of model with the merged SMAP/in-situ soil moisture improved the accuracy of the ensemble streamflow forecast compared to the open loop. However, the level of improvement varies with resolution of merged SMAP/in-situ soil moisture. Overall, the 1-km merged SMAP/in-situ soil moisture resulted in a better ensemble streamflow forecast than the 9-km merged SMAP/in-situ soil moisture. This because the 1-km resolution better reflects the spatial detail of soil moisture. This is in line with previous studies which showed the importance of higher-resolution satellite soil moisture assimilation for improving the streamflow simulation (Abbaszadeh et al., 2020; Le et al., 2022), yet the assimilation of merged satellite/in-situ soil moisture, to the best of our knowledge, is very rare for comparison to our study.

In another experiment, updating the top layer of the model with in-situ soil moisture alone resulted in less improvement in the accuracy of the ensemble streamflow forecast than the merged SMAP/in-situ soil moisture product for some of the forecasts. For example, the CRPSS value is reduced from 0,230 to 0,120 for the ensemble forecast on 6th of August. On the other hand, the MAE reduced from 1,98 to 1,90 m^3/s and from 4,52 to 4,28 m^3/s for the deterministic forecasts on 31st of July and 6th of August, respectively. This mainly because of a very low density of in-situ soil moisture probes, which cannot sufficiently reflect the spatial variability of soil moisture in the au Saumon watershed. Our finding is not inline with the study conducted on Little Washita river experimental watershed where assimilation of in-situ soil moisture significantly improved the streamflow forecasting (Sun et al., 2016). This mainly because there is a high-spatial density of in-situ soil moisture measurements in the Little Washita Watershed than the au Saumon watershed, which adequately represents the spatial variability of soil moisture in the watershed

Similarly, updating the top layer of the model with SMAP enhanced soil moisture (i.e., either the downscaled or original) alone resulted in less improvement in the accuracy of the ensemble streamflow forecast compared to when updating with the merged SMAP/in-situ soil moisture. This is primarily because of the lower quality of SMAP enhanced soil moisture retrievals over the au Saumon watershed. This watershed is heavily forested, which in turn affected the quality of the SMAP soil moisture retrieval as stated before. Because of that, the SMAP tends to overestimate soil moisture and the assimilation of this wet bias tends to reduce the skill of the ensemble streamflow forecast. This agrees with the study by Abbaszadeh et al. (2020) which reported less accurate model predictions due to the assimilation of overestimated SMAP soil moisture because of the presence of lakes in part of their study area which affected the quality the SMAP soil moisture retrieval.

Additional experiments were also conducted to investigate the impact of updating the intermediate (i.e., the second) layer of the model in addition to the top layer. The top and intermediate layers of the model were updated with the surface and vertically extrapolated merged SMAP/in-situ soil moisture, respectively. Updating both layers of the model deteriorated the accuracy of the ensemble streamflow forecast. Similarly, updating either with the vertically extrapolated SMAP enhanced or subsurface in-situ soil moisture alone did not improve the ensemble streamflow forecast.

This might be partly attributed to the addition or removal of water to the soil when updating the second layer of the model, which is then redistributed by the model subsequently affecting the streamflow generation. In addition, the coupling strength between top and second layers of the model could affect the accuracy of the ensemble streamflow forecast. The coupling between the two layers depends on many factors including vegetation, soil properties and climate conditions (Carranza et al., 2018). For example, the dominance of vegetation in the au Saumon watershed reduces the exposure of the ground surface to atmospheric conditions and expected to result in strong coupling between the surface and subsurface soil moisture. However, the expected strong coupling between the top and second layers of the model did not bring improvement in the accuracy of ensemble streamflow forecasts. This might be due to the subsurface physics of HYDROTEL, which was not explicitly designed to consider the vertical coupling between top and second layers of the model.

Overall, only slight further gain in the accuracy of the ensemble streamflow forecast is obtained when the model is updated with the merged SMAP/in-situ soil moisture i compared to when the model is separately updated either with SMAP enhanced or in-situ soil moisture. This could be attributed to the quality of the merged SMAP/in-situ soil moisture, which in turn depends on the quality of the SMAP enhanced soil moisture and in-situ soil moisture. The spatial interpolation of in-situ soil moisture was affected by the paucity of in-situ probes in the au Saumon watershed, while the quality of the SMAP enhanced soil moisture retrieval was affected by the presence of vegetation as previously discussed. These weakness of SMAP and in-situ soil moisture propagates into the merged SMAP/in-situ soil moisture, thereby affecting the accuracy of the ensemble streamflow when updating the model.

6.6 Conclusion

The L-band passive microwave satellites (e.g., the SMOS and SMAP) and in-situ measurements are established methods for estimation of soil moisture. Over the au Saumon watershed, which is dominated by forests, SMAP overestimated soil moisture and less reacts to dryness, while in-situ measurements are well reacted to dryness producing a better dynamic range of soil moisture. On the other hand, SMAP better reproduces the spatial distribution of soil moisture than in-situ measurements. This is because the in-situ measurements are not adequate to capture the spatial variability of soil moisture as the number of probes over the au Saumon watershed are scarce. This highlights the importance of combining the strength of SMAP and in-situ soil moisture to generate soil moisture with the better quality while compensating for their respective weaknesses. Thus, the conditional merging technique was adopted for this purpose.

The merging of SMAP enhanced soil moisture with the in-situ measurements improved the spatio-temporal representation of soil moisture over the au Saumon watershed than any single one of them by preserving the spatial variability of SMAP and keeping the dynamic range of in-situ soil moisture. The 1-km merged SMAP/in-situ soil moisture better represented the spatial detail of soil moisture than the 9-km merged SMAP/in-situ soil moisture.

The assimilation of the merged SMAP/in-situ surface soil moisture only slightly further improved slightly further improved the accuracy of the ensemble streamflow forecast than when the model separately updated either with the SMAP enhanced or in-situ soil moisture alone. On the other hand, when comparing in terms of spatial resolution, the 1-km merged SMAP/in-situ soil moisture produced reasonably better ensemble streamflow forecast than the 9-km merged SMAP/in-situ soil moisture.

The assimilation of the vertically extrapolated merged SMAP/in-situ soil moisture did not bring further improvement to the accuracy of the ensemble streamflow forecast compared to the open loop. This remains true when the model separately updated with the vertically extrapolated SMAP enhanced and subsurface in-situ soil moisture alone.

Besides its contributions, this study also has some limitations which are worth to mention. First, the au Saumon watershed is heavily forested, which subsequently affects the quality of the SMAP soil moisture retrievals. Second, the number of in-situ soil moisture measurement probes are not adequate to represent the spatial variability of soil moisture in the au Saumon watershed. Hence, the lack of spatially dense in-situ soil moisture measurement stations along with the sub-optimal quality of the SMAP soil moisture affects the quality of the merged SMAP/in-situ soil moisture. This consequently affects the accuracy of ensemble streamflow forecast when assimilated.

In future studies, the merging of different satellite soil moisture products with in-situ soil moisture is encouraged. There are several networks of in-situ soil moisture measurements with different density across the globe. Exploring the impact of the density of these networks on the merging with satellite soil moisture and thereby on the accuracy of streamflow forecasting would be interesting.

Exploring different merging techniques is also a good perspective to consider. In addition, the use of real ensemble meteorological forecast for forcing of hydrological model is encouraged while assimilating the merged satellite/in-situ soil moisture. Finally, exploring more advanced data assimilation schemes along with merged products is also encouraged.

Chapter 7 CONCLUSION AND RECOMMENDATIONS

7.1 Conclusion (Français)

L'humidité joue un rôle fondamental en hydrologie et dans ses applications telles que la simulation et la prévision hydrologiques. Il est intéressant de noter qu'au cours des dernières décennies de plus en plus de données satellitaires en micro-ondes passives onté été disponibles. Toutefois, la faible résolution spatiale de ces données est problématique pour les utiliser en hydrologie que ce soit pour des applications à l'échelle locale ou régionale. Ainsi, cette thèse s'intéresse à réduire l'écart d'échelle exitant entre l'humidité du sol issue de SMAP (ayant une résolution spatiale d'environ 36 km) et les besoins pour les applications hydrologiques (requérant une résolution plus fine, par exemple de 1 km). L'humidité du sol issue de la désagrégation d'échelle des données SMAP est ensuite utilisée pour mettre à jour HYDROTEL, modèle hydrologique distribué, à des fins de simulations et prévisions hydrologiques. Pour atteindre les objectifs de cette thèse, le travail de recherche est divisé en trois parties : 1) la désagrégation d'échelle de l'humidité du sol issue de SMAP, 2) l'assimilation de la donnée d'humidité du sol à fine résolution spatiale dans le modèle distribué à base physique HYDROTEL et 3) l'assimilation dans HYDROTEL d'un produit d'humidité du sol combinant données satellitaires SMAP et données d'observation au sol.

La première partie de cette thèse vise à faire une désagrégation d'échelle des données d'humidité du sol issue de SMAP à des résolutions de 1, 3 et 9 km à l'aide d'un modèle d'apprentissage machine : les forêts aléatoires. La méthode de désagrégation d'échelle développée explore l'utilisation de données de surface issues de Sentinel-1A ainsi que celles issues de MODIS et d'autres sources (SRTM DEM, PRISM par exemples). Les sites d'étude sont les bassins versants des rivières Susquehanna et Au Saumon, respectivement considérés comme des bassins moyennement et fortement forestiers. Ils sont tous deux soumis à des étés chauds et humides.

L'utilisation des forêts aléatoires a révélé de bons résultats pour faire la désagrégation d'échelle spatiale de l'humidité du sol issue de SMAP. En effet, cette désagrégation d'échelle permet d'améliorer la représentation de l'hétérogénéité spatiale de l'humidié du sol tout en conservant les tendances à grandes échelles du produit original. En comparant les produits obtenus à 1, 3 et 9 km de résolution, celui à 1 km de résolution a la meilleure représentation de la variabilité spatiale de l'humidite du sol, notamment en termes de réprésentation des détails de fine échelle. Le produit à 1 km a également révélé avoir une meilleure représentation de la dynamique temporelle de l'humidité du sol, plus proche de celle issue de mesures *in situ*, en particulier pour le bassin versant Susquehanna. En revanche, pour le bassin versant Au Saumon, tous les produits dérivés des données satellitaires SMAP surestiment l'humidité du sol par rapport aux données d'observation *in situ*.

Les prédicteurs issus de Sentinel-1A ne permettent pas une amélioration significative du modèle de forêts aléatoires développé, démontrant ainsi leur moindre importance dans le modèle de désagrégation d'échelle pour les deux bassins versants à l'étude. Par ailleurs, d'autres prédciteurs de d'autres sources jouent un rôle importants ; il s'agit par exemple de la température de brillance issue de la polarisation verticale, le NDVI, l'albédo de surface, l'indice de précipitation antécédente et la topographie qui peuvent être venir de SMAP, MODIS ou SRTM. Le faible apport au modèle de désagrégation d'échelle des prédicteurs de Sentinel-1A peut s'expliquer par l'important couvert forestier des bassins versants étudiés, réduisant alors significativement la sensibilité du signal rétropropagé à l'humidité du sol présente.

De plus, le plus long temps de revisite de Sentinel-1A réduit le nombre de données disponibles pour la phase d'apprentissage du modèle de forêts aléatoires et engendre ainsi une discontinuité temporelle dans le modèle de désagrégation d'échelle (les données Sentinel-1A n'étant disponibles qu'aux douze jours). Cependant, en retirant les données Sentinel-1A, l'échantillon d'apprentissage devient plus long et permet d'avoir un modèle plus robuste. Ce retrait présente également l'avantage d'avoir une résolution temporelle plus fine, avec un écart de trois à cinq jours entre deux données d'humidité du sol après descente d'échelle.

La deuxième partie de cette thèse s'intéresse à la mise à jour de l'humidité du sol dans le modèle distribué à base physique HYDROTEL à l'aide du produit SMAP d'origine (à une résolution de 36 km) et ceux issus de la désagrégation d'échelle (à 1, 3 et 9 km de résolution). Pour mettre la jour l'humidité du sol dans les deux premières couches du sol, la méthode d'insertion directe, considérée comme une méthode d'assimilation de données simple, est utilisée. L'humidité du sol dans la couche de surface est directement mise à jour par la donnée SMAP (d'origine ou après descente d'échelle), alors que l'humidité du sol dans la deuxième couche du sol est mise à jour par une extrapolation verticale de l'humidité du sol en surface de SMAP. Pour cette partie de la recherche, le bassin versant Susquehanna a été retenu ainsi que son son sous-bassin principal : le bassin versant de la rivière Upper Susquehanna, un bassin de faible superficie.

La mise à jour de l'humidité du sol dans la couche de surface du modèle avec l'humidité du sol de SMAP (que ce soit le produit d'origine ou après désagrégation d'échelle) améliore la qualité du modèle hydrologique d'ensemble comparé à un modèle sans mise à jour (soit en boucle ouverte), et ce pour les deux bassins versants considérés dans cette partie de l'étude. Cependant, l'amélioration des simulations n'est pas la même selon le produit d'humidité du sol utilisé pour la mise à jour du modèle. Par exemple, pour le bassin versant Susquehanna, le produit d'humidité du sol à 9 km de résolution améliore significativement le modèle hydrologique d'ensemble sans qu'il soit nécessaire d'utiliser un produit à plus fine échelle. Autrement dit, la mise à jour du modèle hydrologique avec des données d'humidité du sol à très fine échelle (à 1 km de résolution en particulier) ne permet pas de relever une quelconque amélioration significative du modèle hydrologique. Pour le bassin de la Upper Susquehanna, la mise à jour de l'humidité du sol avec les produits à 1 et 3 km de résolution permet une meilleure amélioration du modèle hydrologique qu'avec le produit SMAP d'origine ou avec une résolution plus grossière (9 km).

Ce qui précède montre que la taille du bassin versant permet de déterminer la résolution spatiale des données d'humidité du sol à utiliser pour faire la mise à jour du modèle hydrologique. Pour le bassin versant Susquehanna qui est le plus grand, la désagrégation d'échelle à une résolution plus large (9 km) est suffisante pour décrire de manière satisfaisante l'hétérogénéité spatiale de l'humidité du sol et utiliser une résolution beaucoup plus fine (1 km) n'apporte aucune amélioration significative sur la qualité des prévisions hydrologiques d'ensemble. Cependant, pour le bassin versant Upper Susquehanna, qui est un petit bassin versant, l'humidité du sol issue du produit SMAP d'origine à une résolution de 36 km ou de la désagrégation d'échelle la plus large (9 km) ne permet plus de représenter adéquatement l'hétérogénéité spatiale de l'humidité du sol car, avec ces résolutions, seul un faible nombre de pixels suffit pour recouvrir tout le bassin versant. C'est pourquoi il est nécessaire d'utiliser les produits issus d'une désagrégation d'échelle à plus fine résolution (1 et 3 km) permettant ainsi de mieux représenter l'hétérogénité de l'humidité du sol et par suite de fournir des modélisations hdyrologiques de meilleure qualité.

En revanche, la mise à jour des deux premières couches de sol du modèle hydologique par respectivement les données SMAP d'humidité du sol en surface et celles extrapolées en profondeur ne permet pas d'améliorer significativement le modèle hydrologique d'ensemble. Il s'avère que la mise à jour de l'humidité du sol en surface amène de meilleurs résultats. Cela peut en partie s'expliquer par le manque de complexité dans le modèle d'extrapolation verticale utilisé, à la physique des couches profondes du modèle, aux caractéristiques physiographiques du bassun versant (type de sol et occupation du sol notamment), ainsi qu'à un manque de données d'humidité du sol *in situ* pour valider correctement le modèle d'extrapolation développé.

Par rapport à une mise à jour du modèle pour chaque donnée SMAP disponible, la mise à jour du modèle à partir de l'occurrence d'événements de pluie intense permet de réduire la fréquence de mise à jour ainsi que la puissance de calcul nécessaire. Il est intéressant de noter que cette baisse du nombre de mise à jour du modèle améliore la qualité du modèle hydrologique. Cela permet en effet au modèle de mieux représenter le cheminement de l'eau dans le sol entre deux événements précipitants, en particulier le cheminement de l'eau de surface jusque dans les couches profondes.

En plus de ces avantages venant de la mise à jour du modèle hydrologiques à partir de différentes résolutions spatiales de l'humidité du sol issue de SMAP, il est important de souligner certaines limites qui y sont associées. Il s'agit par exemple de la faible qualité relative des données d'humidité du sol de SMAP sur les régions du bassin versant couvertes de forêts, du modèle d'extrapolation verticale utilisé et du manque de données d'humidité du sol *in situ* pour développer un modèle d'extrapolation verticale plus robuste. Cette thèse ne sait pas pencher sur ce problème, qui reste un enjeu majeur en prévision oppérationnelle.

Dans la dernière partie de cette thèse, l'humidité du sol issue de SMAP et celle issue de données *in situ* sont combinées pour fournir une carte d'humidité du sol unique (humidité du sol SMAP/*in situ*) de meilleure qualité qui prend le meilleur des deux sources. Les données satellitaires permettent d'avoir accès à une bonne représentation de l'hétérogénéité spatiale générale sur le territoire, tandis que les données au sol donnent une bonne représentation de la dynamique temporelle. Le produit combiné est utilisé pour mettre à jour l'humidité du sol du modèle HYDROTEL dans le but de faire de la prévision hydrologique. Cette méthode est appliquée sur le bassin versant Au Saumon.

La combinaison des données d'humidité du sol SMAP et *in situ* améliore la représentation spatio-temporelle de l'humidité du sol et fait mieux que les deux produits pris séparemment. Ainsi, la mise à jour de l'humidité du sol du modèle hydrologique avec ce produit combiné améliore la qualité des prévisions d'ensemble du modèle hydrologique comparé à un modèle sans mise à jour de l'humidité du sol. Toutefois, cette amélioration n'est pas significative par rapport à la mise à jour par un seul produit d'humidité du sol, que soit un des produits de descente d'échelle de l'humidité du sol de SMAP ou l'humidité du sol *in situ*. Cela vient surtout de la faible qualité des données d'humidité du sol de SMAP sur le bassin versant Au Saumon qui est très boisé. De plus, le faible nombre de mesures *in situ* affecte le produit combiné.

L'utilisation du produit combiné humitidé du sol SMAP désagrégée à 1 km de résolution et humidité du sol *in situ* permet d'obtenir de meilleures prévisions hydrologiques d'ensemble que lorsque l'humidité du sol issue de SMAP est désagrégée à une résolution de 9 km seulement. D'autre part, la mise à jour du modèle avec l'humidité du sol verticalement extrapolée par le produit combiné SMAP/*in situ* n'apporte pas d'amélioration supplémentaire par rapport à la mise à jour de la seule couche d'humidité du sol de surface du modèle hydrologique. Comme déjà mentionné, ceci pourrait venir de la qualité des données SMAP et du manque de données d'observation au sol.

De manière générale, cette recherche utilise l'humidité du sol mesurée par satellite en micro-onde passive bande L (SMAP) dans l'objetif d'améliorer la modélisation hydrologique. L'accès à une technologie de ce type offre l'accès à des données représentant la variabilité spatiale et temporelle de l'état du bassin versant et en particulier de l'humidité du sol. Tradionellement, les mesures *in situ* sont utilisées. Mais le faible nombre de sites d'observation est problématique et requiert plus de ressources pour avoir une couverture optimale sur de grands territoires (à l'image des bassins versants de cette thèse). Ainsi, les avancées technologiques en télédétection permettent d'approfondir la compréhension des processus hydrologiques et donc améliore les applications qui en sont faites en hydrologie et agriculture. Cependant, il reste encore de nombreuses impasses à surmonter pour utiliser pleinement les ressources offertes par la télédétection dans les applications en hydrologie.

Il est important de continuer à améliorer les modèles hydrologiques pour pouvoir faire face au mieux aux futurs événements qui s'éloignent de l'historique actuel. Il n'est très certainement pas requis de redéfinir l'ensemble de la modélisation hydrologique; mais l'effort devrait se faire sur les modules qui modélisent au mieux les changements à venir. De plus, certains modules ne permettent actuellement pas d'utiliser des données de télédétection comme intrants. L'augmentation de la disponibilité de données de télédétection impose une mise à jour de ces modules. En s'appuyant sur les travaux de cette thèse, il est par exemple recommendé de redéfinir le module d'humidité du sol afin de mieux prendre en compte la mise à jour de l'humidité du sol issue de données satellitaires.

7.2 Recommendations (Français)

De plus en plus de données d'humidité du sol issues de mesures satellitaires en micro-onde passive sont disponibles. C'est pourquoi ces dernières décennies ont vu émerger plusieurs produits d'humidité du sol développés et validés afin d'avoir de meilleures méthodes d'assimilation de l'humidité du sol satellitaires dans les modèles de surface. Cependant, l'étude de ces produits satellitaires pour des applications hydrométéorologiques et agricoles variées est encore jeune et c'est dans ce contexte que cette thèse a été menée. Elle a visé à explorer l'utilisation de l'humidité du sol issue de SMAP pour faire de la modélisation et de la prévision hydrologique. Ainsi, à partir des conclusions de cette thèse, plusieurs recommendations sont proposées pour des recherches futures.

Premièrement, il est très important d'améliorer la qualité de la mesure de l'humidité du sol par SMAP dans les zones modérément ou fortement boisées. En effet, avec ce type d'occupation du sol, SMAP a tendance à surestimer l'humidité du sol et l'utilisation de ces données biaisées ont un effet défavorable sur le modèle de prévision hydrologique. Cette thèse l'a démontré, en particulier sur le bassin versant Au Saumon. Le territoire de ce bassin versant est très majoritairement couvert de forêts. Sur ce type de surface, il est difficile de faire la différence entre le signal émis par le sol et celui provenant de la canopée. Ainsi, au-delà des avancées technologiques en micro-ondes passives, il faut aussi encourager les études qui portent sur l'amélioration des algorithmes déjà existants de traitement de ce type de signal.

Deuxièmement, dans cette thèse, les prédicteurs issus des mesures de radar à synthèse d'ouverture (SAR) n'ont pas apporté une amélioration significative dans la qualité du modèle de désagrégation d'échelle de l'humidité du sol de SMAP. Cela vient en particulier du fait que les bassins versants à l'étude étaient principalement forestiers. Cependant, il est admis que l'utilisation de ces données SAR pour la désagrégation d'échelle de l'humidité du sol issue des micro-ondes passives de SMAP et SMOS est prometteuse pour les sols nus ou couverts de végétation basse. Ainsi, les études futures devraient fortement utiliser les données déjà existantes et à venir des satellites SAR tels que Sentinel-1A/B, la constellation RADARSAT et la bande L de NISAR. L'avantage de ces mesures est qu'elles peuvent être effectuées sans égard des conditions météorologiques et peuvent pénétrer dans la végétation plus profondément que les satellites optiques ou infra-rouges comme MODIS. La mesure par bande L des SAR comme celui de NISAR est la plus prometteuse. Par ailleurs, leur plus fine résolution spatiale offre l'opportunité de faire une désagrégation d'échelle encore plus fine (de l'ordre de plusieurs mètres) de l'humidité du sol issue des micro-ondes passives que celle faite dans cette thèse. Cela bénéficierait sans aucun doute à de nombreuses applications hydrologiques.

Troisièmement, la plupart des méthodes de désagrégation d'échelle ont été développées et appliquées pour les milieurs arides et semi-arides, pour lesquels l'effet de l'occupation du sol dans le modèle de désagrégation d'échelle n'est que peu important. Cette thèse a essayé d'utiliser la méthode des forêts aléatoires pour faire une désagrégation d'échelle de l'humidité du sol de SMAP pour des bassins versants ayant une couverture forestière de modérée jusqu'à forte. Des résultats prometteurs ont été obtenus, mais des études futures sont nécessaires sur des régions ayant les mêmes caractéristiques d'occupation du sol. De plus, le modèle d'ensemble des forêts aléatoires utilisé pour faire la désagrégation d'échelle a tendance a sousestimer les fortes valeurs d'humidité du sol et surestimer les faibles. Ainsi, d'autres recherches sont nécessaires pour améliorer les méthodes d'apprentissage machine utilisées pour la désagrégation d'échelle.

Quatrièmement, cette thèse a fait appel à une méthode d'insertion directe comme technique d'assimilation de données. Celle-ci est la plus simple parmi l'ensemble des techniques d'assimilation de données existantes. Elle fut retenue dans l'objectif de se concentrer sur une évaluation de la capacité d'utilisation des données SMAP pour améliorer la modélisation et la prévision hydrologique. Les études à venir pourraient s'employer à utiliser, pour le même travail que dans cette thèse, des techniques d'assimilation de données plus avancées telles que le filtre particulaire, le filtre de Kalman d'ensemble ou le filtre de Kalman d'ensemble séquentiel.

Cinquièmement, bien que récente, l'assimilation de données continue à gagner en popularité en hydrologie. Cependant, bien que cette thèse et d'autres études montrent qu'il est possible de mettre en place une méthode d'assimilation de données dans un modèle hydrologique, la base physique de la plupart des modèles hydrologiques existants n'a pas été pensée pour l'assimilation de données. Il serait donc important d'améliorer cette base physique ou de développer de nouveaxu modèles orientés vers l'assimilation de données. Par exemple, pour l'assimilation des données d'humidité du sol, la physique des modules sous-surfaces doit être améliorée afin de clairement représenter la connexion entre les différentes couches de sol à partir des propriétés du sol et de l'occupation du sol.

Sixièmement, il est recommendé de faire une extrapolation horizontale de l'humidité du sol de SMAP des zones où elle est de bonne qualité vers les zones à forte végétation. De même, des méthodes robustes et à base physique doivent être développées pour l'extrapolation verticale de l'humidité du sol issue de mesures satellitaires à micro-ondes passives, car les méthodes existantes manquent d'une base physique solide.

Septièmement, cette thèse a scruté le potentiel de l'humidité du sol en modélisation et prévision hydrologique avec un contrôle des données météorologiques extraites d'un historique. La suite de ce travail serait de l'appliquer pour la prévision hydrologique oppérationnele en utilisant de vraies données de prévision météorologique.

Enfin, cette thèse a proposé une fusion de l'humidité du sol issue de SMAP avec des observations *in situ* à l'aide de la méthode de fusion conditionnelle, une technique couramment utilisée pour la combinaison des données de précipitations satellitaires et de pluviomètres. Plusieurs études ont proposé des méthodes pour combiner l'humidité du sol issue de différents satellites. Un exemple d'intérêt est l'ECV_SM (Variables Climatiques Essentielles - Humidité du Sol, *Essential Climate Variables – Soil Moisture*) développé afin de combiné quatre mesures en micro-ondes passives (SMMR, SMM/I, TMI et AMSR-E) et deux en micro-ondes actives (ERS AMI et ASCAR) à large résolution spatiale (Liu et al., 2011; Tomer et al., 2016). Cependant, peu d'études portent sur la combinaison d'humidité du sol issue de mesures satellitaires et de meurues *in situ* et il serait donc utile d'en voir plus.

7.3 Conclusion (English)

Soil moisture plays a key role in hydrological applications such as streamflow simulation and forecasting. Interestingly, its availability from passive microwave remote sensing is increasing over the last decades, yet its low spatial resolution hampers its use for the local and regional scale hydrological applications. Hence, this thesis is designed to reduce the scale disparity between the SMAP soil moisture (roughly 36-km spatial resolution) and the hydrological application requirements (e.g., 1-km resolution) through downscaling. The downscaled SMAP soil moisture is then used for updating a distributed hydrological model, namely HYDROTEL, for the purpose of streamflow simulation and forecasting. Accordingly, this thesis splits into three major consecutive parts: 1) spatial downscaling of the SMAP soil moisture, 2) assimilation of downscaled SMAP into a physically-based distributed hydrological model (i.e., HYDROTEL), and 3) merging SMAP and in-situ soil moisture, and assimilating into HYDROTEL.

In the first part of the thesis the SMAP soil moisture was downscaled to a range of spatial resolutions (i.e., 1-, 3- and 9-km) using random forest-based ML technique. During downscaling, the value of Sentinel-1A derived land surface variables along with those derived from MODIS and other sources (e.g., SRTM DEM, PRISM) were explored for the downscaling of the SMAP soil moisture products over moderately (Susquehanna) to heavily forested (au Saumon) watersheds with warm-summer humid continental climate.

The random forest ML technique demonstrated good capability in improving the spatial resolution of the SMAP soil moisture. Indeed, the downscaling improved spatial variability of the SMAP soil moisture while preserving the spatial pattern of the original SMAP soil moisture. When comparing in terms of the representation of spatial variability, the 1-km downscaled SMAP soil moisture better represented the spatial detail of soil moisture than the 3-km or 9-km downscaled SMAP soil moisture. Similarly, the 1-km downscaled SMAP soil moisture better captured the temporal dynamics of in-situ soil moisture, especially for the Susquehanna watershed. On the other hand, for the au Saumon watershed the SMAP products at various resolution tend to overestimate soil moisture compared to the in-situ soil moisture.

With regards to predictors, the use of predictors derived from Sentinel-1A has only slightly improved the performance of the random forest model, indicating their less importance in the downscaling of the SMAP soil moisture over the Susquehanna and au Saumon watersheds. On the other hand, predictors derived from other sources (e.g., SMAP, MODIS and SRTM), such as vertically polarized brightness temperature, NDVI, surface albedo, antecedent precipitation index and elevation were identified as important predictors. The less importance of the Sentinel-1A predictors might be because of the masking effect of vegetation, which reduces the sensitivity of backscatter signal to soil moisture as the Susquehanna and au Saumon watersheds are dominated by forest.

In addition, the longer revisit time of the Sentinel-1A reduced the data size available for training of random forest and resulted in temporally discontinuities downscaled SMAP soil moisture (i.e., available every 12 days). However, excluding the Sentinel-1 predictors from the list of predictors increased the training size of the data thereby resulting in a more robust model. In addition, the temporal continuity of downscaled soil moisture is increased (i.e., available every 3 to 5 days).

In the second part of the thesis the original (36-km) and resulting downscaled SMAP soil moisture (e.g., at 1-, 3- and 9-km) products were used to update a physically-based distributed hydrological model HYDROTEL. Direct insertion, which is a simple data assimilation technique, was used to update the top and second layers of the model. The top layer was directly updated with the SMAP soil moisture, while the second layer was updated with the vertically extrapolated SMAP soil moisture. We selected the Susquehanna watershed and one of its head sub-watersheds, the Upper Susquehanna watershed (relatively smaller in size) for the experiments.

Updating the top layer of the model with the original and downscaled SMAP soil moisture at all resolutions improved the ensemble streamflow simulations compared to the non-updated model (i.e., open loop) for both watersheds. However, the degree of improvement varies with the spatial resolution used to update the model. For example, for the Susquehanna watershed, the 9-km downscaled SMAP soil moisture significantly improved the accuracy of the ensemble streamflow simulations without the need to go to higher-spatial resolutions. Putting it differently, when the model is updated with the downscaled SMAP soil moisture at higher-spatial resolutions, e.g., 1-km, no further significant improvement was noticed. On the other hand, for the Upper Susquehanna watershed, the downscaled SMAP at higher-spatial resolutions, e.g., at 1- and 3-km, produced better ensemble streamflow simulations than the original and downscaled SMAP at coarser spatial resolution (i.e., 9-km).

From the above insights, it is inferred that that the size of the watershed determines the spatial resolution of SMAP soil moisture suitable for updating of the model. For the larger watershed (i.e., the Susquehanna watershed), the downscaled SMAP at coarser resolution (i.e., 9-km) adequately described the spatial heterogeneity of soil moisture and going to higher-resolution (e.g., 1-km) did not bring significant gain on the accuracy of the ensemble streamflow simulations. However, for the smaller watershed (i.e., the Upper Susquehanna watershed) the coarser resolution downscaled SMAP (9-km) or the original SMAP with resolution of 36-km could not sufficiently describe the spatial heterogeneity of soil moisture as only few pixels cover the entire watershed. For this reason, the downscaled SMAP at higher-spatial resolutions (1- and 3-km) better reflected the spatial heterogeneity of soil moisture and resulted in better streamflow simulation.

On the other hand, updating the top and second layers of the model with surface and vertically extrapolated SMAP soil moisture, respectively, only slightly further improved the accuracy of the ensemble streamflow simulations compared to when only the top layer of the model is updated. This could be partly attributed to the simplicity of the assumption adopted for developing the vertical extrapolation method, the model subsurface physics, physiographic characteristics of the watershed (e.g., soil type and land cover) and lack of the in-situ soil moisture data for validation of the adopted extrapolation approach.

Updating the model based on high rainfall events reduced the frequency of updating as well as the computational time needed compared to when the model updated based on the availability of the SMAP soil moisture with the full watershed coverage. Interestingly, the reduction in the frequency of updating improved the accuracy of the ensemble streamflow simulations. This is because reducing the frequency of updating offers more time for soil moisture to propagate to the rootzone between two subsequent updating times.

Apart from the benefits obtained from updating the model with the SMAP soil moisture with a range of spatial resolutions, it is worthwhile to mention some of the limitations, such as the sub-optimal quality of the SMAP soil moisture retrieval notably over forested part of the watershed, the assumption made for vertical extrapolation of the SMAP soil moisture and paucity of in-situ soil moisture for implementation of a more robust vertical extrapolation method. Our study did not investigate this issue, which should be addressed particularly in the context of operational streamflow forecasting.

In the last part of the thesis, the soil moisture derived from the SMAP and in-situ measurements are merged to generated improved soil moisture maps (SMAP/in-situ soil moisture) by combining their respective strengths. SMAP has the advantage of providing the spatial heterogeneity of soil moisture over a large area, whereas in-situ measurements have the advantage of preserving the temporal dynamics of in-situ soil moisture. The merged product (i.e., SMAP/in-situ soil moisture) was then used to update HYDROTEL for the purpose of streamflow forecasting. This study was implemented in the au Saumon watershed.

Merging of the SMAP and in-situ soil moisture improved the spatio-temporal representation of soil moisture compared to any single one of them. In this context, updating the model with the merged product (i.e., SMAP/in-situ soil moisture) improved the accuracy of the ensemble streamflow forecasts compared to the non-updated model. However, the improvement was not that significant compared to when the model is updated with either downscaled SMAP or in-situ soil moisture alone. This is primarily because of the sub-optimal quality of the SMAP soil moisture retrieval over the au Saumon watershed which is heavily forested. Besides that, the sparsity of in-situ soil moisture measurements in the au Saumon watershed affects the merging of the SMAP and in-situ soil moisture.

With regards to the benefit of spatial resolutions, updating the model with the 1-km SMAP/in-situ soil moisture resulted in better ensemble streamflow forecasts than the 9-km SMAP/in-situ soil moisture. On the other hand, updating the model with the vertical extrapolated SMAP/in-situ soil moisture did not bring further improvement compared to updating only the top layer of the model. This could be attributed to the quality of SMAP, and paucity of in-situ soil moisture as stated previously.

Generally, this study used soil moisture measured based on L-band satellite passive microwave remote sensing technologies (e.g., SMAP) for the purpose of improving hydrological simulations. Such technologies help to measure the spatial and temporal variability of watershed state, mainly soil moisture. Traditionally, in-situ measurements are used, but they are scarce, time consuming and expensive to install over a large area (e.g., watershed). Thus, advancement of remote sensing technologies greatly helps the improvement of our understanding of hydrological processes and also for hydrological and agricultural applications. However, still there is long way to go in hydrology to use remotely sensed information for hydrological applications.

Frequent refining of hydrological models is important to be able to extrapolate to changing conditions. It may not be needed to refine the entire model, but some of the modules that simulate changing hydrological processes could be refined. In addition, some of the modules of hydrological models were not designed for incorporation of remotely sensed information. Thus, with the increase of availability of remotely sensed information refining some of the modules of the models are important. For example, in the case of this thesis it is recommended to refine the soil moisture module of the model to better accommodate the assimilation of satellite soil moisture.

7.4 Recommendations (English)

The availability of satellite passive microwave soil moisture products is increasing. Hence, over the last decades the validation and generation of optimal soil moisture products through assimilation of satellite soil moisture into land surface model have been active research areas. However, the study on their use for different hydrometeorological and agricultural applications is at its early stage and that is why this thesis focused on exploring the value of the SMAP soil moisture for the purpose of streamflow simulation and forecasting. Thus, from the perspective of this thesis the following research areas are recommended.

First, improving the quality of the SMAP soil moisture retrieval over moderately to heavily forested areas is very important. This is because over such land surface conditions the SMAP tends to overestimate soil moisture and the use of such a wet bias would affect the model predictions. This is what is observed in this thesis particularly for the au Saumon watershed, which is heavily forested. Forest covers a considerable portion of the land surface. Over such an area it is difficult to discriminate the emission from soil surface and vegetation canopy. Thus, besides the upcoming advanced satellite passive remote sensing technologies, studies which focus on improving the existing retrieval algorithms are encouraged to address this issue over such an area.

Second, in this thesis, the SAR derived predictors did not bring significant improvement in the accuracy of the downscaled SMAP soil moisture mainly because of the forested nature of the study watersheds. However, it is deemed that the use of SAR data for the downscaling of satellite passive microwave soil moisture products including the SMAP and SMOS is promising over bare to low vegetated areas. Hence, future studies should capitalize on the use of existing and upcoming SAR satellites such as Sentinel-1A/B, the Canadian RADARSAT constellation mission (RCM) and the L-band NASA-ISRO Synthetic Aperture Radar (NISAR). This is because SAR satellites have the potential to operate under all-weather conditions and penetrate through vegetation to some extent than the optical\thermal based satellites such as MODIS, notably the upcoming L-band SAR such as NISAR is the most promising one. Besides, their higher-spatial resolution allows to downscale the satellite passive microwave soil moisture even to higher-spatial resolutions (i.e., in the order of meters) which is crucial for local scale hydrological applications.

Third, most of the downscaling techniques were developed and applied in the arid and semi-arid climate setting where the effect of land cover on the downscaling is minimal. This thesis attempted to use random forest ML technique to downscale the SMAP soil moisture for moderate to heavily forested watersheds. Encouraging results were obtained, yet further studies are needed for similar land and climate setting. In addition, downscaling based on the ensemble ML such as random forest inherently tends to under- and overestimate higher and lower soil moisture values, respectively. Thus, further studies are needed for improving ML based downscaling techniques.

Fourth, in this thesis the direct insertion data assimilation technique was used. This technique is the simplest of all data assimilation techniques and adopted with the objective to evaluate the value of satellite passive microwave soil moisture (SMAP) in improving streamflow simulation and forecasting. In future studies, one can expand on this work by using more advanced data assimilation techniques such as Particle Filter, Ensemble Kalman Filter and sequential Ensemble Kalman Filter.

Fifth, data assimilation is rather recent but continues to gain popularity in hydrology. However, the physics of most of the existing hydrological models were not designed for data assimilation, even though they have been proved effective in this thesis and many other studies. However, it would be important to improve the physics of the existing hydrological models or develop new hydrological models for the purpose of data assimilation. For soil moisture assimilation, for example, the model subsurface physics needs to be improved to clearly represent the connection between the soil layers based on the soil properties and land covers.

Sixth, it is also advisable to spatially extrapolate the SMAP soil moisture from part of the watershed where it has a good quality to the part where it has a low quality. For example, in heavily vegetated areas the SMAP soil moisture is poor for the Susquehanna watershed, but in open and agricultural areas the quality is good. Similarly, robust and physically-based method need to be developed for vertical extrapolation of satellite passive microwave soil moisture as the exponential filter used in this study lack a solid physical basis.

Seventh, this thesis scrutinized the value of soil moisture in hindcasting mode for simulation of streamflow. Extending this work for operational streamflow forecasts would be encouraged by using the real forecasted meteorological data.

Finally, this thesis attempted to merge the SMAP soil moisture with the in-situ observations using conditional merging technique, a technique often used for merging satellite precipitation with rain-gauge observations. There are many studies that have been carried out to merge soil moisture derived from different satellites. On good example is ECV_SM (Essential Climate Variables (ECV_SM)) developed by merging four passive (SMMR, SSM/I, TMI, and AMSR-E) and two active (ERS AMI and ASCAT) coarse resolution microwave_sensors (Liu et al., 2011; Tomer et al., 2016). However, the study on the merging of satellite and in-situ soil moisture is rather rare, which suggests further studies.

Appendix

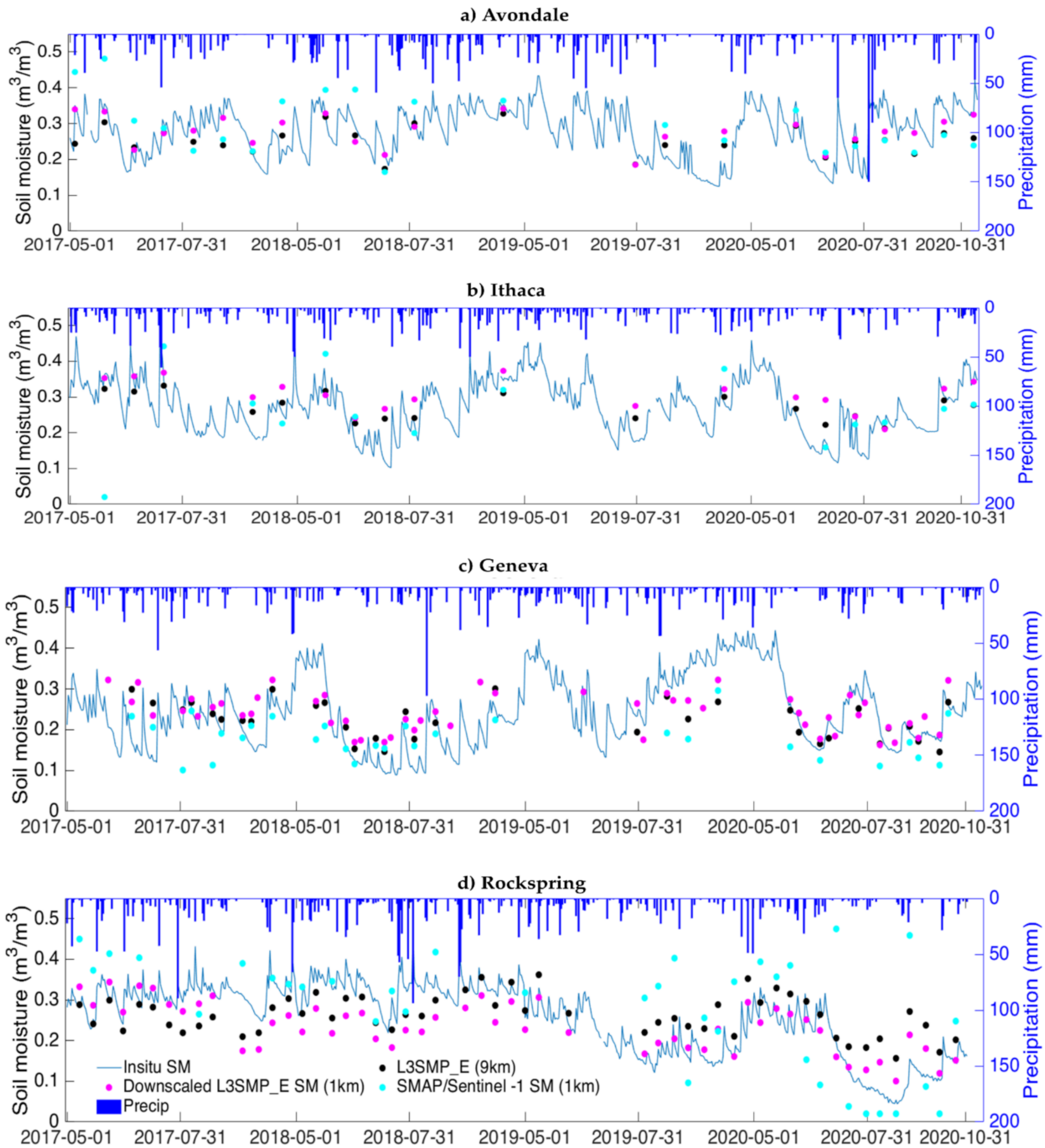

Figure A.0.1 Time series of downscaled L3S L3SMP_E SM, L3SMP_E, L2_SM_SP, daily rainfall and in situ SM for (a) Avondale, (b) Ithaca, (c) Geneva and (d) Rockspring sites

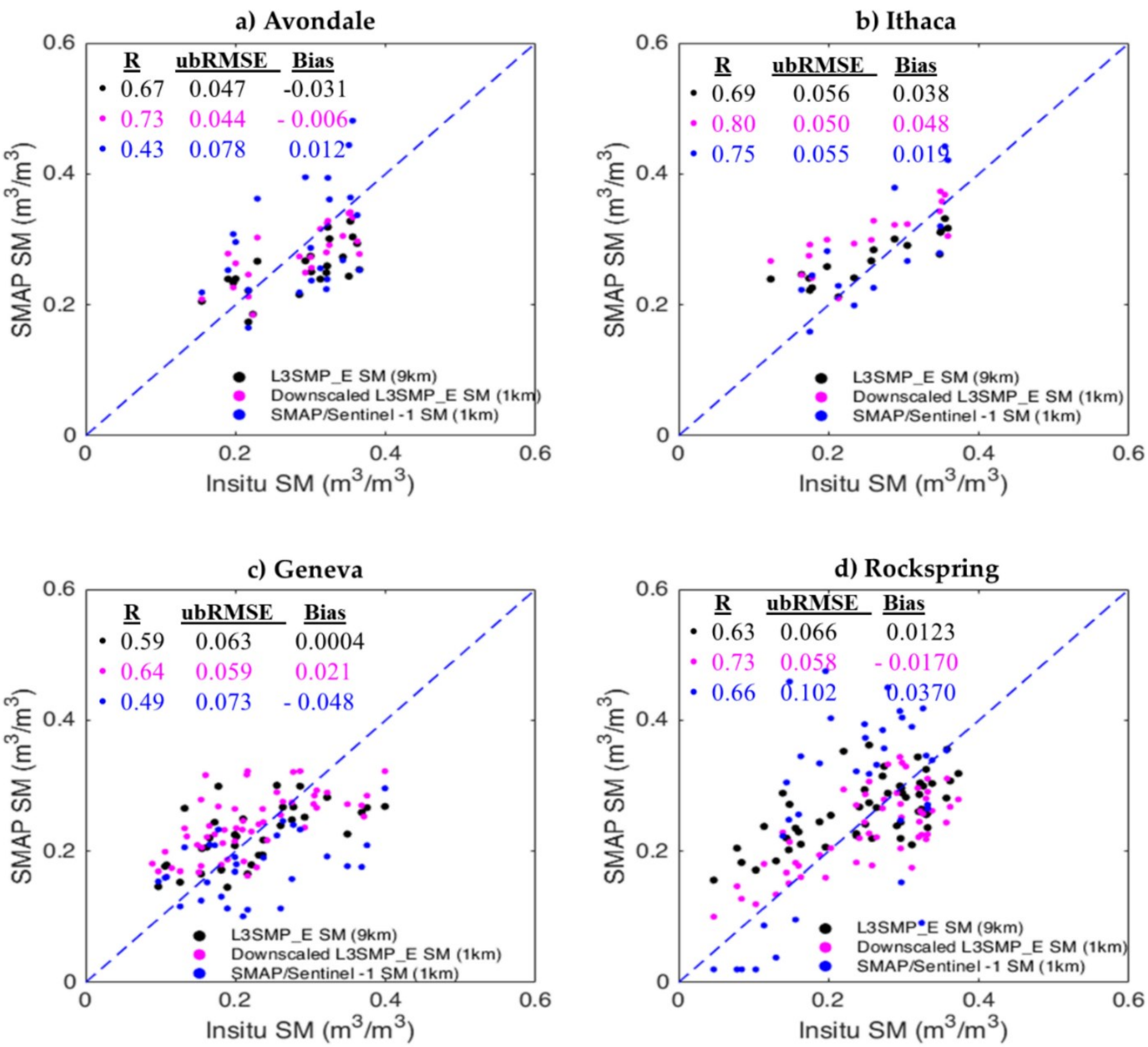

Figure A.0.2 Scatter plot of L3SMP_E, L2_SM_SP and downscaled SM versus in situ soil moisture observation for (a) Avondale, (b) Ithaca, (c) Geneva and (d) Rockspring sites

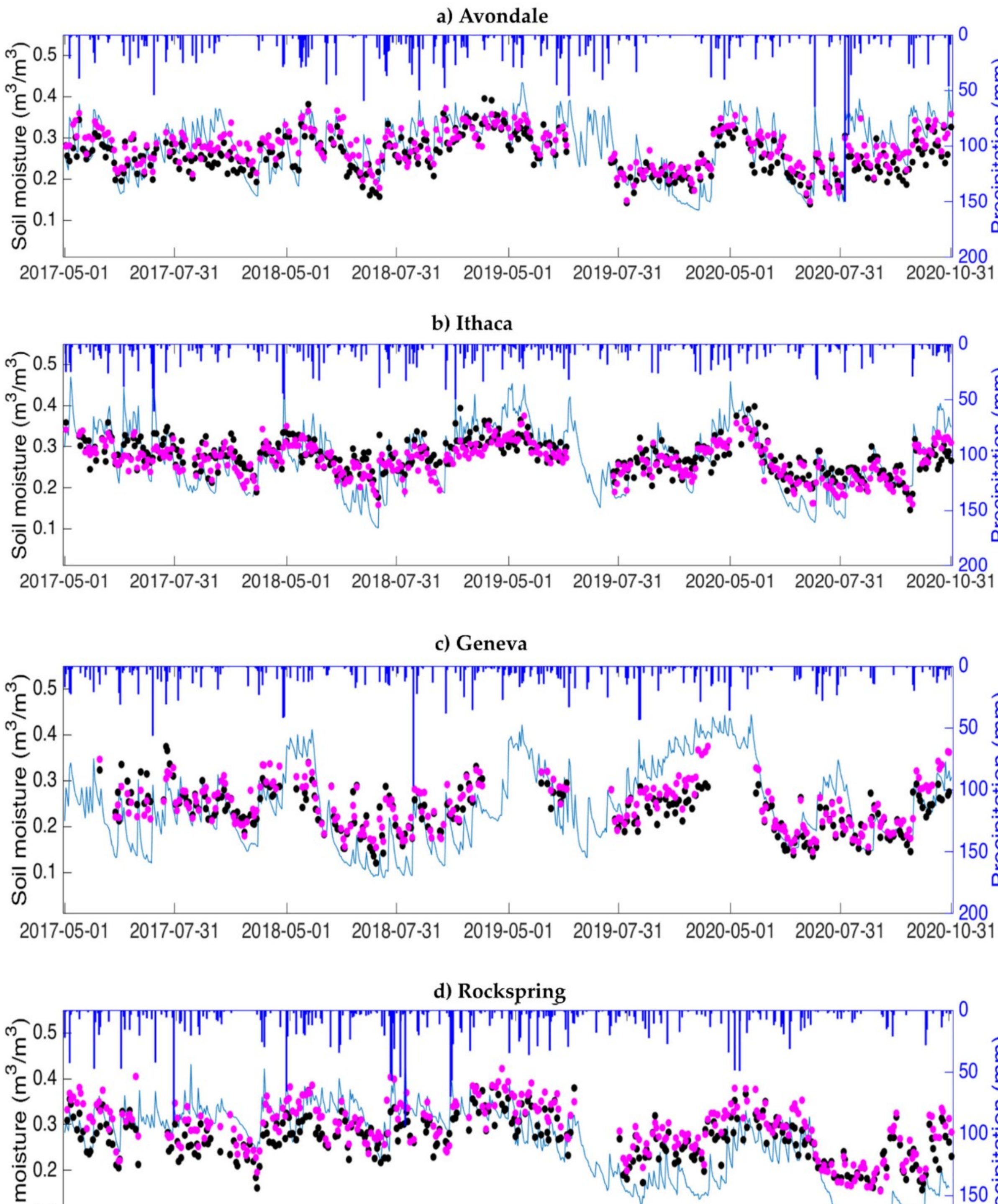

Figure A.0.3 Time series of the downscaled L3SMP_E SM, L3SMP_E SM, L2_SM_SP, daily rainfall and in situ observation for (a) Avondale, (b) Ithaca, (c) Geneva and (d) Rockspring sites

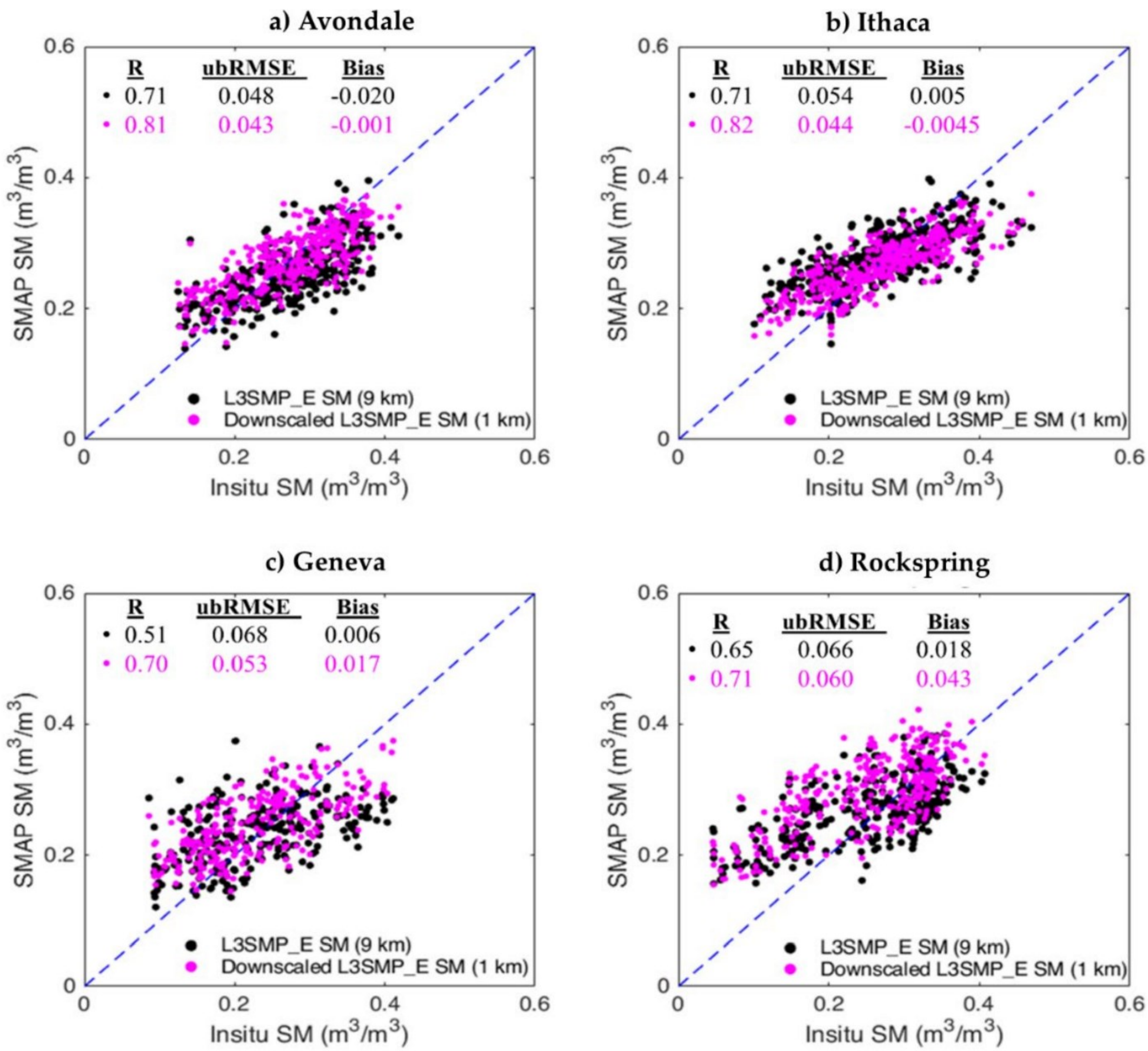

Figure A.0.4 Scatter plot of L3SMP_E, and downscaled SM versus in situ SM observation for (a) Avondale, (b) Ithaca, (c) Geneva and (d) Rockspring sites

List of References

Abbaszadeh, P., Gavahi, K., & Moradkhani, H. (2020). Multivariate remotely sensed and in-situ data assimilation for enhancing community WRF-Hydro model forecasting. *Advances in Water Resources*, *145*(December 2019), 103721. https://doi.org/10.1016/j.advwatres.2020.103721

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). https://doi.org/10.1029/2018WR023354

Abowarda, A. S., Bai, L., Zhang, C., Long, D., Li, X., Huang, Q., & Sun, Z. (2021). Generating surface soil moisture at 30 m spatial resolution using both data fusion and machine learning toward better water resources management at the field scale. *Remote Sensing of Environment*, *255*. https://doi.org/10.1016/j.rse.2021.112301

Albergel, C., Rüdiger, C., Pellarin, T., Calvet, J. C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., Piguet, B., & Martin, E. (2008). From near-surface to root-zone soil moisture using an exponential filter: An assessment of the method based on in-situ observations and model simulations. *Hydrology and Earth System Sciences*, *12*(6), 1323–1337. https://doi.org/10.5194/hess-12-1323-2008

Alvarez-Garreton, C., Ryu, D., Western, A. W., Crow, W. T., & Robertson, D. E. (2014). The impacts of assimilating satellite soil moisture into a rainfall-runoff model in a semi-arid catchment. *Journal of Hydrology*, *519*(PD), 2763–2774. https://doi.org/10.1016/j.jhydrol.2014.07.041

Alvarez-Garreton, C., Ryu, D., Western, A. W., Su, C. H., Crow, W. T., Robertson, D. E., & Leahy, C. (2015). Improving operational flood ensemble prediction by the assimilation of satellite soil moisture: Comparison between lumped and semi-distributed schemes. *Hydrology and Earth System Sciences*, *19*(4), 1659–1676. https://doi.org/10.5194/hess-19-1659-2015

Alvarez-Garreton, Camila, Ryu, D., Western, A. W., Crow, W. T., Su, C. H., & Robertson, D. R. (2016). Dual assimilation of satellite soil moisture to improve streamflow prediction in data-scarce catchments. *Water Resources Research*, *52*(7), 5357–5375. https://doi.org/10.1002/2015WR018429

Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., Santhi, C., Harmel, R. D., Van Griensven, A., Van Liew, M. W., Kannan, N., & Jha, M. K. (2012). SWAT: Model use, calibration, and validation. *Transactions of the ASABE*, *55*(4).

Asano, Y., & Uchida, T. (2010). Is representative elementary area defined by a simple mixing of variable small streams in headwater catchments? *Hydrological Processes*, *24*(5). https://doi.org/10.1002/hyp.7589

Attarzadeh, R., Amini, J., Notarnicola, C., & Greifeneder, F. (2018). Synergetic use of Sentinel-1 and Sentinel-2 data for soil moisture mapping at plot scale. *Remote Sensing*, *10*(8). https://doi.org/10.3390/rs10081285

Attema, E. P. W., & Ulaby, F. T. (1978). Vegetation modeled as a water cloud. *Radio Science*, *13*(2), 357–364. https://doi.org/10.1029/RS013i002p00357

Azimi, S., Dariane, A. B., Modanesi, S., Bauer-Marschallinger, B., Bindlish, R., Wagner, W., & Massari, C. (2020). Assimilation of Sentinel 1 and SMAP – based satellite soil moisture retrievals into SWAT hydrological model: the impact of satellite revisit time and product spatial resolution on flood simulations in small basins. *Journal of Hydrology*, *581*(October 2019), 124367. https://doi.org/10.1016/j.jhydrol.2019.124367

Babaeian, E., Sadeghi, M., Jones, S. B., Montzka, C., Vereecken, H., & Tuller, M. (2019). Ground, Proximal, and Satellite Remote Sensing of Soil Moisture. In *Reviews of Geophysics* (Vol. 57, Issue 2). https://doi.org/10.1029/2018RG000618

Bai, J., Cui, Q., Zhang, W., & Meng, L. (2019). An approach for downscaling SMAP soil moisture by combining Sentinel-1 SAR and MODIS data. *Remote Sensing*, *11*(23). https://doi.org/10.3390/rs11232736

Baldwin, D., Manfreda, S., Keller, K., & Smithwick, E. A. H. (2017). Predicting root zone soil moisture with soil properties and satellite near-surface moisture data across the conterminous United States. *Journal of Hydrology*, *546*.

https://doi.org/10.1016/j.jhydrol.2017.01.020

Balenzano, A., Mattia, F., Satalino, G., Lovergine, F. P., Palmisano, D., Peng, J., Marzahn, P., Wegmüller, U., Cartus, O., Dąbrowska-Zielińska, K., Musial, J. P., Davidson, M. W. J., Pauwels, V. R. N., Cosh, M. H., McNairn, H., Johnson, J. T., Walker, J. P., Yueh, S. H., Entekhabi, D., … Jackson, T. J. (2021). Sentinel-1 soil moisture at 1 km resolution: a validation study. *Remote Sensing of Environment*, *263*(May). https://doi.org/10.1016/j.rse.2021.112554

Barrett, B. W., Dwyer, E., & Whelan, P. (2009). Soil moisture retrieval from active spaceborne microwave observations: An evaluation of current techniques. *Remote Sensing*, *1*(3), 210–242. https://doi.org/10.3390/rs1030210

Beck, H. E., Van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., & De Roo, A. (2017). MSWEP: 3-hourly 0.25° global gridded precipitation (1979-2015) by merging gauge, satellite, and reanalysis data. *Hydrology and Earth System Sciences*, *21*(1). https://doi.org/10.5194/hess-21-589-2017

Belda, M., Holtanová, E., Halenka, T., & Kalvová, J. (2014). Climate classification revisited: From Köppen to Trewartha. *Climate Research*, *59*(1). https://doi.org/10.3354/cr01204

Belgiu, M., & Drăgu, L. (2016). Random forest in remote sensing: A review of applications and future directions. In *ISPRS Journal of Photogrammetry and Remote Sensing* (Vol. 114). https://doi.org/10.1016/j.isprsjprs.2016.01.011

Bergeron, J., Royer, A., Turcotte, R., & Roy, A. (2014). Snow cover estimation using blended MODIS and AMSR-E data for improved watershed-scale spring streamflow simulation in Quebec, Canada. *Hydrological Processes*, *28*(16). https://doi.org/10.1002/hyp.10123

Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, *320*(1–2), 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007

Bolten, J. D., Crow, W. T., Jackson, T. J., Zhan, X., & Reynolds, C. A. (2010). Evaluating the Utility of Remotely Sensed Soil Moisture Retrievals for Operational Agricultural Drought Monitoring. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *3*(1). https://doi.org/10.1109/JSTARS.2009.2037163

Boucher, M. A., Anctil, F., Perreault, L., & Tremblay, D. (2011). A comparison between ensemble and deterministic hydrological forecasts in an operational context. *Advances in Geosciences*, *29*, 85–94. https://doi.org/10.5194/adgeo-29-85-2011

Bowers, S. A., & Hanks, R. J. (1965). Reflection of radiant energy from soils. *Soil Science*, *100*(2). https://doi.org/10.1097/00010694-196508000-00009

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1). https://doi.org/10.1023/A:1010933404324

Brocca, L., Ciabatta, L., Moramarco, T., Ponziani, F., Berni, N., & Wagner, W. (2016). Use of Satellite Soil Moisture Products for the Operational Mitigation of Landslides Risk in Central Italy. In *Satellite Soil Moisture Retrieval: Techniques and Applications*. https://doi.org/10.1016/B978-0-12-803388-3.00012-7

Brocca, L., Melone, F., Moramarco, T., Wagner, W., Naeimi, V., Bartalis, Z., & Hasenauer, S. (2010). Improving runoff prediction through the assimilation of the ASCAT soil moisture product. *Hydrology and Earth System Sciences*, *14*(10). https://doi.org/10.5194/hess-14-1881-2010

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

Brocca, Luca, Melone, F., Moramarco, T., & Wagner, W. (2011). What perspective in remote sensing of soil moisture for hydrological applications by coarse-resolution sensors. *Remote Sensing for Agriculture, Ecosystems, and Hydrology XIII*, *8174*. https://doi.org/10.1117/12.898034

Brocca, Luca, Moramarco, T., Melone, F., Wagner, W., Hasenauer, S., & Hahn, S. (2012). Assimilation of surface- and root-zone ASCAT soil moisture products into rainfall-runoff modeling. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(7 PART1). https://doi.org/10.1109/TGRS.2011.2177468

Brown, M. E., Escobar, V., Moran, S., Entekhabi, D., O'Neill, P. E., Njoku, E. G., Doorn, B., & Entin, J. K. (2013). NASA's soil moisture active passive (SMAP) mission and opportunities for applications users. *Bulletin of the American Meteorological Society*, *94*(8). https://doi.org/10.1175/BAMS-D-11-00049.1

Capehart, W. J., & Carlson, T. N. (1997). Decoupling of surface and near-surface soil water content: A remote sensing perspective. *Water Resources Research*, *33*(6). https://doi.org/10.1029/97WR00617

Carlson, T. (2007). An overview of the "triangle method" for estimating surface evapotranspiration and soil moisture from satellite imagery. In *Sensors* (Vol. 7, Issue 8). https://doi.org/10.3390/s7081612

Carlson, T. N., Capehart, W. J., & Gillies, R. R. (1995). A new look at the simplified method for remote sensing of daily evapotranspiration. *Remote Sensing of Environment*, *54*(2). https://doi.org/10.1016/0034-4257(95)00139-R

Carranza, C. D. U., Van Der Ploeg, M. J., & Torfs, P. J. J. F. (2018). Using lagged dependence to identify (de)coupled surface and subsurface soil moisture values. *Hydrology and Earth System Sciences*, *22*(4). https://doi.org/10.5194/hess-22-2255-2018

Cenci, L., Laiolo, P., Gabellani, S., Campo, L., Silvestro, F., Delogu, F., Boni, G., & Rudari, R. (2016). Assimilation of H-SAF Soil Moisture Products for Flash Flood Early Warning Systems. Case Study: Mediterranean Catchments. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(12), 5634–5646. https://doi.org/10.1109/JSTARS.2016.2598475

Cenci, L., Pulvirenti, L., Boni, G., Chini, M., Matgen, P., Gabellani, S., Squicciarino, G., & Pierdicca, N. (2017). An evaluation of the potential of Sentinel 1 for improving flash flood predictions via soil moisture-data assimilation. *Advances in Geosciences*, *44*, 89–100. https://doi.org/10.5194/adgeo-44-89-2017

Chaubell, J., Yueh, S., Entekhabi, D., & Peng, J. (2016). Resolution enhancement of SMAP radiometer data using the Backus Gilbert optimum interpolation technique. *International Geoscience and Remote Sensing Symposium (IGARSS)*, *2016-November*. https://doi.org/10.1109/IGARSS.2016.7729065

Chauhan, N. S., Miller, S., & Ardanuy, P. (2003). Spaceborne soil moisture estimation at high resolution: A microwave-optical/IR synergistic approach. *International Journal of Remote Sensing*, *24*(22). https://doi.org/10.1080/0143116031000156837

Chen, F., Crow, W. T., Starks, P. J., & Moriasi, D. N. (2011a). Improving hydrologic predictions of a catchment model via assimilation of surface soil moisture. *Advances in Water Resources*, *34*(4), 526–536. https://doi.org/10.1016/j.advwatres.2011.01.011

Chen, F., Crow, W. T., Starks, P. J., & Moriasi, D. N. (2011b). Improving hydrologic predictions of a catchment model via assimilation of surface soil moisture. *Advances in Water Resources*, *34*(4). https://doi.org/10.1016/j.advwatres.2011.01.011

Chen, K. S., Yen, S. K., & Huang, W. P. (1995). A simple model for retrieving bare soil moisture from radar-scattering coefficients. *Remote Sensing of Environment*, *54*(2), 121–126. https://doi.org/10.1016/0034-4257(95)00129-O

Chen, L., Xing, M., He, B., Wang, J., Shang, J., Huang, X., & Xu, M. (2021). Estimating Soil Moisture over Winter Wheat Fields during Growing Season Using Using RADARSAT-2 Data. *Remote Sensing*, *14*. https://doi.org/10.1109/JSTARS.2021.3067890

Chen, Q., Miao, F., Wang, H., Xu, Z. X., Tang, Z., Yang, L., & Qi, S. (2020). Downscaling of Satellite Remote Sensing Soil Moisture Products Over the Tibetan Plateau Based on the Random Forest Algorithm: Preliminary Results. *Earth and Space Science*, *7*(6). https://doi.org/10.1029/2020EA001265

Chen, Q., Miao, F., Xu, Z. X., Wang, H., Yang, L., & Tang, Z. (2019). Downscaling of Remote Sensing Soil Moisture Products Based on TVDI in Complex Terrain Areas. *Proceedings of 2019 International Conference on Meteorology Observations, ICMO 2019*. https://doi.org/10.1109/ICMO49322.2019.9026012

Chen, S., Useya, J., & Mugiyo, H. (2020). Decision-level fusion of Sentinel-1 SAR and Landsat 8 OLI texture features for crop discrimination and classification: case of Masvingo, Zimbabwe. *Heliyon*, *6*(11). https://doi.org/10.1016/j.heliyon.2020.e05358

Choi, M., & Hur, Y. (2012). A microwave-optical/infrared disaggregation for improving spatial representation of soil moisture using AMSR-E and MODIS products. *Remote Sensing of Environment*, *124*. https://doi.org/10.1016/j.rse.2012.05.009

Colliander, A., Jackson, T. J., Bindlish, R., Chan, S., Das, N., Kim, S. B., Cosh, M. H., Dunbar, R. S., Dang, L., Pashaian, L., Asanuma, J., Aida, K., Berg, A., Rowlandson, T., Bosch, D., Caldwell, T., Caylor, K., Goodrich, D., al Jassar, H., … Yueh, S. (2017). Validation of SMAP surface soil moisture products with core validation sites. *Remote Sensing of Environment*, *191*. https://doi.org/10.1016/j.rse.2017.01.021

Colliander, Andreas, Cosh, M. H., Kelly, V. R., Kraatz, S., Bourgeau-Chavez, L., Siqueira, P., Roy, A., Konings, A. G., Holtzman, N., Misra, S., Entekhabi, D., O'Neill, P., & Yueh, S. H. (2020). SMAP Detects Soil Moisture Under Temperate Forest Canopies. *Geophysical Research Letters*, *47*(19). https://doi.org/10.1029/2020GL089697

Crow, W. T., Chen, F., Reichle, R. H., & Liu, Q. (2017a). L band microwave remote sensing and land data assimilation improve the representation of prestorm soil moisture conditions for hydrologic forecasting. *Geophysical Research Letters*, *44*(11). https://doi.org/10.1002/2017GL073642

Crow, W. T., Chen, F., Reichle, R. H., & Liu, Q. (2017b). L band microwave remote sensing and land data assimilation improve the representation of prestorm soil moisture conditions for hydrologic forecasting. *Geophysical Research Letters*, *44*(11), 5495–5503. https://doi.org/10.1002/2017GL073642

Crow, W. T., & Ryu, D. (2009). A new data assimilation approach for improving runoff prediction using remotely-sensed soil moisture retrievals. *Hydrology and Earth System Sciences*, *13*(1). https://doi.org/10.5194/hess-13-1-2009

Crow, Wade T., Berg, A. A., Cosh, M. H., Loew, A., Mohanty, B. P., Panciera, R., De Rosnay, P., Ryu, D., & Walker, J. P. (2012a). Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products. *Reviews of Geophysics*, *50*(2), 1–20. https://doi.org/10.1029/2011RG000372

Crow, Wade T., Berg, A. A., Cosh, M. H., Loew, A., Mohanty, B. P., Panciera, R., De Rosnay, P., Ryu, D., & Walker, J. P. (2012b). 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

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., George H. Taylor, A., Curtisb, J., & Pasterisb, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, *2029*(March 2008), 2011–2029. https://doi.org/10.1002/joc

Das, Entekhabi, D., Dunbar, R. S., Chaubell, M. J., Colliander, A., Yueh, S., Jagdhuber, T., Chen, F., Crow, W., O'Neill, P. E., Walker, J. P., Berg, A., Bosch, D. D., Caldwell, T., Cosh, M. H., Collins, C. H., Lopez-Baeza, E., & Thibeault, M. (2019). The SMAP and Copernicus Sentinel 1A/B microwave active-passive high resolution surface soil moisture product. *Remote Sensing of Environment*, *233*. https://doi.org/10.1016/j.rse.2019.111380

Das, Narendra N., Entekhabi, D., & Njoku, E. G. (2011). An algorithm for merging SMAP radiometer and radar data for high-resolution soil-moisture retrieval. *IEEE Transactions on Geoscience and Remote Sensing*, *49*(5). https://doi.org/10.1109/TGRS.2010.2089526

Das, Narendra Narayan, Entekhabi, D., Dunbar, R. S., Njoku, E. G., & Yueh, S. H. (2016). Uncertainty Estimates in the SMAP Combined Active-Passive Downscaled Brightness Temperature. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(2). https://doi.org/10.1109/TGRS.2015.2450694

Das, Narendra Narayan, Entekhabi, D., Njoku, E. G., Shi, J. J. C., Johnson, J. T., & Colliander, A. (2014). Tests of the SMAP combined radar and radiometer algorithm using airborne field campaign observations and simulated data. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(4). https://doi.org/10.1109/TGRS.2013.2257605

De Jeu, R. A. M., Holmes, T. R. H., Parinussa, R. M., & Owe, M. (2014). A spatially coherent global soil moisture product with improved temporal resolution. *Journal of Hydrology*, *516*. https://doi.org/10.1016/j.jhydrol.2014.02.015

De Lannoy, G. J. M., & Reichle, R. H. (2016). Assimilation of SMOS brightness temperatures or soil moisture retrievals into a land surface model. *Hydrology and Earth System Sciences*, *20*(12). https://doi.org/10.5194/hess-20-4895-2016

DePhilip, M., & Moberg, T. (2010). Ecosystem flow recommendations for the Susquehanna River Basin. *… , Report to the Susquehanna River Basin …*, *December*, 191. http://www.nature.org/media/pa/tnc-final-susquehanna-river-ecosystem-flows-study-report.pdf

Di Luzio, M., Johnson, G. L., Daly, C., Eischeid, J. K., & Arnold, J. G. (2008). Constructing retrospective gridded daily precipitation and temperature datasets for the conterminous United States. *Journal of Applied Meteorology and Climatology*, *47*(2). https://doi.org/10.1175/2007JAMC1356.1

Dickinson, E., Henderson-Sellers, A., & Kennedy, J. (1993). Biosphere-atmosphere Transfer Scheme (BATS) Version 1e as Coupled to the NCAR Community Climate Model. *NCAR Tech. Rep. NCAR/TN-3871STR, 72*, *August*.

Djamai, N., Magagi, R., Goita, K., Merlin, O., Kerr, Y., & Walker, A. (2015). Disaggregation of SMOS soil moisture over the Canadian Prairies. *Remote Sensing of Environment*, *170*. https://doi.org/10.1016/j.rse.2015.09.013

Dobriyal, P., Qureshi, A., Badola, R., & Hussain, S. A. (2012). A review of the methods available for estimating soil moisture and its implications for water resource management. In *Journal of Hydrology* (Vols. 458–459). https://doi.org/10.1016/j.jhydrol.2012.06.021

Dobson, M. C., Ulaby, F. T., Hallikainen, M. T., & El-Rayes, M. A. (1985). Microwave Dielectric Behavior of Wet Soil-Part II: Dielectric Mixing Models. *IEEE Transactions on Geoscience and Remote Sensing*, *GE-23*(1), 35–46. https://doi.org/10.1109/TGRS.1985.289498

Dorigo, W. A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., Gruber, A., Drusch, M., Mecklenburg, S., Van Oevelen, P., Robock, A., & Jackson, T. (2011). The International Soil Moisture Network: A data hosting facility for global in situ soil moisture measurements. *Hydrology and Earth System Sciences*, *15*(5). https://doi.org/10.5194/hess-15-1675-2011

Dubois, P. C., Jakob, van Z., & Engman, T. (1995). Measuring Soil Moisture with Imaging Radars. *IEEE Transactions on Geoscience and Remote Sensing*, *33*(6), 915–926. https://doi.org/10.1109/TGRS.1995.477194

Dunne, T., Moore, T., & Taylor, C. H. (1975). *Recognition and prediction of runoff-producing* *zones in humid regions*.

El Sharif, H., Wang, J., & Georgakakos, A. P. (2015). Modeling regional crop yield and irrigation demand using SMAP type of soil moisture data. *Journal of Hydrometeorology*, *16*(2). https://doi.org/10.1175/JHM-D-14-0034.1

Entekhabi, D. (1995). Recent advances in land-atmosphere interaction research. *Reviews of Geophysics*, *33*(2 S). https://doi.org/10.1029/95RG01163

Entekhabi, D., Njoku, E., Neill, P. O., Spencer, M., Jackson, T., Entin, J., Im, E., & Kellogg, K. (2008). THE SOIL MOISTURE ACTIVE / PASSIVE MISSION (SMAP) Massachusetts Institute of Technology , Cambridge , MA 02139 California Institute of Technology , Jet Propulsion Laboratory , Pasadena , CA 91109-8099 NASA Goddard Space Flight Center , Greenbelt MD 2077. *Jet Propulsion*, 3–6.

Falloon, P., Jones, C. D., Ades, M., & Paul, K. (2011). Direct soil moisture controls of future global soil carbon changes: An important source of uncertainty. *Global Biogeochemical Cycles*, *25*(3). https://doi.org/10.1029/2010GB003938

Fan, S. Di, Hu, Y. M., Wang, L., Liu, Z. H., Shi, Z., Wu, W. Bin, Pan, Y. C., Wang, G. X., Zhu, A. X., & Li, B. (2018). Improving spatial soil moisture representation through the integration of SMAP and PROBA-V products. *Sustainability (Switzerland)*, *10*(10). https://doi.org/10.3390/su10103459

Fang, B., Lakshmi, V., Bindlish, R., Jackson, T. J., Cosh, M., & Basara, J. (2013). Passive Microwave Soil Moisture Downscaling Using Vegetation Index and Skin Surface Temperature. *Vadose Zone Journal*, *12*(3). https://doi.org/10.2136/vzj2013.05.0089

Fortin, J.-P., Richard, T., Serge, M., Roger, M., Jose, F., & Jean-Pierre, V. (2001). DISTRIBUTED WATERSHED MODEL COMPATIBLE WITH REMOTE SENSING AND GIS DATA.I:DESCRIPTION OF MODEL. *Manager*, *6*(December), 515–523.

Fortin, V., Chahinian, N., Montanari, A., Moretti, G., & Moussa, R. (2006). Distributed hydrological modelling with lumped inputs. *IAHS-AISH Publication*, *307*.

Gabriel, G., & Virginia, V. (2017). *Downscaling SMAP soil moisture images with wavelets transform and PCA method*. https://doi.org/10.1109/grss-chile.2017.7996007

Gan, Y., Zhang, Y., Liu, Y., Kongoli, C., & Grassotti, C. (2022). Assimilation of blended in situ-satellite snow water equivalent into the National Water Model for improving hydrologic simulation in two US river basins. *Science of the Total Environment*, *838*(May), 156567. https://doi.org/10.1016/j.scitotenv.2022.156567

Guan, X., Huang, J., Guo, N., Bi, J., & Wang, G. (2009). Variability of soil moisture and its relationship with surface albedo and soil thermal parameters over the Loess Plateau. *Advances in Atmospheric Sciences*, *26*(4). https://doi.org/10.1007/s00376-009-8198-0

Hajj, M. El, Baghdadi, N., Zribi, M., & Bazzi, H. (2017). Synergic use of Sentinel-1 and Sentinel-2 images for operational soil moisture mapping at high spatial resolution over agricultural areas. *Remote Sensing*, *9*(12). https://doi.org/10.3390/rs9121292

Hajnsek, I., Jagdhuber, T., Schön, H., & Papathanassiou, K. P. (2009). Potential of estimating soil moisture under vegetation cover by means of PolSAR. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(2). https://doi.org/10.1109/TGRS.2008.2009642

Hamed Alemohammad, S., Kolassa, J., Prigent, C., Aires, F., & Gentine, P. (2018). Global downscaling of remotely sensed soil moisture using neural networks. *Hydrology and Earth System Sciences*, *22*(10). https://doi.org/10.5194/hess-22-5341-2018

Han, E., Merwade, V., & Heathman, G. C. (2012). Implementation of surface soil moisture data assimilation with watershed scale distributed hydrological model. *Journal of Hydrology*, *416–417*. https://doi.org/10.1016/j.jhydrol.2011.11.039

Haralick, R. M., Dinstein, I., & Shanmugam, K. (1973). Textural Features for Image Classification. *IEEE Transactions on Systems, Man and Cybernetics*, *SMC-3*(6). https://doi.org/10.1109/TSMC.1973.4309314

He, L., Hong, Y., Wu, X., Ye, N., Walker, J. P., & Chen, X. (2018). Investigation of SMAP Active-Passive Downscaling Algorithms Using Combined Sentinel-1 SAR and SMAP Radiometer Data. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(8). https://doi.org/10.1109/TGRS.2018.2842153

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J. N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730). https://doi.org/10.1002/qj.3803

Hu, F., Wei, Z., Zhang, W., Dorjee, D., & Meng, L. (2020). A spatial downscaling method for SMAP soil moisture through visible and shortwave-infrared remote sensing data. *Journal of Hydrology*, *590*. https://doi.org/10.1016/j.jhydrol.2020.125360

Hutengs, C., & Vohland, M. (2016). Downscaling land surface temperatures at regional scales with random forest regression. *Remote Sensing of Environment*, *178*. https://doi.org/10.1016/j.rse.2016.03.006

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). https://doi.org/10.1007/s12665-016-5917-6

Jackson, T. J., & Schmugge, T. J. (1991). Vegetation effects on the microwave emission of soils. *Remote Sensing of Environment*, *36*(3). https://doi.org/10.1016/0034-4257(91)90057-D

Jackson, Thomas J., Cosh, M. H., Bindlish, R., Starks, P. J., Bosch, D. D., Seyfried, M., Goodrich, D. C., Moran, M. S., & Du, J. (2010). Validation of advanced microwave scanning radiometer soil moisture products. *IEEE Transactions on Geoscience and Remote Sensing*, *48*(12). https://doi.org/10.1109/TGRS.2010.2051035

Jackson, Thomas J., & Schmugge, T. J. (1989). Passive Microwave Remote Sensing System for Soil Moisture: Some Supporting Research. *IEEE Transactions on Geoscience and Remote Sensing*, *27*(2). https://doi.org/10.1109/36.20301

Jadidoleslam, N., Mantilla, R., & Krajewski, W. F. (2021). Data assimilation of satellite-based soil moisture into a distributed hydrological model for streamflow predictions. *Hydrology*, *8*(1). https://doi.org/10.3390/hydrology8010052

Jougla, R., & Leconte, R. (2022). Short-Term Hydrological Forecast Using Artificial Neural Network Models with Different Combinations and Spatial Representations of Hydrometeorological Inputs. *Water (Switzerland)*, *14*(4).

https://doi.org/10.3390/w14040552

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). https://doi.org/10.1016/S0034-4257(02)00084-6

Karami, A., Moradi, H. R., Mousivand, A., van Dijk, A. I. J. M., & Renzullo, L. (2022). Using ensemble learning to take advantage of high-resolution radar backscatter in conjunction with surface features to disaggregate SMAP soil moisture product. *International Journal of Remote Sensing*, *43*(3). https://doi.org/10.1080/01431161.2021.2022239

Karthikeyan, L., Pan, M., Wanders, N., Kumar, D. N., & Wood, E. F. (2017). Four decades of microwave satellite soil moisture observations: Part 1. A review of retrieval algorithms. *Advances in Water Resources*, *109*, 106–120. https://doi.org/10.1016/j.advwatres.2017.09.006

Kelly, R. E. J., Davie, T. J. A., & Atkinson, P. M. (2003). Explaining temporal and spatial variation in soil moisture in a bare field using SAR imagery. *International Journal of Remote Sensing*, *24*(15), 3059–3074. https://doi.org/10.1080/01431160110115771

Kerr, Y. H., Waldteufel, P., Wigneron, J. P., Martinuzzi, J. M., Font, J., & Berger, M. (2001). Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission. *IEEE Transactions on Geoscience and Remote Sensing*, *39*(8). https://doi.org/10.1109/36.942551

Kim, D., Lee, J., Kim, H., & Choi, M. (2016). Spatial composition of AMSR2 soil moisture products by conditional merging technique with ground soil moisture data. *Stochastic Environmental Research and Risk Assessment*, *30*(8). https://doi.org/10.1007/s00477-016-1300-0

Kim, J., & Hogue, T. S. (2012a). Through Integration of AMSR-E and MODIS Products. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(2), 446–460.

Kim, J., & Hogue, T. S. (2012b). Improving spatial soil moisture representation through integration of AMSR-E and MODIS products. *IEEE Transactions on Geoscience and*

Remote Sensing, *50*(2), 446–460. https://doi.org/10.1109/TGRS.2011.2161318

Kizito, F., Campbell, C. S., Campbell, G. S., Cobos, D. R., Teare, B. L., Carter, B., & Hopmans, J. W. (2008). Frequency, electrical conductivity and temperature analysis of a low-cost capacitance soil moisture sensor. *Journal of Hydrology*, *352*(3–4). https://doi.org/10.1016/j.jhydrol.2008.01.021

Kolassa, J., Aires, F., Polcher, J., Prigent, C., Jimenez, C., & Pereira, J. M. (2013). Soil moisture retrieval from multi-instrument observations: Information content analysis and retrieval methodology. *Journal of Geophysical Research Atmospheres*, *118*(10). https://doi.org/10.1029/2012JD018150

Kornelsen, K. C., & Coulibaly, P. (2013). McMaster Mesonet soil moisture dataset: Description and spatio-temporal variability analysis. *Hydrology and Earth System Sciences*, *17*(4). https://doi.org/10.5194/hess-17-1589-2013

Kornelsen, Kurt C., & Coulibaly, P. (2013). Advances in soil moisture retrieval from synthetic aperture radar and hydrological applications. *Journal of Hydrology*, *476*, 460–489. https://doi.org/10.1016/j.jhydrol.2012.10.044

Kumar, S. V., Peters-Lidard, C. D., Mocko, D., Reichle, R., Liu, Y., Arsenault, K. R., Xia, Y., Ek, M., Riggs, G., Livneh, B., & Cosh, M. (2014). Assimilation of remotely sensed soil moisture and snow depth retrievals for drought estimation. *Journal of Hydrometeorology*, *15*(6). https://doi.org/10.1175/JHM-D-13-0132.1

Lakhankar, T., Krakauer, N., & Khanbilvardi, R. (2009). Applications of microwave remote sensing of soil moisture for agricultural applications. *International Journal of Terraspace Science and Engineering*, *2*(1).

Landau, L. D., Lifshitz, E. M., & King, A. L. (1961). Electrodynamics of Continuous Media. *American Journal of Physics*, *29*(9). https://doi.org/10.1119/1.1937882

Le, M.-H., Nguyen, B. Q., Pham, H. T., Patil, A., Do, H. X., Ramsankaran, R., Bolten, J. D., & Lakshmi, V. (2022). Assimilation of SMAP Products for Improving Streamflow Simulations over Tropical Climate Region—Is Spatial Information More Important Than Temporal Information? *Remote Sensing*, *14*(7), 1607. https://doi.org/10.3390/rs14071607

Leroux, D. J., Pellarin, T., Vischel, T., Cohard, J. M., Gascon, T., Gibon, F., Mialon, A., Galle, S., Peugeot, C., & Seguis, L. (2016). Assimilation of SMOS soil moisture into a distributed hydrological model and impacts on the water cycle variables over the Ouémé catchment in Benin. *Hydrology and Earth System Sciences*, *20*(7), 2827–2840. https://doi.org/10.5194/hess-20-2827-2016

Li, J., Wang, S., Gunn, G., Joosse, P., & Russell, H. A. J. (2018). A model for downscaling SMOS soil moisture using Sentinel-1 SAR data. *International Journal of Applied Earth Observation and Geoinformation*, *72*. https://doi.org/10.1016/j.jag.2018.07.012

Li, Z. L., Leng, P., Zhou, C., Chen, K. S., Zhou, F. C., & Shang, G. F. (2021). Soil moisture retrieval from remote sensing measurements: Current knowledge and directions for the future. *Earth-Science Reviews*, *218*(March), 103673. https://doi.org/10.1016/j.earscirev.2021.103673

Lievens, H., Reichle, R. H., Liu, Q., De Lannoy, G. J. M., Dunbar, R. S., Kim, S. B., Das, N. N., Cosh, M., Walker, J. P., & Wagner, W. (2017). Joint Sentinel-1 and SMAP data assimilation to improve soil moisture estimates. *Geophysical Research Letters*, *44*(12). https://doi.org/10.1002/2017GL073904

Lievens, H., Tomer, S. K., Al Bitar, A., De Lannoy, G. J. M., Drusch, M., Dumedah, G., Hendricks Franssen, H. J., Kerr, Y. H., Martens, B., Pan, M., Roundy, J. K., Vereecken, H., Walker, J. P., Wood, E. F., Verhoest, N. E. C., & Pauwels, V. R. N. (2015). SMOS soil moisture assimilation for improved hydrologic simulation in the Murray Darling Basin, Australia. *Remote Sensing of Environment*, *168*, 146–162. https://doi.org/10.1016/j.rse.2015.06.025

Lin, D. S., Wood, E. F., Beven, K., & Saatchi, S. (1994). Soil moisture estimation over grass-covered areas using AIRSAR. *International Journal of Remote Sensing*, *15*(11). https://doi.org/10.1080/01431169408954246

Liu, Y. Y., Parinussa, R. M., Dorigo, W. A., De Jeu, R. A. M., Wagner, W., M. Van Dijk, A. I. J., McCabe, M. F., & Evans, J. P. (2011). Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals. *Hydrology and Earth System Sciences*, *15*(2). https://doi.org/10.5194/hess-15-425-2011

Liu, Yangxiaoyue, Jing, W., Wang, Q., & Xia, X. (2020). Generating high-resolution daily soil moisture by using spatial downscaling techniques: a comparison of six machine learning algorithms. *Advances in Water Resources*, *141*. https://doi.org/10.1016/j.advwatres.2020.103601

Liu, Yangxiaoyue, Xia, X., Yao, L., Jing, W., Zhou, C., Huang, W., Li, Y., & Yang, J. (2020). Downscaling Satellite Retrieved Soil Moisture Using Regression Tree-Based Machine Learning Algorithms Over Southwest France. *Earth and Space Science*, *7*(10). https://doi.org/10.1029/2020EA001267

Liu, Yangxiaoyue, Yang, Y., Jing, W., & Yue, X. (2018). Comparison of different machine learning approaches for monthly satellite-based soil moisture downscaling over Northeast China. *Remote Sensing*, *10*(1). https://doi.org/10.3390/rs10010031

Liu, Yuqiong, & Gupta, H. V. (2007). Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resources Research*, *43*(7), 1–18. https://doi.org/10.1029/2006WR005756

Lobell, D. B., & Asner, G. P. (2002). Moisture Effects on Soil Reflectance. *Soil Science Society of America Journal*, *66*(3). https://doi.org/10.2136/sssaj2002.7220

Loizu, J., Massari, C., Álvarez-Mozos, J., Tarpanelli, A., Brocca, L., & Casalí, J. (2018a). On the assimilation set-up of ASCAT soil moisture data for improving streamflow catchment simulation. *Advances in Water Resources*, *111*(November 2017), 86–104. https://doi.org/10.1016/j.advwatres.2017.10.034

Loizu, J., Massari, C., Álvarez-Mozos, J., Tarpanelli, A., Brocca, L., & Casalí, J. (2018b). On the assimilation set-up of ASCAT soil moisture data for improving streamflow catchment simulation. *Advances in Water Resources*, *111*(December 2016), 86–104. https://doi.org/10.1016/j.advwatres.2017.10.034

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*. https://doi.org/10.1016/j.rse.2019.111364

López López, P., Wanders, N., Schellekens, J., Renzullo, L. J., Sutanudjaja, E. H., & Bierkens, M. F. P. (2016). Improved large-scale hydrological modelling through the assimilation of streamflow and downscaled satellite soil moisture observations. *Hydrology and Earth System Sciences*, *20*(7), 3059–3076. https://doi.org/10.5194/hess-20-3059-2016

Lv, A., Zhang, Z., & Zhu, H. (2021). A neural-network based spatial resolution downscaling method for soil moisture: Case study of qinghai province. *Remote Sensing*, *13*(8). https://doi.org/10.3390/rs13081583

Ma, C., Li, X., & McCabe, M. F. (2020). Retrieval of high-resolution soil moisture through combination of Sentinel-1 and Sentinel-2 data. *Remote Sensing*, *12*(14), 1–28. https://doi.org/10.3390/rs12142303

Magagi, R., Berg, A. A., Goita, K., Belair, S., Jackson, T. J., Toth, B., Walker, A., McNairn, H., O'Neill, P. E., Moghaddam, M., Gherboudj, I., Colliander, A., Cosh, M. H., Burgin, M., Fisher, J. B., Kim, S. B., Mladenova, I., Djamai, N., Rousseau, L. P. B., … Merzouki, A. (2013). Canadian experiment for soil moisture in 2010 (CanEx-SM10): Overview and preliminary results. *IEEE Transactions on Geoscience and Remote Sensing*, *51*(1). https://doi.org/10.1109/TGRS.2012.2198920

Magagi, R. D., & Kerr, Y. H. (1997). Retrieval of soil moisture and vegetation characteristics by use of ERS-1 wind scatterometer over arid and semi-arid areas. *Journal of Hydrology*, *188–189*(1–4), 361–384. https://doi.org/10.1016/S0022-1694(96)03166-6

Mao, Y., Crow, W. T., & Nijssen, B. (2019). A framework for diagnosing factors degrading the streamflow performance of a soil moisture data assimilation system. *Journal of Hydrometeorology*, *20*(1), 79–97. https://doi.org/10.1175/JHM-D-18-0115.1

Martínez-Fernández, J., & Ceballos, A. (2005). Mean soil moisture estimation using temporal stability analysis. *Journal of Hydrology*, *312*(1–4). https://doi.org/10.1016/j.jhydrol.2005.02.007

Massari, C., Brocca, L., Tarpanelli, A., & Moramarco, T. (2015). Data assimilation of satellite soil moisture into rainfall-runoffmodelling: A complex recipe? In *Remote Sensing* (Vol. 7, Issue 9). https://doi.org/10.3390/rs70911403

Massari, C., Camici, S., Ciabatta, L., & Brocca, L. (2018a). Exploiting satellite-based surface soil moisture for flood forecasting in the Mediterranean area: State update versus rainfall correction. *Remote Sensing*. https://doi.org/10.3390/rs10020292

Massari, C., Camici, S., Ciabatta, L., & Brocca, L. (2018b). Exploiting satellite-based surface soil moisture for flood forecasting in the Mediterranean area: State update versus rainfall correction. *Remote Sensing*, *10*(2). https://doi.org/10.3390/rs10020292

McGarragh, G., Poulsen, C., Povey, A., Thomas, G., Christensen, M., Sus, O., Schlundt, C., Stapelberg, S., Stengel, M., Grainger, D., McGarragh, G., Poulsen, C., Povey, A., Thomas, G., Christensen, M., Sus, O., Schlundt, C., Stapelberg, S., Stengel, M., & Grainger, D. (2015). SNAP (Sentinel Application Platform) and the ESA Sentinel 3 Toolbox. *ESASP*, *734*.

Meng, S., Xie, X., & Liang, S. (2017). Assimilation of soil moisture and streamflow observations to improve flood forecasting with considering runoff routing lags. *Journal of Hydrology*, *550*. https://doi.org/10.1016/j.jhydrol.2017.05.024

Merlin, O., Rüdiger, C., Al Bitar, A., Richaume, P., Walker, J. P., & Kerr, Y. H. (2012). Disaggregation of SMOS soil moisture in Southeastern Australia. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(5 PART 1). https://doi.org/10.1109/TGRS.2011.2175000

Mironov, V. L., Dobson, M. C., Kaupp, V. H., Komarov, S. A., & Kleshchenko, V. N. (2004). Generalized refractive mixing dielectric model for moist soils. *IEEE Transactions on Geoscience and Remote Sensing*, *42*(4), 773–785. https://doi.org/10.1109/TGRS.2003.823288

Mohseni, F., Mirmazloumi, S. M., Mokhtarzade, M., Jamali, S., & Homayouni, S. (2022). Global Evaluation of SMAP/Sentinel-1 Soil Moisture Products. *Remote Sensing*, *14*(18), 1–31. https://doi.org/10.3390/rs14184624

Montzka, C., Rotzer, K., Bogena, H. R., & Vereecken, H. (2017). Sub-grid soil moisture variability of SMAP, SMOS and ASCAT predi cted on basis of soil texture, links to NetCDF files. In *Supplement to: Montzka, C et al. (submitted): A new soil moisture downscaling approach for SMAP, SMOS and ASCAT by predicting sub-grid variability.*

Remote Sensing.

Moran, M. S., Clarke, T. R., Inoue, Y., & Vidal, A. (1994). Estimating crop water deficit using the relation between surface-air temperature and spectral vegetation index. *Remote Sensing of Environment*, *49*(3). https://doi.org/10.1016/0034-4257(94)90020-5

Muñoz-Carpena, R. (1969). Field Devices For Monitoring Soil Water Content. *Edis*, *2004*(8). https://doi.org/10.32473/edis-ae266-2004

Narayan, U., Lakshmi, V., & Jackson, T. J. (2006). High-resolution change estimation of soil moisture using L-band radiometer and radar observations made during the SMEX02 experiments. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(6). https://doi.org/10.1109/TGRS.2006.871199

Neelam, M., & Mohanty, B. P. (2020). On the radiative transfer model for soil moisture across space, time and hydro-climates. *Remote Sensing*, *12*(16). https://doi.org/10.3390/RS12162645

Njoku, E. G., & Entekhabi, D. (1996). Passive microwave remote sensing of soil moisture. *Journal of Hydrology*, *184*(1–2). https://doi.org/10.1016/0022-1694(95)02970-2

Nocita, M., Stevens, A., Noon, C., & Van Wesemael, B. (2013). Prediction of soil organic carbon for different levels of soil moisture using Vis-NIR spectroscopy. *Geoderma*, *199*. https://doi.org/10.1016/j.geoderma.2012.07.020

Noilhan, J., & Mahfouf, J. F. (1996). The ISBA land surface parameterisation scheme. *Global and Planetary Change*, *13*(1–4). https://doi.org/10.1016/0921-8181(95)00043-7

Oh, Y., Sarabandi, K., & Ulaby, F. T. (1992). An Empirical Model and an Inversion Technique for Radar Scattering from Bare Soil Surfaces. *IEEE Transactions on Geoscience and Remote Sensing*, *30*(2). https://doi.org/10.1109/36.134086

Ojo, E. R., Bullock, P. R., L'Heureux, J., Powers, J., McNairn, H., & Pacheco, A. (2015). Calibration and Evaluation of a Frequency Domain Reflectometry Sensor for Real-Time Soil Moisture Monitoring. *Vadose Zone Journal*, *14*(3). https://doi.org/10.2136/vzj2014.08.0114

Owe, M., De Jeu, R., & Walker, J. (2001). A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index. *IEEE Transactions on Geoscience and Remote Sensing*, *39*(8). https://doi.org/10.1109/36.942542

Pablos, M., González-Zamora, A., Sánchez, N., Portal, G., Vall-Llossera, M., & Martínez-Fernández, J. (2018). Surface soil moisture estimation from MODIS apparent thermal inertia: A comparison with SMOS and SMAP soil moisture products. *International Geoscience and Remote Sensing Symposium (IGARSS)*, *2018-July*. https://doi.org/10.1109/IGARSS.2018.8518847

Palace, M., Keller, M., Asner, G. P., Hagen, S., & Braswell, B. (2008). Remote Sensing: Models and Methods for Image Processing, 3rd ed. *Biotropica*, *40*(2).

Paloscia, S., Pettinato, S., Santi, E., Notarnicola, C., Pasolli, L., & Reppucci, A. (2013). Soil moisture mapping using Sentinel-1 images: Algorithm and preliminary validation. *Remote Sensing of Environment*, *134*, 234–248. https://doi.org/10.1016/j.rse.2013.02.027

Parajka, J., Naeimi, V., Blöschl, G., Wagner, W., Merz, R., & Scipal, K. (2006). Assimilating scatterometer soil moisture data into conceptual hydrologic models at the regional scale. *Hydrology and Earth System Sciences*, *10*(3), 353–368. https://doi.org/10.5194/hess-10-353-2006

Park, S., Park, S., Im, J., Rhee, J., Shin, J., & Park, J. D. (2017). Downscaling GLDAS Soil moisture data in East Asia through fusion of Multi-Sensors by optimizing modified regression trees. *Water (Switzerland)*, *9*(5). https://doi.org/10.3390/w9050332

Pasolli, L., Notarnicola, C., Bruzzone, L., Bertoldi, G., Della Chiesa, S., Niedrist, G., Tappeiner, U., & Zebisch, M. (2012). Polarimetric RADARSAT-2 imagery for soil moisture retrieval in alpine areas. *Canadian Journal of Remote Sensing*, *37*(5). https://doi.org/10.5589/m11-065

Patil, A., & Ramsankaran, R. A. A. J. (2017). Improving streamflow simulations and forecasting performance of SWAT model by assimilating remotely sensed soil moisture observations. *Journal of Hydrology*, *555*, 683–696. https://doi.org/10.1016/j.jhydrol.2017.10.058

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., 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*, *12*.

Peng, J., 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). https://doi.org/10.5194/hess-19-4765-2015

Peng, Jian, Albergel, C., Balenzano, A., Brocca, L., Cartus, O., Cosh, M. H., Crow, W. T., Dabrowska-Zielinska, K., Dadson, S., Davidson, M. W. J., de Rosnay, P., Dorigo, W., Gruber, A., Hagemann, S., Hirschi, M., Kerr, Y. H., Lovergine, F., Mahecha, M. D., Marzahn, P., … Loew, A. (2021). A roadmap for high-resolution satellite soil moisture applications – confronting product characteristics with user requirements. In *Remote Sensing of Environment* (Vol. 252). https://doi.org/10.1016/j.rse.2020.112162

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

Petropoulos, G., Carlson, T. N., Wooster, M. J., & Islam, S. (2009). A review of Ts/VI remote sensing based methods for the retrieval of land surface energy fluxes and soil surface moisture. In *Progress in Physical Geography* (Vol. 33, Issue 2). https://doi.org/10.1177/0309133309338997

Petropoulos, G. P., Ireland, G., & Barrett, B. (2015). Surface soil moisture retrievals from remote sensing: Current status, products & future trends. In *Physics and Chemistry of the Earth* (Vols. 83–84). https://doi.org/10.1016/j.pce.2015.02.009

Piles, M., & Sánchez, N. (2016). Spatial Downscaling of Passive Microwave Data With Visible-to-Infrared Information for High-Resolution Soil Moisture Mapping. In *Satellite Soil Moisture Retrieval: Techniques and Applications*. https://doi.org/10.1016/B978-0-12-803388-3.00006-1

Piles, María, Camps, A., Vall-Llossera, M., Corbella, I., Panciera, R., Rudiger, C., Kerr, Y. H., & Walker, J. (2011). Downscaling SMOS-derived soil moisture using MODIS visible/infrared data. *IEEE Transactions on Geoscience and Remote Sensing*, *49*(9), 3156–3166. https://doi.org/10.1109/TGRS.2011.2120615

Piles, Maria, 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). https://doi.org/10.1109/TGRS.2009.2022088

Poe, G. A., & Conway, R. W. (1990). A Study of the Geolocation Errors of the Special Sensor Microwave/Imager (SSM/I). *IEEE Transactions on Geoscience and Remote Sensing*, *28*(5). https://doi.org/10.1109/36.58965

Rao, K. S., Raju, S., & Wang, J. R. (1993). Estimation of Soil Moisture and Surface Roughness Parameters from Backscattering Coefficient. *IEEE Transactions on Geoscience and Remote Sensing*, *31*(5), 1094–1099. https://doi.org/10.1109/36.263781

Rao, P., Wang, Y., Wang, F., Liu, Y., Wang, X., & Wang, Z. (2022). Daily soil moisture mapping at 1 km resolution based on SMAP data for desertification areas in northern China. *Earth System Science Data*, *14*(7). https://doi.org/10.5194/essd-14-3053-2022

Reichle, R. H., & Koster, R. D. (2004). Bias reduction in short records of satellite soil moisture. *Geophysical Research Letters*, *31*(19). https://doi.org/10.1029/2004GL020938

Reichle, Rolf H. (2008). Data assimilation methods in the Earth sciences. *Advances in Water Resources*, *31*(11), 1411–1418. https://doi.org/10.1016/j.advwatres.2008.01.001

Renzullo, L. J., van Dijk, A. I. J. M., Perraud, J. M., Collins, D., Henderson, B., Jin, H., Smith, A. B., & McJannet, D. L. (2014). Continental satellite soil moisture data assimilation improves root-zone moisture analysis for water resources assessment. *Journal of Hydrology*, *519*(PD). https://doi.org/10.1016/j.jhydrol.2014.08.008

Reynolds, S. G. (1970). The gravimetric method of soil moisture determination Part I A study of equipment, and methodological problems. *Journal of Hydrology*, *11*(3). https://doi.org/10.1016/0022-1694(70)90066-1

Rudiger, C., Su, C. H., Ryu, D., & Wagner, W. (2016). Disaggregation of Low-Resolution L-Band Radiometry Using C-Band Radar Data. *IEEE Geoscience and Remote Sensing*

Letters, *13*(10). https://doi.org/10.1109/LGRS.2016.2583433

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

Samuel, J., Rousseau, A. N., Abbasnezhadi, K., & Savary, S. (2019). Development and evaluation of a hydrologic data-assimilation scheme for short-range flow and inflow forecasts in a data-sparse high-latitude region using a distributed model and ensemble Kalman filtering. *Advances in Water Resources*, *130*. https://doi.org/10.1016/j.advwatres.2019.06.004

Sánchez-Ruiz, S., Piles, M., Sánchez, N., Martínez-Fernández, J., Vall-llossera, M., & Camps, A. (2014). Combining SMOS with visible and near/shortwave/thermal infrared satellite data for high resolution soil moisture estimates. *Journal of Hydrology*, *516*. https://doi.org/10.1016/j.jhydrol.2013.12.047

Santi, E., Paloscia, S., Pettinato, S., Notarnicola, C., Pasolli, L., & Pistocchi, A. (2013). Comparison between SAR soil moisture estimates and hydrological model simulations over the Scrivia test site. *Remote Sensing*, *5*(10). https://doi.org/10.3390/rs5104961

Sarker, M. L. R., Nichol, J., Iz, H. B., Ahmad, B. Bin, & Rahman, A. A. (2013). Forest biomass estimation using texture measurements of high-resolution dual-polarization C-band SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, *51*(6). https://doi.org/10.1109/TGRS.2012.2219872

Schmugge, T. (1976). Remote sensing of soil moisture. *In Symp. C., Cospar Meeting*, *1999*(December), 1–6.

Schmugge, T., O'neill, P. E., & Wanga, J. R. (1986). Passive Microwave Soil Moisture Research. *IEEE Transactions on Geoscience and Remote Sensing*, *GE-24*(1). https://doi.org/10.1109/TGRS.1986.289584

Senanayake, I. P., Yeo, I. Y., Tangdamrongsub, N., Willgoose, G. R., Hancock, G. R., Wells, T., Fang, B., & Lakshmi, V. (2017). Downscaling SMAP and SMOS soil moisture retrievals over the Goulburn River Catchment, Australia. *Proceedings - 22nd*

International Congress on Modelling and Simulation, MODSIM 2017. https://doi.org/10.36334/modsim.2017.h6.senanayake

Senanayake, I. P., Yeo, I. Y., Willgoose, G. R., & Hancock, G. R. (2021). Disaggregating satellite soil moisture products based on soil thermal inertia: A comparison of a downscaling model built at two spatial scales. *Journal of Hydrology*, *594*. https://doi.org/10.1016/j.jhydrol.2020.125894

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., & Teuling, A. J. (2010a). Investigating soil moisture-climate interactions in a changing climate: A review. *Earth-Science Reviews*, *99*(3–4), 125–161. https://doi.org/10.1016/j.earscirev.2010.02.004

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., & Teuling, A. J. (2010b). Investigating soil moisture-climate interactions in a changing climate: A review. In *Earth-Science Reviews* (Vol. 99, Issues 3–4). https://doi.org/10.1016/j.earscirev.2010.02.004

Sharma, P. K., Kumar, D., Srivastava, H. S., & Patel, P. (2018). Assessment of Different Methods for Soil Moisture Estimation : A Review. *Journal of Remote Sensing & GIS*, *9*(1), 57–73.

Shi, Y., Song, L., Xia, Z., Lin, Y., Myneni, R. B., Choi, S., Wang, L., Ni, X., Lao, C., & Yang, F. (2015). Mapping annual precipitation across Mainland China in the period 2001-2010 from TRMM3B43 product using spatial downscaling approach. *Remote Sensing*, *7*(5). https://doi.org/10.3390/rs70505849

Shoshany, M., Svoray, T., Svoray, T., Curran, P. J., Foody, G. M., & Perevolotsky, A. (2000). The relationship between ERS-2 SAR backscatter and soil moisture: Generalization from a humid to semi-arid transect. *International Journal of Remote Sensing*, *21*(11), 2337–2343. https://doi.org/10.1080/01431160050029620

Sinclair, S., & Pegram, G. (2005). Combining radar and rain gauge rainfall estimates using conditional merging. *Atmospheric Science Letters*, *6*(1). https://doi.org/10.1002/asl.85

Soliman, A., Heck, R. J., Brenning, A., Brown, R., & Miller, S. (2013). Remote sensing of soil moisture in vineyards using airborne and ground-based thermal inertia data. *Remote Sensing*, *5*(8). https://doi.org/10.3390/rs5083729

Srivastava, P. K., Han, D., Ramirez, M. R., & Islam, T. (2013). Machine Learning Techniques for Downscaling SMOS Satellite Soil Moisture Using MODIS Land Surface Temperature for Hydrological Application. *Water Resources Management*, *27*(8). https://doi.org/10.1007/s11269-013-0337-9

Stroud, P. (1999). A recursive exponential filter for time-sensitive data. *Rep. LAUR-99*.

Sun, H., Cai, C., Liu, H., & Yang, B. (2019). Microwave and Meteorological Fusion: A method of Spatial Downscaling of Remotely Sensed Soil Moisture. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(4). https://doi.org/10.1109/JSTARS.2019.2901921

Sun, H., & Cui, Y. (2021). Evaluating downscaling factors of microwave satellite soil moisture based on machine learning method. *Remote Sensing*, *13*(1). https://doi.org/10.3390/rs13010133

Sun, H., Zhou, B., & Liu, H. (2019). Spatial evaluation of soil moisture (SM), land surface temperature (LST), and LST-Derived SM indexes dynamics during SMAPVEX12. *Sensors (Switzerland)*, *19*(5). https://doi.org/10.3390/s19051247

Sun, L., Seidou, O., Nistor, I., Goïta, K., & Magagi, R. (2016). Simultaneous assimilation of in situ soil moisture and streamflow in the SWAT model using the Extended Kalman Filter. *Journal of Hydrology*, *543*. https://doi.org/10.1016/j.jhydrol.2016.10.040

Susha Lekshmi, S. U., Singh, D. N., & Shojaei Baghini, M. (2014). A critical review of soil moisture measurement. *Measurement: Journal of the International Measurement Confederation*, *54*, 92–105. https://doi.org/10.1016/j.measurement.2014.04.007

Tang, G., Clark, M. P., & Papalexiou, S. M. (2022). EM-Earth: The Ensemble Meteorological Dataset for Planet Earth. *Bulletin of the American Meteorological Society*, 996–1018. https://doi.org/10.1175/bams-d-21-0106.1

Tolson, B. A., & Shoemaker, C. A. (2007). Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research*, *43*(1).

https://doi.org/10.1029/2005WR004723

Tomer, S. K., Al Bitar, A., Sekhar, M., Zribi, M., Bandyopadhyay, S., Sreelash, K., Sharma, A. K., Corgne, S., & Kerr, Y. (2015). Retrieval and multi-scale validation of Soil Moisture from multi-temporal SAR Data in a semi-arid tropical region. *Remote Sensing*, *7*(6), 8128–8153. https://doi.org/10.3390/rs70608128

Tomer, S. K., Bitar, A. Al, Sekhar, M., Zribi, M., Bandyopadhyay, S., & Kerr, Y. (2016). MAPSM: A spatio-temporal algorithm for merging soil moisture from active and passive microwave remote sensing. *Remote Sensing*, *8*(12). https://doi.org/10.3390/rs8120990

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B. Ö., Floury, N., Brown, M., Traver, I. N., Deghaye, P., Duesmann, B., Rosich, B., Miranda, N., Bruno, C., L'Abbate, M., Croci, R., Pietropaolo, A., … Rostan, F. (2012). GMES Sentinel-1 mission. *Remote Sensing of Environment*, *120*. https://doi.org/10.1016/j.rse.2011.05.028

Tromp-van Meerveld, H. J., & McDonnell, J. J. (2006). On the interrelations between topography, soil depth, soil moisture, transpiration rates and species distribution at the hillslope scale. *Advances in Water Resources*, *29*(2). https://doi.org/10.1016/j.advwatres.2005.02.016

Tuttle, S. E., & Salvucci, G. D. (2014). A new approach for validating satellite estimates of soil moisture using large-scale precipitation: Comparing AMSR-E products. *Remote Sensing of Environment*, *142*. https://doi.org/10.1016/j.rse.2013.12.002

Velpuri, N. M., Senay, G. B., & Morisette, J. T. (2015). Evaluating New SMAP Soil Moisture for Drought Monitoring in the Rangelands of the US High Plains. *Rangelands*. https://doi.org/10.1016/j.rala.2016.06.002

Vereecken, H., Huisman, J. A., Bogena, H., Vanderborght, J., Vrugt, J. A., & Hopmans, J. W. (2008). On the value of soil moisture measurements in vadose zone hydrology: A review. *Water Resources Research*, *46*(4). https://doi.org/10.1029/2008WR006829

Vereecken, H., Huisman, J. A., Hendricks Franssen, H. J., Brüggemann, N., Bogena, H. R., Kollet, S., Javaux, M., Van Der Kruk, J., & Vanderborght, J. (2015). Soil hydrology:

Recent methodological advances, challenges, and perspectives. *Water Resources Research*, *51*(4). https://doi.org/10.1002/2014WR016852

Verseghy, D. L. (1991). Class—A Canadian land surface scheme for GCMS. I. Soil model. *International Journal of Climatology*, *11*(2). https://doi.org/10.1002/joc.3370110202

Wagener, T. (2003). Evaluation of catchment models. *Hydrological Processes*, *17*(16), 3375–3378. https://doi.org/10.1002/hyp.5158

Wagner, W., Lemoine, G., & Rott, H. (1999). A method for estimating soil moisture from ERS Scatterometer and soil data. *Remote Sensing of Environment*, *70*(2). https://doi.org/10.1016/S0034-4257(99)00036-X

Wagner, W., Noll, J., Borgeaud, M., & Rott, H. (1999). Monitoring soil moisture over the canadian prairies with the ERS scatterometer. *IEEE Transactions on Geoscience and Remote Sensing*, *37*(1 PART 1), 206–216. https://doi.org/10.1109/36.739155

Wakigari, S. A., & Leconte, R. (2022). Enhancing Spatial Resolution of SMAP Soil Moisture Products through Spatial Downscaling over a Large Watershed: A Case Study for the Susquehanna River Basin in the Northeastern United States. *Remote Sensing*, *14*(3), 776. https://doi.org/10.3390/rs14030776

Walker, J. P., Willgoose, G. R., & Kalma, J. D. (2001). One-dimensional soil moisture profile retrieval by assimilation of near-surface observations: A comparison of retrieval algorithms. *Advances in Water Resources*, *24*(6), 631–650. https://doi.org/10.1016/S0309-1708(00)00043-9

Wanders, N., Karssenberg, D., De Roo, A., De Jong, S. M., & Bierkens, M. F. P. (2014). The suitability of remotely sensed soil moisture for improving operational flood forecasting. *Hydrology and Earth System Sciences*. https://doi.org/10.5194/hess-18-2343-2014

Wang, H., Magagi, R., & Goita, K. (2017). Comparison of different polarimetric decompositions for soil moisture retrieval over vegetation covered agricultural area. *Remote Sensing of Environment*, *199*. https://doi.org/10.1016/j.rse.2017.07.008

Wang, J. R., & Choudhury, B. J. (1981). Remote sensing of soil moisture content over bare field at 1.4 GHz frequency. *Journal of Geophysical Research*, *86*(C6).

https://doi.org/10.1029/JC086iC06p05277

Wang, James R., & Schmugge, T. J. (1980). An empirical model for the complex dielectric permittivity of soils as a function of water content. *IEEE Transactions on Geoscience and Remote Sensing*, *GE-18*(4), 288–295. https://doi.org/10.1109/TGRS.1980.350304

Wang, L., & Qu, J. J. (2009). Satellite remote sensing applications for surface soil moisture monitoring: A review. *Frontiers of Earth Science in China*, *3*(2), 237–247. https://doi.org/10.1007/s11707-009-0023-7

Wang, Y., Peng, J., Song, X., Leng, P., Ludwig, R., & Loew, A. (2018). Surface soil moisture retrieval using optical/thermal infrared remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(9). https://doi.org/10.1109/TGRS.2018.2817370

Wei, Z., Meng, Y., Zhang, W., Peng, J., & Meng, L. (2019). Downscaling SMAP soil moisture estimation with gradient boosting decision tree regression over the Tibetan Plateau. *Remote Sensing of Environment*, *225*. https://doi.org/10.1016/j.rse.2019.02.022

Weidong, L., Baret, F., Xingfa, G., Qingxi, T., Lanfen, Z., & Bing, Z. (2002). Relating soil surface moisture to reflectance. *Remote Sensing of Environment*, *81*(2–3). https://doi.org/10.1016/S0034-4257(01)00347-9

Wen, F., Zhao, W., Wang, Q., & Sanchez, N. (2020). A Value-Consistent Method for Downscaling SMAP Passive Soil Moisture with MODIS Products Using Self-Adaptive Window. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(2). https://doi.org/10.1109/TGRS.2019.2941696

Wen, F., Zhao, W., & Wang, W. (2019). A Downscaling Scheme for Deriving Spatially Continuous Fine-Resolution Soil Moisture Data Based on Gap-Free Land Surface Temperature. *International Geoscience and Remote Sensing Symposium (IGARSS)*. https://doi.org/10.1109/IGARSS.2019.8899200

Wen, Y., Zhao, J., Zhu, G., Xu, R., & Yang, J. (2021). Evaluation of the rf-based downscaled smap and smos products using multi-source data over an alpine mountains basin, northwest china. *Water (Switzerland)*, *13*(20). https://doi.org/10.3390/w13202875

Western, A. W., Grayson, R. B., & Blöschl, G. (2002). Scaling of soil moisture: A hydrologic perspective. *Annual Review of Earth and Planetary Sciences*, *30*, 149–180. https://doi.org/10.1146/annurev.earth.30.091201.140434

William C., S. J., Joseph B., K., Jimy, D., David O., G., Dale M., B., Wang, W., & Jordan G., P. (2017). A Description of the Advanced Research WRF Version 2. *Building Energy Efficiency*, *46*(317), 56–67.

Wolfensberger, D., Gabella, M., Boscacci, M., Germann, U., & Berne, A. (2020). RainForest: A random forest algorithm for quantitative precipitation estimation over Swizerland. *Atmospheric Measurement Techniques Discussions*, *September*, 1–35. https://doi.org/10.5194/amt-2020-284

Wu, X., Walker, J. P., Rüdiger, C., Panciera, R., Monerris, A., & Das, N. N. (2011). Towards medium-resolution brightness temperature retrieval from active and passive microwave. *MODSIM 2011 - 19th International Congress on Modelling and Simulation - Sustaining Our Future: Understanding and Living with Uncertainty*.

Wu, Xiaoling, Walker, J. P., Rüdiger, C., & Panciera, R. (2015). Effect of land-cover type on the SMAP active/passive soil moisture downscaling algorithm performance. *IEEE Geoscience and Remote Sensing Letters*, *12*(4). https://doi.org/10.1109/LGRS.2014.2364049

Wu, Xiaoling, Walker, J. P., Rudiger, C., Panciera, R., & Gao, Y. (2016). Intercomparison of Alternate Soil Moisture Downscaling Algorithms Using Active-Passive Microwave Observations. *IEEE Geoscience and Remote Sensing Letters*, *14*(2). https://doi.org/10.1109/LGRS.2016.2633521

Wu, Xiaoling, Walker, J. P., Rüdiger, C., Panciera, R., & Gao, Y. (2017). Medium-Resolution Soil Moisture Retrieval Using the Bayesian Merging Method. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(11). https://doi.org/10.1109/TGRS.2017.2728808

Xing, M., He, B., Ni, X., Wang, J., An, G., Shang, J., & Huang, X. (2019). Retrieving surface soil moisture over wheat and soybean fields during growing season using modified water cloud model from RADARSAT-2 SAR data. *Remote Sensing*, *11*(16). https://doi.org/10.3390/rs11161956

Xu, T., & Liang, F. (2021). Machine learning for hydrologic sciences: An introductory overview. *Wiley Interdisciplinary Reviews: Water*, *8*(5), 1–29. https://doi.org/10.1002/wat2.1533

Ye, N., Walker, J. P., Guerschman, J., Ryu, D., & Gurney, R. J. (2015). Standing water effect on soil moisture retrieval from L-band passive microwave observations. *Remote Sensing of Environment*, *169*. https://doi.org/10.1016/j.rse.2015.08.013

Yee, M. S., Walker, J. P., Monerris, A., Rüdiger, C., & Jackson, T. J. (2016). On the identification of representative in situ soil moisture monitoring stations for the validation of SMAP soil moisture products in Australia. *Journal of Hydrology*, *537*. https://doi.org/10.1016/j.jhydrol.2016.03.060

Yin, J., Zhan, X., Liu, J., Moradkhani, H., Fang, L., & Walker, J. P. (2020). Near-real-time one-kilometre Soil Moisture Active Passive soil moisture data product. *Hydrological Processes*, *34*(21). https://doi.org/10.1002/hyp.13857

Young, R., Walker, J. P., Yeoh, N., Smith, A., Ellett, K., Merlin, O., & Western, A. W. (2008). *Soil moisture and meteorological observations from the Murrumbidgee catchment*. *0343778*, 1–55. http://www.oznet.org.au/documentation/

Zazueta, F., & Xin, J. (1994). *Soil Moisture Sensors| Netafim*. https://www.netafimindia.com/digital-farming/netbeat/Monitor/soil-moisture-sensors/

Zeng, C., Long, D., Shen, H., Wu, P., Cui, Y., & Hong, Y. (2018). A two-step framework for reconstructing remotely sensed land surface temperatures contaminated by cloud. *ISPRS Journal of Photogrammetry and Remote Sensing*, *141*. https://doi.org/10.1016/j.isprsjprs.2018.04.005

Zhan, X., Fang, L., Liu, J., Hain, C., Yin, J., Schull, M., Cosh, M., Wen, J., Lakhankar, T., Yang, K., & Walker, J. (2017). Fusing microwave and optical satellite observations for high resolution soil moisture data products. *International Geoscience and Remote Sensing Symposium (IGARSS)*, *2017-July*. https://doi.org/10.1109/IGARSS.2017.8127507

Zhan, X., Houser, P. R., Walker, J. P., & Crow, W. T. (2006). A method for retrieving high-resolution surface soil moisture from hydros L-band radiometer and radar observations. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(6). https://doi.org/10.1109/TGRS.2005.863319

Zhang, G., & Lu, Y. (2012). Bias-corrected random forests in regression. *Journal of Applied Statistics*, *39*(1). https://doi.org/10.1080/02664763.2011.578621

Zhao, L., Yang, K., Qin, J., & Chen, Y. (2013). Optimal exploitation of AMSR-E signals for improving soil moisture estimation through land data assimilation. *IEEE Transactions on Geoscience and Remote Sensing*, *51*(1). https://doi.org/10.1109/TGRS.2012.2198483

Zhao, W., & Duan, S. B. (2020). Reconstruction of daytime land surface temperatures under cloud-covered conditions using integrated MODIS/Terra land products and MSG geostationary satellite data. *Remote Sensing of Environment*, *247*. https://doi.org/10.1016/j.rse.2020.111931

Zhao, W., & Li, A. (2013). A downscaling method for improving the spatial resolution of AMSR-E derived soil moisture product based on MSG-SEVIRI data. *Remote Sensing*, *5*(12), 6790–6811. https://doi.org/10.3390/rs5126790

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

Zhao, W., Wen, F., Song, L., Li, X., & Li, A. (2019). Downscaling SMAP Passive Soil Moisture Product with MODIS Products over Mountainous Region. *International Geoscience and Remote Sensing Symposium (IGARSS)*. https://doi.org/10.1109/IGARSS.2019.8898591

Zheng, J., Lü, H., Crow, W. T., Zhao, T., Merlin, O., Rodriguez-Fernandez, N., Shi, J., Zhu, Y., Su, J., Kang, C. S., Wang, X., & Gou, Q. (2021). Soil moisture downscaling using multiple modes of the DISPATCH algorithm in a semi-humid/humid region. *International Journal of Applied Earth Observation and Geoinformation*, *104*, 102530. https://doi.org/10.1016/j.jag.2021.102530