UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Escola Tècnica Superior d'Enginyeria
de Telecomunicació de Barcelona

telecos
BCN

IMPROVING ROOT ZONE SOIL MOISTURE ESTIMATION FROM SURFACE REMOTE SENSING MEASUREMENTS

A Master's Thesis

Submitted to the Faculty of the

Escola Tècnica d'Enginyeria de Telecomunicació de Barcelona

Universitat Politècnica de Catalunya

by

GAO MENGYI

In partial fulfilment

of the requirements for the degree of

MASTER IN ADVANCED TELECOMMUNICATION TECHNOLOGIES

Advisor: Mercè Vall-llossera Ferran, Christoph Josef Herbert

Barcelona, October 2021

Title of the thesis:

Improving root zone soil moisture estimation from surface remote sensing measurements.

Author:

GAO, MENGYI

Advisor:

Mercè Vall-llossera Ferran
Christoph Josef Herbert

Abstract

Satellite soil moisture provides a wider range of spatial data than in-situ observations. Satellites are sensitive to measuring soil moisture in a few centimetres of the soil layer. However, root zone soil moisture is more important, especially in vegetated areas. Therefore, it is necessary to invert the near-surface soil moisture collected by satellite missions to estimate root zone soil moisture. This study quantified the association between near-surface and root zone soil moisture using in-situ data from the REMEDHUS network in Spain and Soil Moisture Ocean Salinity (SMOS) mission from 2014-2018, comparing SMOS with in-situ observations based on an exponential filter approach, optimizing an optimal characteristic time length (T_{opt}) by computing the correlation coefficients (R) of the filtered satellite time series with in-situ time series. The results show that the exponential filter's main factor T is sensitive to soil moisture and varies depending on the climate. When using this method to infer root zone soil moisture, the eight stations' average R-value reached 0.71, meeting the "Strong correlation" standard. The root zone soil moisture obtained by SMOS surface inversion is compared with in situ soil moisture. The mean R-value of soil moisture in the root zone based on SMOS is 0.58, which is approximate to the "baseline error (0.6)" -- the accuracy of the surface soil moisture estimated by SMOS. However, attention should be paid to the effects of soil moisture season and soil texture on the experimental results.

Acknowledgements

Firstly, I would like to thank both my advisors, Mercè Vall-llossera Ferran and Christoph Josef Herbert, without them I would not have done this work. Thank them for introducing me to this research topic and their constant assistance during these past eight months.

Secondly, I want to thank my family. They still give me support and encouragement, although they are not around. Of course, I am also grateful to the teachers, friends and classmates I met in Barcelona. They studied and lived with me, allowing me to gain happiness and warmth in this city far from home.

Most importantly, thank yourself. The virus broke the original study plan, but you still persevere. Thank you for the growth that this experience has given, academically or in life. Your initial intention to come to UPC was to explore more possibilities in your major. The change makes you feel anxious and frustrated when you just came here. But fortunately, you have overcome them one by one and came here.

将军已下马，各自奔前程。

Revision history and approval record

Revision	Date	Purpose

Written by:		Reviewed and approved by:	
Date		Date	
Name		Name	
Position		Position	

Table of content

List of Figures

List of Tables

1. Introduction

1.1. Motivation and Goals

Soil moisture is an essential climate variable (ECV) and of great importance in processes at the atmosphere-vegetation-soil boundary. Previous studies point out that soil moisture influences land-atmosphere interactions by modifying the boundary layer's energy moisture fluxes (Basara et al., 2002; Taylor et al., 2007). Koster (2004) demonstrated that the variation of the soil moisture could predict drought in semi-arid environments such as North America Great Plains. Frye and Mote (2010) found that under weather conditions that are not conducive to convection, soil moisture and soil moisture gradients significantly influence convective initiation. Taylor and his colleagues (2012) found that the convective precipitation in the Sahel region of Africa in the afternoon preferentially falls on the drier soil, which is probably caused by the abnormally low soil moisture.

Soil moisture can also reflect the water demand of crops, which has important guiding significance for agricultural production. Soil textures influence water storage and soil moisture content. Therefore, soil moisture affects the dissolution and transfer of nutrients in the soil and the activities of microorganisms. The amount of soil moisture is related to the growth of plants and crop yield. Timely monitoring is helpful to grasp the diverse requirements of plants on soil water content to provide suggestions for various agricultural measures such as tillage, irrigation and fertilization, which will directly affect the growth of plants and crops.

It can be seen that the measurement of soil moisture plays an indispensable role in human production and the natural environment. However, in some cases, such as drought monitoring and agricultural modelling, the estimate of root zone soil moisture is essential because the root zone is the reservoir of available water for plants. One of the methods used to estimate root zone soil moisture is based on the Soil Moisture Index (SWI), an exponential filter developed by Wagner (1999) and improved by Albergel (2008). The method uses only one parameter T (timescale of the soil moisture variation) to estimate soil moisture from surface observations and studies the delay between surface and root zone soil moisture. In addition, Gill M K et al. (2006) used support vector

machines to make the soil moisture prediction. After that, Carranza C et al. (2021) used the random forest method to estimate the root zone soil moisture.

In recent years, with the development of satellite missions, many researchers have begun to measure soil moisture through satellite data. Compared with in-situ soil moisture monitoring, satellite missions have the advantages of all-weather, wide coverage area and convenient operation. For example, Brocca L et al. (2010) improved runoff prediction by assimilating the ASCAT (Advanced Scatterometer, deployed on Meteorological Operational satellite) soil moisture product. Wagner W et al. (1999) used European Remote Sensing Satellites (ERS-1 and ERS-2) data monitoring soil moisture over the Canadian Prairies. Jackson et al. (2011) inverted in-situ soil moisture values and SMOS data. The results show that SMOS estimates are similar to in-situ observations. Sanchez et al. (2012) inverted that SMOS can detect temporal anomalies and the temporal evolution of surface soil moisture.

Soil moisture products are derived from microwave remote sensing missions, which need to be verified through in-situ soil moisture observations (Rüdiger et al., 2009). Many in-situ soil moisture networks can be operated at present. The relevant information can be found on the website of the International Soil Moisture Network (ISMN). Such as the Goulburn River experimental catchment area in Australia (Rüdiger et al., 2007) or SMOSMANIA in southwestern France (Soil Moisture Observing System – Meteorological Automatic Network Integrated Application, Calvet et al., 2007; Albergel et al., 2008).

In this study, REMEDHUS network soil moisture data will be used as a network to compare satellite and in-situ time series, using soil moisture index model and exponential filtering to examine the fluctuation of main parameter T, including its dependence on space, depth, soil moisture season and soil conditions. Thus, improving root zone soil moisture estimation from surface remote sensing measurements. Finally, the evaluate the accuracy of root zone soil moisture predicted from SMOS surface soil moisture based on exponential filter and confirmed the feasibility of SMOS mission for root zone soil moisture prediction.

1.2. Thesis Outline

This master's thesis consists of six chapters. The work is organized as follows:

Chapter 2 introduces the surface and root zone soil moisture.

Chapter 3 accounts material and methodology of this study.

Chapter 4 presents the most relevant obtained results for this study.

Chapter 5 discusses the result along with suggesting future work and lines of research.

Chapter 6 provides the main conclusion drawn from this thesis.

2. Surface and Root Zone Soil Moisture: in-situ and satellite measurement

2.1. Soil Moisture Measurement

The methods of soil moisture measurement can be divided into non-invasive geophysical measurement and invasive in-situ measurement. Non-invasive geophysical measurement is represented by airborne and spaceborne remote sensing technologies, which uses remote sensing to obtain soil moisture data from the air with the Ground Penetrating Radar (GPR) and Electromagnetic Induction (EMI). Spaceborne remote sensing techniques for soil moisture are less labour-intense, which cannot be achieved by in-situ measurements (Jackson, 2002). Remote sensing also has the advantage of large coverage areas (Ochsner et al., 2013), providing information about soil moisture temporal and spatial variability. However, the non-invasive geophysical measurement has a limited investigation depth, making it difficult to get moisture data from the deep soil layer. Moreover, satellite missions have fixed operation periods, which makes data exist collection intervals. Another soil moisture measurement (invasive in-situ measurement) has the advantage is that soil moisture can be continuously monitored through point-scale multi-depth measurements, with better time resolution and real-time uninterrupted data acquisition. This method can obtain direct soil moisture data at different depths by the probes. However, it is undeniable that the limited station selection and the cost of labour force limit the universality of this method. Figure 1 is the diagrammatic drawing of this method, which generally requires the selection of observation areas and stations, and the installation of detection devices in the soil to collect soil moisture data of the network.

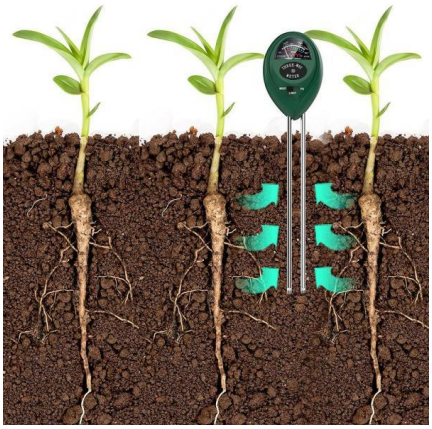


Figure 1 Diagrammatic drawing of invasive in-situ measurement.

Mission	Temporal Resolution	Spatial Resolution (km)	Type	EMR-Band
SSM/I	Daily	25	Passive	C
TRMM TM	Daily	50-56	Passive	C
Aqua AMSR-E	Daily	56	Passive	C
ERS 1-2 SCAT	35 days	25-50	Active	C
SMOS	3 days	50	Passive	L
SMAP	2-3 days	10-40	Both	L
MetOp ASCAT	29 days	50	Active	C

Table 1 Synopsis of recent and future satellite soil moisture missions.

Many efforts have been made by researchers to get ideal measurement data from satellites. Thirty-five years ago, Schmugge et al. (1986) found that the L-band frequency range was identified as the best choice for soil moisture retrieval using microwave radiometers. The first spaceborne mission that mainly focused on collecting soil moisture data was the Soil Moisture Ocean Salinity (SMOS) satellite, which was initiated by the European Space Agency (ESA) in 2009 and used microwave radiometry to estimate soil moisture at global coverage (Kerr et al., 2010). The Y-shaped MIRAS (Microwave Imaging Radiometer by Aperture Synthesis) instrument on SMOS can be clearly identified in the left picture of Figure 2 left. Soil Moisture Active Passive (SMAP) is the second spaceborne mission to monitor soil moisture (Figure 2 right). The SMAP mission uses L-band radar and radiometer instruments sharing a rotating 6-meter mesh reflector antenna to provide high-resolution and high-precision global maps of soil moisture and freeze-thaw every 2 or 3 days. It was launched by the National Aeronautics and Space Administration (NASA), but the active sensor failed after three months of operation. Table 1 summarizes the main satellite mission

information used for soil moisture. It can be seen that most missions use the C-band (4-8 GHz), only SMOS and SMAP use the L-band (1-2 GHz).

Figure 2 3D render of SMOS satellite (left) and SMAP satellite (right).

2.2. Root Zone Soil Moisture

Soil depth is the most direct factor affecting soil moisture. In general, the 0-100cm soil layer is divided into the following three characteristic layers. The topmost layer is a variable layer (0-20cm), which is usually the surface soil of the ground. This part of the soil is greatly affected by meteorological conditions and agricultural technical measures. The content depends on many other factors such as soil texture, season, layering, topography vegetation, etc. The soil moisture in this layer has a significant variation, and the coefficient of variation is also large. The middle part is the layer with slowly changing (20-50cm). This layer is less affected by meteorological conditions and agricultural technical measures, and the water coefficient of variation is also small. But it has a more significant impact on the water supply for crop growth. The deepest layer is homogeneous (50-100cm). Soil moisture in this layer is in a relatively stable state.

Due to the main percentage of the plants' roots being usually located in 20-100 cm, this study used the soil located in this depth interval as root zone soil based on the actual situation in the study area. The soil moisture in this part is more stable, and the water exchange and absorption of roots and soil are more active, which directly affects the plant's growth.

Numerous studies have also proved that it is possible to estimate root zone soil moisture based on

near-surface soil moisture, and most studies will also use surface soil moisture to explore root soil moisture. Georgakakos and Bae (1994) evaluated soil moisture variability in the Midwestern United States in their research. They found the stability of root zone soil moisture is much better than near-surface. Mahmood and Hubbard (2007) indicated that the correlation between surface and root zone soil moisture shows high variability due to differences in soil texture, land use, and climate conditions. Mahmoud and his colleagues (2012) also studied the predictability of soil moisture at different depths. They found that the soil moisture in the root zone can be estimated with observations of 10 cm near the surface, but this is also affected by the climate conditions. Wu and Dickinson (2004) and some researchers examined the variability in soil moisture through the National Center for Atmospheric Research Community Climate Model. They found that seasons affect the correlation between near-surface and root zone soil moisture.

3. Material and Methodology

3.1. Study Area

This study uses the REMEDHUS network which set up by Centro Hispano Luso de Investigaciones Agrarias (CIALE) Universidad de Salamanca. It was deployed in 2005 and has been in use ever since. REMEDHUS area is located in Spain's Duero Basin, a semi-arid flat agricultural area with about 1,300 square kilometres between 41.1-41.5°N and 5.1-5.7°W, with an average elevation of about 800 meters. Statistics show that 80% of the area is winter wheat and barley cereals, 12% is forest pastures, 5% and 3% is irrigated crops and vineyards. The soil textures are mainly sandy loam and sandy clay loam. This study selects five years from 2014 to 2018 as the research period. Various data used in this study will be introduced as follows.

3.1.1. REMEDHUS soil moisture

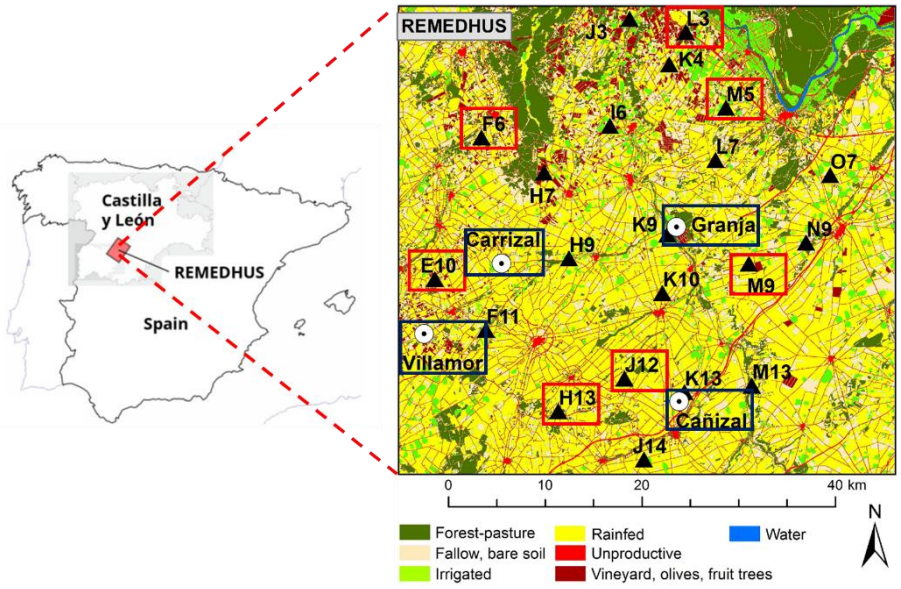


Figure 3 Location of the REMEDHUS network.

The REMEDHUS network has 24 automated stations equipped with capacitance probes (Hydra Probes, Stevens Water monitoring System, Inc.) and environmental probes (Sentek Pty). This study selected 8 (E10, F6, H13, J12, L3, M5, M9 and CAR) as research stations. The location of each station is present in Figure 3. The capacitance probe measures the soil moisture at 5 cm, 25

cm, 50 cm and 100 cm per hour, with a nominal accuracy of 0.01 m^3m^{-3}. REMEDHUS is part of the ISMN International Soil Moisture Network, and it has been used in many validation studies of remote sensing soil moisture products (Sánchez et al., 2012; González Res-Zamora et al., 2015) includes SWI research (Ceballos et al., 2005; Paulik et al., 2014). Furthermore, due to the different soil volume monitored by the two probe types, a weighted average of the 0–50 cm soil moisture was also calculated (Martínez-Fernández et al., 2015):

$$SM_{0-50cm} = \frac{SM_{5cm}}{5} + \frac{2SM_{25cm}}{5} + \frac{2SM_{50cm}}{5} \quad (1)$$

Therefore, four in-situ time series from the eight stations resulted for 5, 25 and 50 cm soil depths, together with the weighted 0–50 cm. Soil water parameters: field capacity (θ_{FC}), wilting point (θ_{WP}) and total water capacity (θ_{TWC}) of each station and soil layer (Table 2) were obtained from each station's water holding retention curve. The retention curves were estimated using sandboxes and pressure membrane by applying the Van Genuchten (1980) method, measuring the soil moisture contents at nine soil–water potential values (from 0 to 1500 kPa).

Station	Land use	Depth (cm)	Texture	θ_{FC}	θ_{WP}	θ_{TWC}
E10	Vineyard	5	Sandy loam	0.088	0.028	0.410
		25	Sandy loam	0.108	0.047	0.367
		50	Sandy clay loam	0.193	0.099	0.397
F6	Vineyard	5	Sandy loam	0.229	0.111	0.324
		25	Sandy clay loam	0.207	0.113	0.294
		50	Sandy loam	0.108	0.063	0.347
H13	Forest-Pasture	5	Sandy loam	0.158	0.075	0.424
		25	Sandy loam	0.138	0.071	0.446
		50	Sandy loam	0.113	0.076	0.447
J12	Rainfed	5	Sandy clay loam	0.236	0.096	0.483
		25	Sandy clay loam	0.228	0.113	0.456
		50	Sandy clay loam	0.265	0.168	0.415
L3	Vineyard	5	Loamy sand	0.125	0.04	0.427
		25	Loamy sand	0.146	0.056	0.348
		50	Loamy sand	0.130	0.043	0.370
M5	Rainfed	5	Loamy sand	0.100	0.057	0.357
		25	Loamy sand	0.125	0.042	0.406
		50	Loamy sand	0.071	0.043	0.507
M9	Rainfed	5	Sandy clay loam	0.226	0.137	0.519
		25	Sandy clay loam	0.238	0.124	0.527
		50	Loam	0.214	0.146	0.508

CAR	Rainfed	5	Loam	0.256	0.137	0.505
		25	Sandy clay loam	0.239	0.127	0.515
		50	Sandy clay loam	0.218	0.109	0.500

Table 2 Land use, texture, field capacity (θ_{FC}), wilting point (θ_{WP}) and total water capacity (θ_{TWC}) at the different depths of each station used in this study.

3.1.2. REMEDHUS climate

The REMEDHUS region has a semi-arid continental climate with an average annual temperature of 12°C. The annual precipitation in the past ten years is 385±100 mm. However, the intensity and time of weather conditions vary during the year. Climate factors combined with the growing and dormant periods can define seasons with similar soil moisture recharge and utilization processes. In this study, the estimation of the soil moisture season is based on the relationship between PET (potential evapotranspiration) and precipitation. The comparison results of PET and precipitation are given in section 4.3.1. PET is calculated based on the Penman-Monteith equation as a reference for the direct estimation of actual crop evapotranspiration. It is derived from solar irradiance, relative humidity, average temperature and wind speed. The average PET values were calculated by four climate stations (Villamor, Granja, Canizal and Carrizal). The location of each station is present in Figure 3. Figure 4 and Figure 5 show the annual changes in PET and precipitation observed at four stations in 2017-2018. This method has been used as a standard method by the United Nations Food and Agriculture Organization.

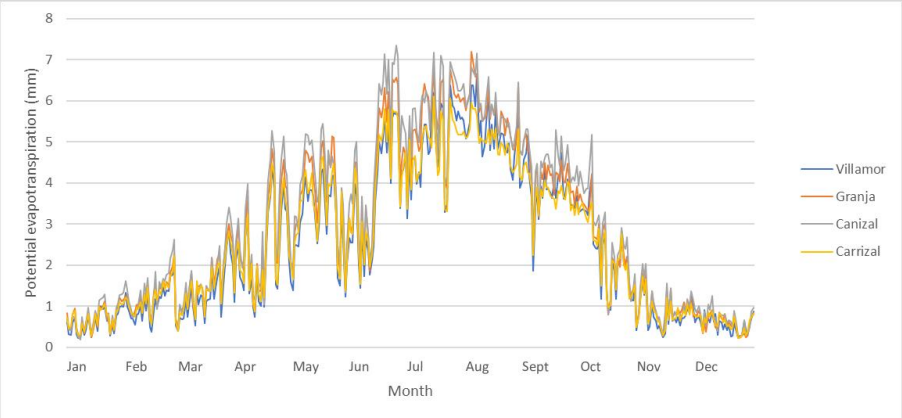


Figure 4 Potential evapotranspiration (PET) change in REMEDHUS network from 2017 to 2018.

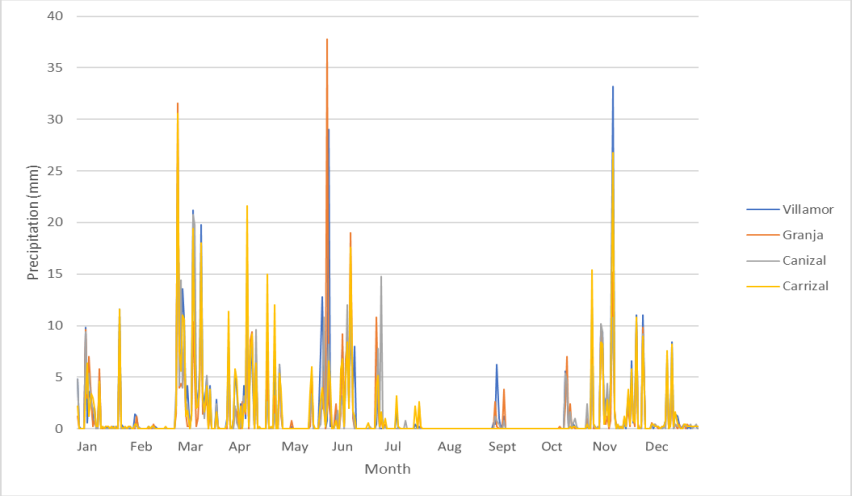


Figure 5 Precipitation changes in REMEDHUS network from 2017 to 2018.

3.1.3. REMEDHUS location

Table 3 details the longitude and latitude of each station and is used to match the stations with the SMOS data to analyze soil moisture.

Station	Latitude	Longitude
E10	41.27473	-5.5919
F6	41.37338	-5.5485
H13	41.18264	-5.47708
J12	41.20548	-5.41558
L3	41.44649	-5.3587
M5	41.39392	-5.32146
CAR	41.28671	-5.52843

Table 3 Latitude and longitude of each station used in this study.

3.2. SMOS Soil Moisture

According to the measurement, application and processing degree of SMOS, several products listed in Table 4 are available to use.

Level of data	Data introduce
L0	SMOS Payload data in their original format (CCSDS packets)
L1A	Reformatted and calibrated observation and housekeeping data in engineering units.
L1B	The output of the image reconstruction of the SMOS. It consist of Fourier components of brightness temperatures in the antenna polarisation reference frame.
L1C	Multi-incidence angle brightness temperatures at the top of the atmosphere, geolocated in an equal-area grid system.

L2	Soil Moisture Near Real Time Neural Network data
L3	Sea-ice product
L4	Downscaled versions of the original resolution to a 1km grid.

Table 4 Synopsis of SMOS data at all levels

Barcelona Expert Center (BEC) regularly distributes its Soil Moisture and Salinity research products on its web page (http://bec.icm.csic.es). Including high-resolution Soil Moisture maps from the Iberian Peninsula at 1 Km and L3 and L4 global salinity maps at several space resolutions. This study used the 1 km high-resolution SMOS L4 soil moisture product (2014-2018) researched and produced by the BEC. It is obtained by merging microwave and visible infrared sensor data with different spatial resolutions and modelling data through a linear downscaling algorithm. The relationship between SMOS L3 soil moisture products and auxiliary data is derived from the so-called universal triangle that uses adaptive moving Windows to ensure similarity of climatic conditions. Auxiliary data includes SMOS L1C brightness temperature, MODIS NDVI and European Center for Medium-Range Weather Forecast (ECMWF) surface temperature. Some studies have indicated a slight difference between the SMOS brightness temperature measured by the ascending and descending orbits (Martín-Neira et al., 2016). In this study, ascending orbits data were used.

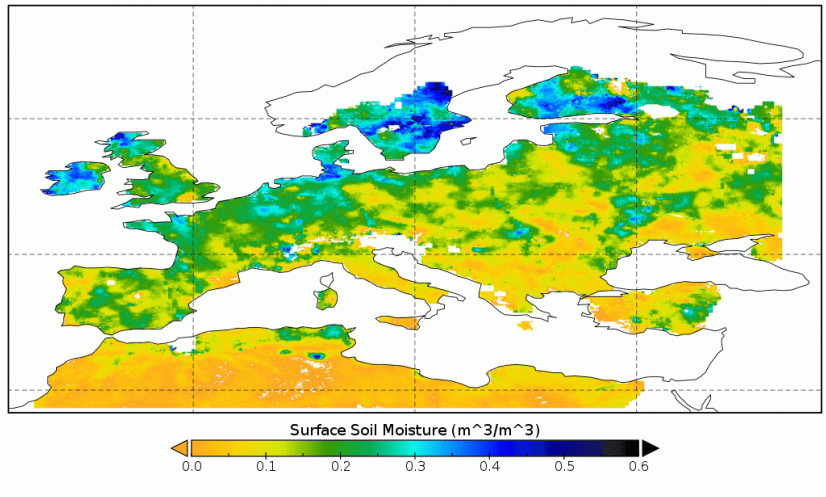


Figure 6 1 km high-resolution SMOS L4 soil moisture product researched and produced by the BEC (Barcelona Expert Center).

3.3. Calculation of Soil Water Index

It is mentioned in chapter 2 that root zone soil moisture is related to soil moisture near the surface. To further explore the connection between them, Albergel et al. (2008) used a recursive exponential filter to predict root zone soil moisture from near-surface observations. There is a key parameter T in the model, which can be interpreted as the characteristic time length. It is related to the diffusivity of water. Sometimes, it represents the wetting front after a rain event. The experimental accuracy of predicting root zone soil moisture will change with the T value. There is an optimal T-value (T_{opt}) for each station, which has the highest prediction accuracy. Besides, different metrics are needed to evaluate the T-value and lock the T_{opt}. In the whole process, the recursive formulation is used to predict the soil wetness index (SWI) and standardizes volumetric soil water content by the minimum and maximum values attained over the entire period of record at each location. SWI is calculated using the following recursive formula. (The filter is initialized to $SWI_{(1)} = SM(t_1)$ and $K_1 = 1$.)

$$SWI_n = SWI_{(n-1)} + K_n(SM(t_n) - SWI_{(n-1)}) \quad (2)$$

Where $SWI_{(n-1)}$ is the estimated value of predicted root zone soil moisture at t_{n-1}, $SM(t_n)$ is the estimated value of surface soil moisture at t n, and the gain K_n at t_n can be obtained by

$$K_n = \frac{K_{n-1}}{K_{n-1} + e^{-\frac{t_n - t_{n-1}}{T}}} \quad (3)$$

Where T is the key parameter required in this study represents the timescale of the soil moisture variation (in days), this parameter can be interpreted as the characteristic time length at each station and depth, which increases with the depth and decreases with the soil diffusion coefficient constant. There are three methods for obtaining the optimal T value (T_{opt}). The first is compares remotely sensed SWI with in-situ soil moisture measurements at different depths. The second is to compare remotely sensed SWI with simulated soil moisture data at different depths. The third is to compare the SWI from the surface in-situ soil moisture measurement with the deep layer in-situ soil moisture measurement. This research belongs to the last group. Alberger (2008) showed that each research could be found an optimal T, characterized by the highest prediction accuracy

evaluated based on the Nash-Sutcliffe (NS) score. In this study, a variety of metrics were used to conduct a comprehensive exploration of T_{opt}. Because of the maximum period of the seasons, the range is from 1 to 120 days. T_{opt} was determined using these metrics to evaluate the comparison between the in-situ observations at the 25cm, 50cm, and 0-50cm layers at eight stations and the SWI calculated using surface soil moisture. The SWI calculated according to the in-situ surface soil moisture measurement is compared to soil moisture 25cm, 50cm and 0-50cm obtaining the different metrics for each T used. The T_{opt} is selected based on the better metric in the different cases.

3.4. Metric of Soil Moisture

This study will use correlation coefficient (R), root mean square difference (RMSD), centred root mean square difference (cRMSD), Bias and Nash–Sutcliffe (NS) score as the validation metrics to select T_{opt}. For R and NS, T_{opt} corresponds to the highest value. For RMSD and cRMSD, T_{opt} corresponds to the lowest value. For the Bias, T opt corresponds to the deviation value closest to zero. Finally, the most appropriate T_{opt} estimation metric method will be selected and applied in the subsequent research according to the obtained results. Each of these metrics is described below.

3.4.1. Correlation coefficient (R)

The correlation coefficient is a statistical index used to reflect a measure of linear correlation between two sets of data, generally represented by *r*. The correlation coefficient can be defined in various ways, and the most commonly used Pearson correlation coefficient is used in this study.

Pearson correlation coefficient between two variables is defined as the quotient of covariance and standard deviation.

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_X)]}{\sigma_X \sigma_Y} \quad (4)$$

The above formula defines the overall correlation coefficient, usually represented by the Greek

lowercase letter ρ. Pearson correlation coefficient can be obtained by estimating the covariance and standard deviation of samples, which is generally defined by the *r*:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\text{samples,}} \tag{5}$$

In general, the correlation strength of variables can be judged by the following value range (0< *r* <1) :

0.8-1.0: Exceeding strong correlation

0.6-0.8: Strong correlation

0.4-0.6: Medium correlation

0.2-0.4: Weak correlation

0.0-0.2: Exceeding weak correlation or no correlation

3.4.2. Root mean square difference (RMSD)

RMSD is the root-mean-square deviation between filtered satellite observations and the corresponding in-situ measurement, which can be used to evaluate the similarity between two samples and can be expressed as:

$$RMSD = \sqrt{\frac{\sum_{t=1}^{T}(\hat{y}_t - y_t)^2}{T}} \tag{6}$$

RMSD is very sensitive to very large or minor errors in a set of measurements, so the Root mean square error is a good indicator of measurement precision. cRMSD (Centered Root Mean Square Difference) has the similar function and computational process as RMSD. It is also used to measure the deviation between the observed value and the actual value. The smaller the result, the more similar the samples.

3.4.3. Bias

Bias is a measure of the dispersion of the sample distribution. It is a non-symmetric error caused

by a systematic deviation. Generally used to measure the degree of deviation of the sample value from the arithmetic mean, representing the degree of agreement between the results of multiple determinations of a sample. The smaller the deviation, the less the sample deviates from the average. Bias is generally expressed as:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2} \tag{7}$$

Where N is the sample amount, and μ represents the mean value of the sample population X. The closer the deviation value is to 0, the higher the precision of the measurement result.

3.4.4. Nash–Sutcliffe (NS) score

NS Score is often used to quantify the prediction accuracy of simulation models (such as hydrological models), which can be expressed as

$$NSE = 1 - \frac{\sum(y_i - y_i^{pred})^2}{\sum(y_i - \bar{y})^2} \tag{8}$$

Where, y_i^{pred} is the predicted value of variables by the model. For a perfect model, the variance of the estimated error is 0, then NSE=1. Conversely, if the variance of the estimated error generated by a model is equal to the variance of the observed value, then NSE=0. When the variance of the estimation error obtained by the prediction model is significantly greater than that of the observed value, NSE< 0. Therefore, the value of NSE ranges from $(-\infty, 1]$. When the NSE value is closer to 1, the model's prediction ability is better.

3.5. Sample Cross-Correlation

The cross-correlation function measures the similarity between a time series and lagged versions of another time series as a function of the lag.

$$c_{y_1y_2(k)} = \begin{cases} \frac{1}{T}\sum_{t=1}^{T-k}(y_{1t} - \bar{y}_1)(y_{2,t+k} - \bar{y}_2); k = 0, 1, 2, \ldots \\ \frac{1}{T}\sum_{t=1}^{T+k}(y_{2t} - \bar{y}_2)(y_{1,t-k} - \bar{y}_1); k = 0, -1, -2, \ldots \end{cases} \quad (9)$$

where y_1 and y_2 are the sample means of the series, the sample standard deviations of the series are:

$$s_{y_1} = \sqrt{c_{y_1y_1(0)}}, \text{ where } c_{y_1y_1(0)} = Var_{(y_1)}. \quad (10)$$

$$s_{y_2} = \sqrt{c_{y_2y_2(0)}}, \text{ where } c_{y_2y_2(0)} = Var_{(y_2)}. \quad (11)$$

An estimate of the cross-correlation is

$$r_{y_1y_1(k)} = \frac{c_{y_1y_2(k)}}{s_{y_1}s_{y_2}}; k = 0, \pm 1, \pm 2, \ldots \quad (12)$$

The ordinary correlation coefficient first aligns the two sequences, then subtracts the average value of the samples in the same position from the average value of the respective sequences, then multiplies the difference, and finally sums it according to the number of samples. Cross-correlation is to stagger the two series by *k* positions to perform the above steps. *T* is the length of the series. After subtracting the staggered *k* positions, only *T* - *k* elements are left in a one-to-one correspondence. Now, *k* represents the lag. The difference between the correlation coefficient and the ordinary correlation coefficient is reflected in the numerator of the formula. The two series are not calculated for the corresponding elements but shifted, and the shifted overlapping elements are combined and calculated.

3.6. Coefficient of Variation

The coefficient of variation, like range, standard deviation and variance, is the absolute value reflecting the degree of data dispersion. When it is necessary to compare the dispersion degree of soil moisture data of two stations, if the measurement scale difference between two series of data is too significant, it is not appropriate to directly use standard deviation for comparison. The coefficient of variation is affected by the degree of dispersion of variable values and the average level of varying values. It is defined as the ratio of standard deviation to mean:

$$CV = \frac{\sigma}{\mu} \tag{13}$$

The advantage of the coefficient of variation over standard deviation is that it does not need to refer to the mean of the data. The coefficient of variation is a dimensionless quantity, so when comparing two series of data with different dimensions or different mean values, the coefficient of variation rather than the standard deviation should be used as a reference for comparison. However, when the average value is close to 0, small perturbations will also have a massive impact on the coefficient of variation, resulting in poor accuracy.

3.7. Soil Moisture Season

SM Season	Period (In general)	Criterion	Prevailing Processes	SM Condition
Recharge	November-February	Precipitation>PET; initial plant growth	Precipitation	SM storage increases
Utilization	February - Mid June	Precipitation<PET; main growing season	Strong root-water uptake and evapotranspiration	SM decreases due to consumption
Deficit	Mid June-October	Precipitation<<PET; crops are harvested	Evaporation at maximum	Continuous drying; SM at minimum in the end

Table 5 Estimated soil moisture seasons from 5-year, and the typical growing season within REMEDHUS, including the prevailing processes of SM change.

The soil moisture season divides the year into three seasons: recharge, utilization and deficit. It is defined by the relationship between monthly accumulated precipitation and potential evapotranspiration (PET). Recharge begins when the cumulative precipitation is above PET and roughly coincides with crop harvest dates (around November). Although the amount of evaporation is less in winter due to the climate, the soil moisture starts to recharge and has been used by early plant development and transpiration. The utilization period begins when the PET value exceeds the precipitation (usually in spring). The amount of water available for precipitation cannot make up for the loss caused by plant water evapotranspiration. At this time, the vegetation mainly uses stored groundwater. If the rainfall decreases and the plant growing season continues to consume water, the soil will begin to dry out at this time. The boundary between utilization and deficit is defined as approximately mid-June, when plants' soil moisture in the root zone is

intensively used, with inter-annual variability. When water becomes scarce in the autumn, crops are harvested quickly. The summer is usually too dry, with insufficient moisture in the underground soil, to grow crops on a pure rain basis. Table 5 summarizes the main characteristic information of the soil moisture season.

After the series is divided according to soil moisture season, the original complete series would be decomposed into three main categories of soil moisture season (recharge, utilization and deficit). Each year will contain these three stages, but this will cause the same soil moisture seasons of different years to be discontinuous with each other, so T_{opt} estimation cannot be carried out. There are two methods to solve this problem, numerical average and time series splicing.

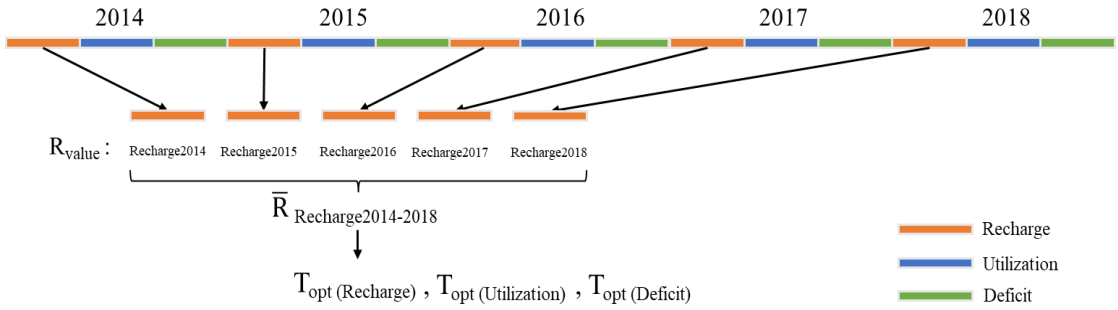


Figure 7 Diagram of numerical average.

As the Figure 7, the numerical average method is to use metric standard R to calculate each soil moisture season of each year for (15 soil moisture seasons in 5 years), and then classify the results according to different seasons and take average value to generate, and finally, select T_{opt}.

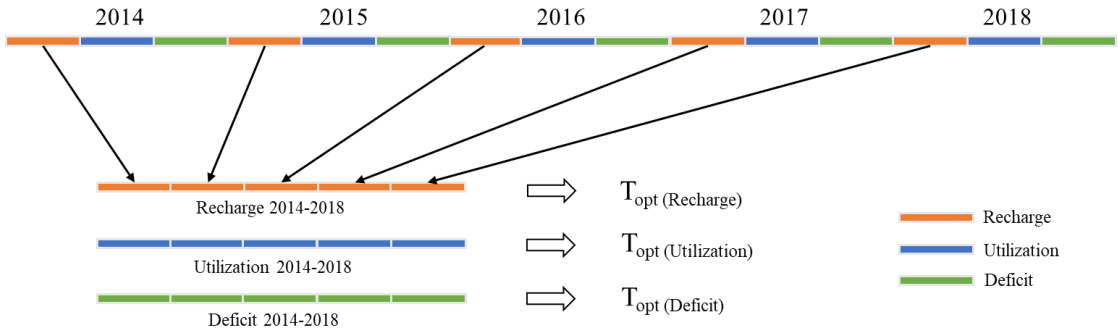


Figure 8 Diagram of time series splicing.

The time series splicing method splices five time series of the same season into a long time series (Figure 8). This new time series contains all the sub-series of the same season in different years of a single station. Then, estimating T_{opt} for the long series.

3.8. Root Zone Soil Moisture Estimation

Figure 9 shows the research flow in this part. In general, it is divided into three steps. First, the near-surface error is calculated by using in-situ observation data and SMOS. One of the original goals of the SMOS mission was to collect surface soil moisture data, so this process provides "baseline error". Second, SMOS root zone SWI and in-situ root zone SWI were used to calculate the matching accuracy. In this process, the previous phase T_{opt} value will be used (obtained by SWI from the surface in-situ soil moisture measurement with the root zone layer in-situ soil moisture measurement). Finally, the two precision values are compared. If the final results are similar to the "baseline error", it proves that using SMOS data to estimate root zone soil moisture is feasible.

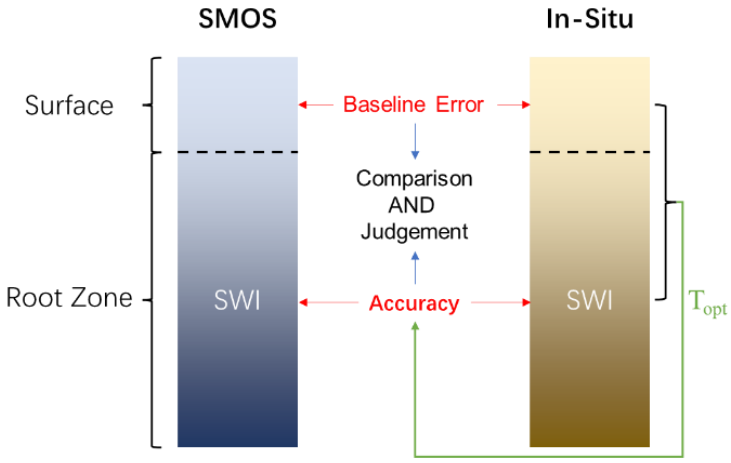


Figure 9 Estimating root zone soil moisture using SMOS.

4. Results

4.1. Depth Effect on Soil Moisture

This part will analyze the influence of the depth of in-situ measurement on soil moisture. It will explore the difference in soil moisture at depths of 5cm, 25cm and 50cm. Stations L3, M5, J12 and H13 will be picked as example stations.

4.1.1. SMOS and in-situ soil moisture

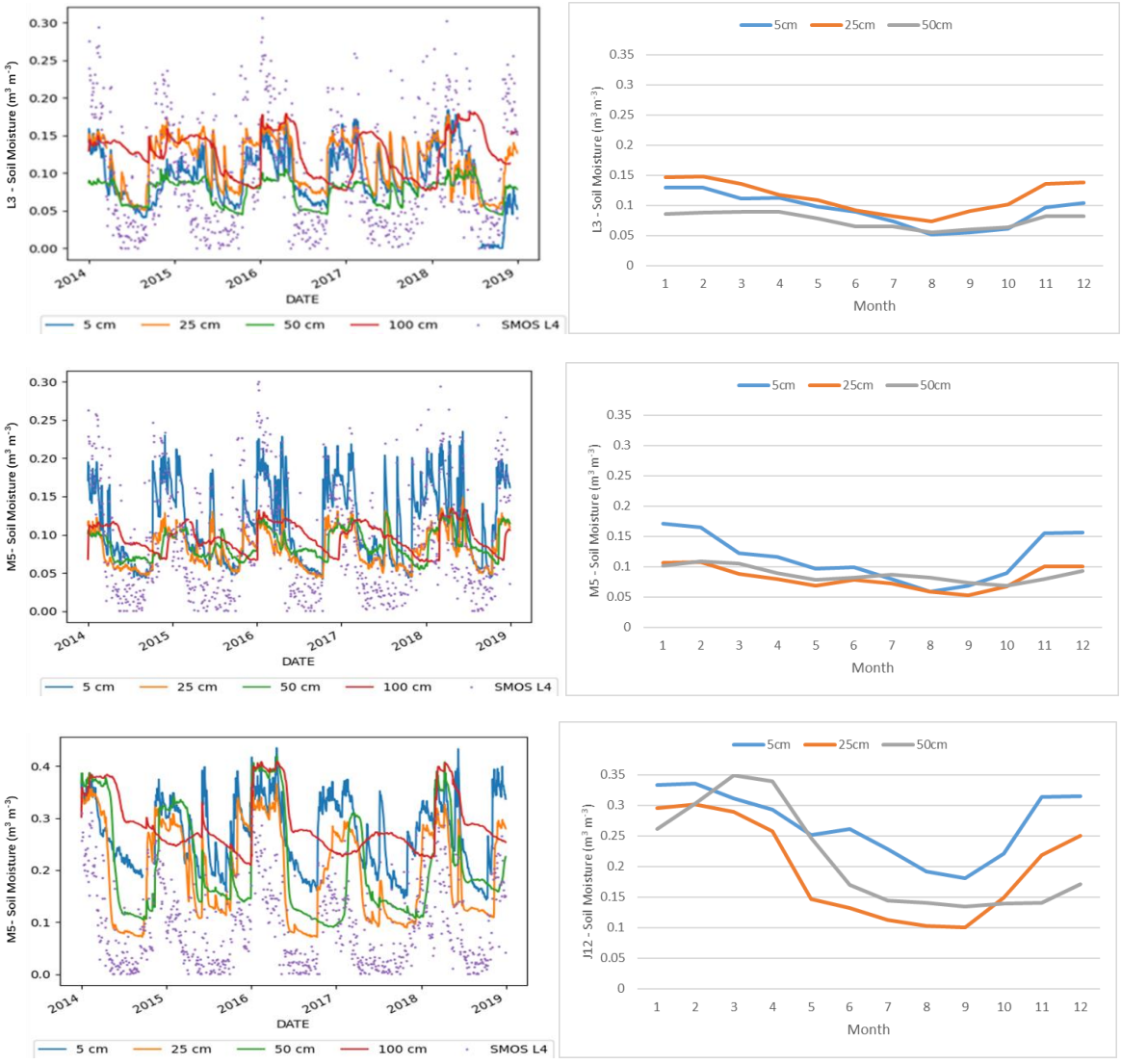


Figure 10 In situ soil moisture time series and SMOS time series at L3, M5 and J12 from 2014 to 2018 (left), and in situ monthly average soil moisture for each station (right).

In-situ soil moisture data were presented together with SMOS mission measurements (Figure 10 left). It can be seen that the SMOS series has a more extensive dynamic range than the in-situ soil moisture. For the time series of the three stations in-situ measurements, as expected, the depth of 5 cm has a higher dynamic range, which is more evident at the M5 and L3 stations. The deeper the measurement depth, the more limited and smoother the fluctuation range of the curve.

Station	Depth	Average	Maximum	Minimum	Range	CV
J12	5cm	0.303	0.440	0.194	0.245	0.203
	25cm	0.158	0.276	0.096	0.179	0.392
	50cm	0.230	0.390	0.152	0.237	0.277
L3	5cm	0.077	0.131	0.020	0.111	0.347
	25cm	0.109	0.168	0.061	0.106	0.263
	50cm	0.068	0.090	0.043	0.047	0.219
M5	5cm	0.141	0.218	0.089	0.130	0.247
	25cm	0.092	0.137	0.058	0.079	0.247
	50cm	0.078	0.105	0.049	0.056	0.220

Table 6 Station L3, M5 and J12 maximum, minimum, range and CV of in situ soil moisture at each depth from 2014 to 2018.

Figure 10 right provides the time series of the monthly average soil moisture of the three stations near the surface (5 cm) and the root zone (25 and 50 cm) from 2014 to 2018. Generally speaking, near-surface soil moisture content has a strong correlation with root zone soil moisture. The monthly average soil moisture content of the J12 station is higher than the M5 and L3, and the fluctuation range is also extensive. The monthly average soil moisture of M5 showed a similar pattern to that of L3. However, the change of J12 soil moisture in early spring was a few weeks later than that of M5 and L3, and it has had a relatively steady performance over the summer, while the soil moisture at each depth of M5 and L3 was decreasing at this time. The above phenomenon may be that the J12 station is far from the other two stations, which is caused by different geological and climatic reasons. This situation can also be seen from Table 6 that the coefficient of variation (CV) of J12 is higher than that of the others, which confirms that the range of soil moisture at J12 is more extensive. It can be seen from Figure 6 that both the SMOS soil moisture data and the soil moisture obtained from in-situ observations show apparent seasonality, which corresponds to the climate conditions of the region. The monthly mean soil moisture at each station peaks in early spring, and then it is dry during the summer. High temperature and surface

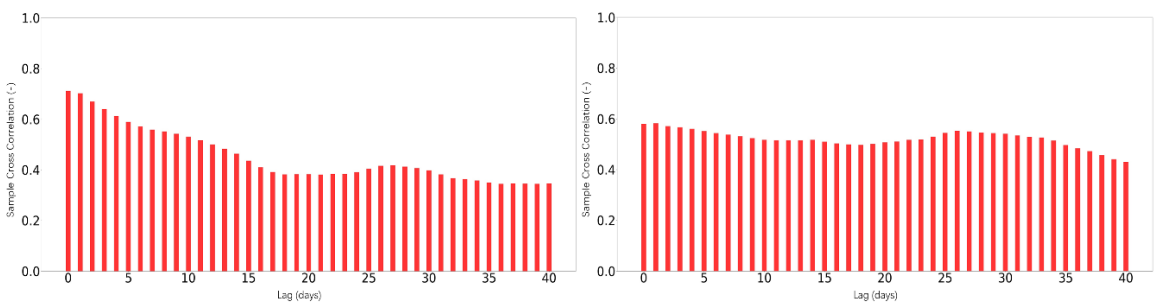
evaporation keep soil moisture at a higher content, and the difference in in-situ soil moisture at different depths was more significant than the cold and rainy periods.

In contrast, shortly after rainfall in autumn and winter, soil moisture rechargement begins. Due to the time delay to reach deeper layers, the water at 5 cm is more than 25 cm or even 50 cm. These patterns are similar to the report of Illston and his colleagues (2004). Mahmood et al. (2012) did a similar study and finally found that the coupling between the Nebraska root zone and near-surface soil moisture is stronger in locations where the climate is wetter.

4.1.2. Sample cross-correlation

The relationship between soil moisture is not instant, and there may be a certain time lag. It can see the degree of correlation between the two series under different lags by calculating the cross-correlation coefficient. It is more valuable to find the relationship between two series.

The 5-25 cm cross-correlation of H13 has a pronounced intensity peak, and it peaked on the first day (Figure 11 left-top), while the cross-correlation of 5-50 cm (Figure 11 right-top) is weak and lags 1-2 days to reach its peak. Not surprisingly, the maximum cross-correlation of 5-50 cm is usually weaker than 5-25 cm. This supports the findings of Wu (2002). The coupling strength between soil layers decreases as the depth increases. The decrease in cross-correlation strength with depth is consistent with the findings of Mahmood (2012). A similar situation also occurs at the M5 station, where the 5-25 cm cross-correlation lags about 1-2 days (Figure 11 left-bottom), and the 5-50 cm lags about 3-4 days (Figure 11 right-bottom). This is similar to the result of H13, as the depth increases, the cross-correlation tends to decrease, and the lag time tends to increase.



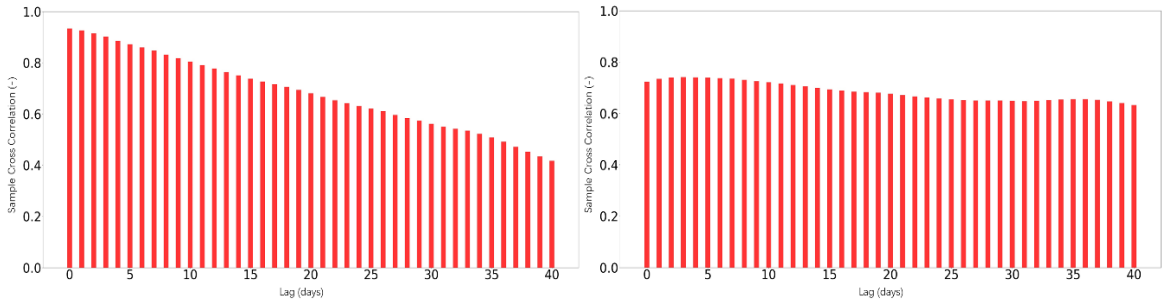


Figure 11 Sample cross-correlations of 5–25 cm (left) and 5–50 cm (right) at H13 and M5 stations. H13 at the top and M5 at the bottom.

Other stations cross-correlation of soil moisture are also calculated, founding that is similar to H13 and M5. At 5-25 cm, most autocorrelation functions show strong autocorrelation of surface soil moisture on a time scale of 1-5 days, but the autocorrelation intensity decreases significantly after ten days. This suggests that the strong cross-correlation between surface and root zone soil layers is partly due to the strong serial autocorrelation of surface soil moisture. However, autocorrelation quickly decreases beyond several days, and cross-correlations are more strongly a function of the coupling between soil layers.

4.2. T-optimal Estimation

This part presents the experimental results of soil moisture estimation in the root zone based on in-situ 5cm surface soil moisture. The aim is to find the most appropriate validation metric for estimating root zone soil moisture and whether T_{opt} results were affected by the target depth.

As an example, Figure 12 illustrates the time evolution of soil moisture at the depth of 25 cm from the ground at station M5 and the results of SWI with T parameters of 8, 50, and 100 days. In this case, it can be seen that as the T parameter increases, the time series curve becomes smoother. Therefore, the correlation coefficient reflecting the time series similarity seems to be suitable for estimating T_{opt}. As shown, the smaller the T value (8 days) calculated for SWI, the better the definition of in-situ evolution of soil moisture at 25 cm.

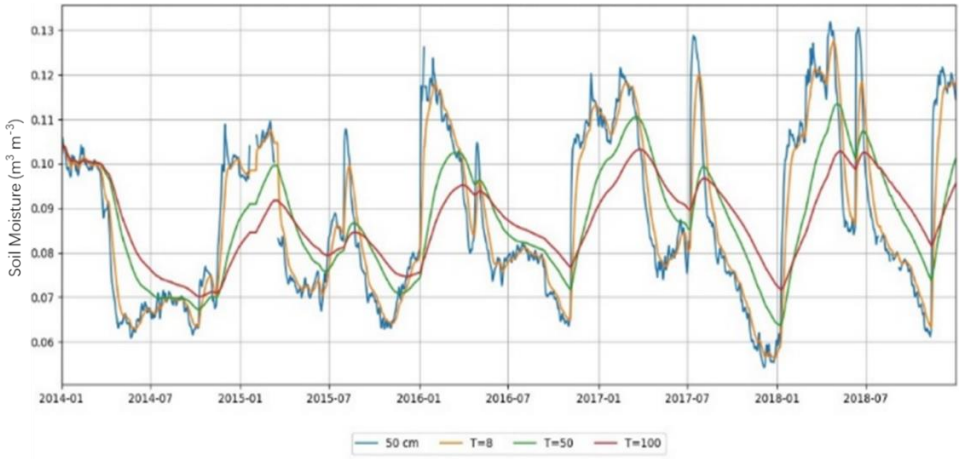
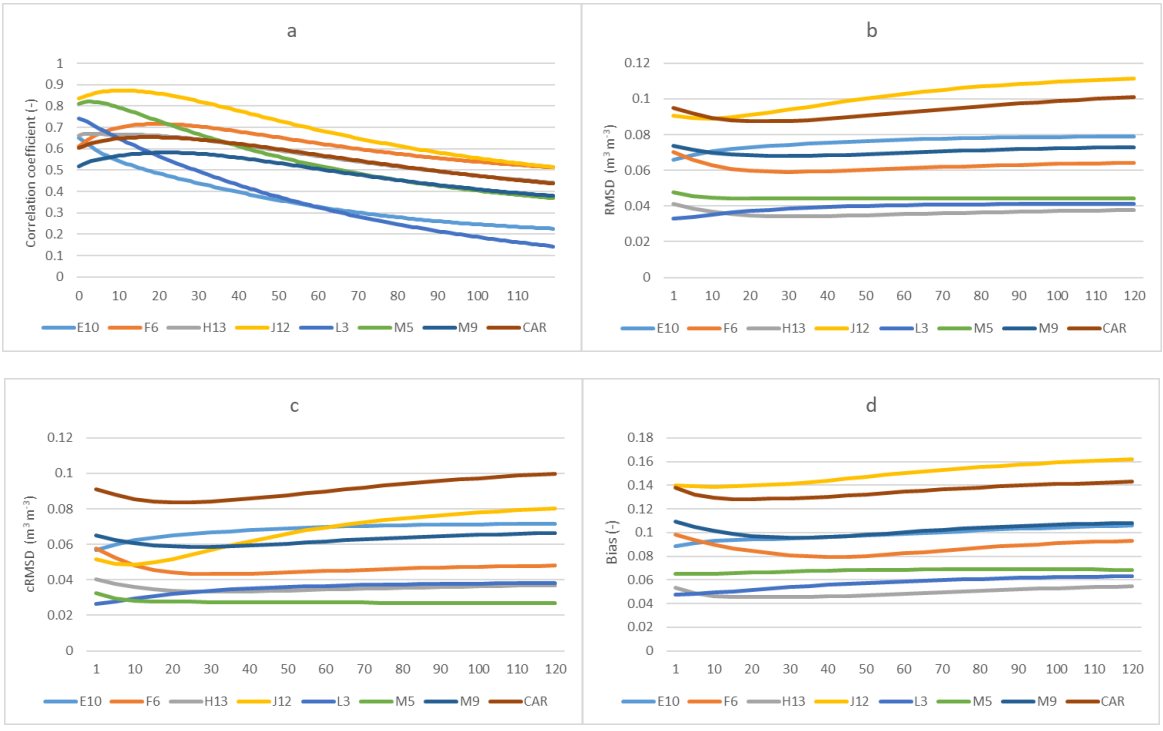


Figure 12 In situ soil moisture measurements and SWI time series from M5 station calculated with different T (8, 50 and 100 days) at 25 cm depth.

To discriminate T_{opt} with different metrics, it is necessary to compare in-situ soil moisture at different depths. In addition, the T range of calculating SWI is 1-120 days. Table 7-9 shows the different T_{opt} obtained by different metrics at each station with different depths. Overall, the T value increases with the soil depth (Tables 7 and 8), the same as previous model assumptions and research results of SWI (Albergel et al., 2008; Paulik et al., 2014). However, the T_{opt} of 0-50 cm (Table 9) shows similar results to those obtained at 25 cm (Table 7). It is most likely because the weighted average of the 0-50 cm estimate (1) gives more weight to 0 and 25 cm depths.



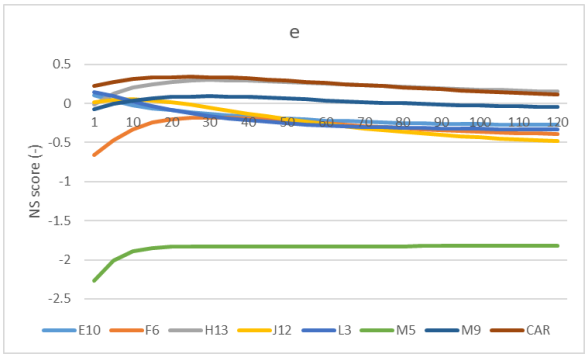


Figure 13 T parameter following the different metrics after the comparison between SWI in-situ and in situ soil moisture measurements at 25 cm depth.

Regarding the validation metrics used, when R is used as a decision metric, the value of T_{opt} is lower. For each station, T_{opt} based R values for 25 cm depth ranged from 1 to 23, from 1 to 61 for 50 cm, and from 1 to 27 for 0-50 cm. For example, only R can help the L3 station to determine its T_{opt} value. There is no variation of T_{opt} value obtained by the remaining four metrics. T_{opt} = 1 at 25cm, T_{opt} = 120 at 50cm, and T_{opt} = 1 at 0-50cm. The same problem that produced the same T_{opt} results occurred on F6 and E10 stations when using different metrics. It was confusing. To sum up, the T_{opt} obtained based on other metrics ranges from 1 to 120 and does not propose an ideal T. For the depth of 50 cm (Table 8), the T_{opt} values of the R and NS scores increase with depth. With RMSD, cRMSD, and Bias, the T_{opt} of some stations remained constant at all depths, or the variation was so irregular that it was difficult to refer to.

Figure 13 depicts the evolution of T values from 1 day to 120 days. When R is used to measure the optimum T value (Figure 13a), all stations show similar curve shapes and have similar R ranges. All the stations are divided into two types. L3 and E10 belong to the first type. T optimal appears when T=1, and then R-value decreases gradually with the increase of T, presenting a downward trend on the whole. The remaining stations belong to the second type. They do not obtain the optimal T when t=1, and the R-value increases with the increase of T and reaches the peak at a certain point to meet the T_{opt}, then begins to decline. The R-value curves of these stations all show an increase first and then a decrease. All the above shows the consistency of R as a T_{opt} measurement method. For most application scenarios, the correct time model for obtaining the soil moisture in the root zone may be more important than the absolute value, making R (Figure 13a)

the most appropriate measurement metric. In contrast, the degree of dispersion of other decision metrics is very high.

When using RMSD and cRMSD as the metric, there is a lack of apparent criteria. Because in most stations, the range between the maximum and minimum values of RMSD and cRMSD is tiny (approximately 0.02), making the task of choosing T_{opt} tough (Figure 13b and c). The same situation also appears in obtaining T_{opt} by using Bias and NS scores as the measurement metric. Most stations make no discernible difference at all depths (Table 7-9, Figure 13d and e). When the Bias is used to estimate the entire period, the difference between the SWI estimate and in situ soil moisture is negligible, resulting in a curve that looks like a straight line. It can be inferred that this may be due to the balance of negative and positive values. Therefore, RMSD, cRMSD, and Bias are considered unsuitable for calculating T_{opt}, so the possibility of further analysis of them is abandoned.

	E10	F6	H13	J12	L3	M5	M9	CAR
R	1	20	3	12	1	4	23	19
RMSD	1	30	30	10	1	120	30	25
cRMSD	1	30	30	10	1	120	25	20
Bias	1	40	25	10	1	5	30	15
NS-score	1	30	30	10	1	120	30	25

Table 7 T opt obtained by the different metrics at 25 cm depth.

	E10	F6	H13	J12	L3	M5	M9	CAR
R	1	57	44	61	8	26	42	59
RMSD	5	120	45	40	120	120	70	110
cRMSD	120	120	45	40	120	120	55	70
Bias	1	120	120	25	120	75	120	120
NS-score	5	120	45	40	120	120	70	115

Table 8 T opt obtained by the different metrics at 50 cm depth.

	E10	F6	H13	J12	L3	M5	M9	CAR
R	1	20	18	24	1	4	24	27
RMSD	1	45	30	20	1	35	35	40
cRMSD	1	45	30	20	1	40	30	35
Bias	1	120	25	10	1	25	35	25
NS-score	1	45	30	20	1	35	35	40

Table 9 T opt obtained by the different metrics at 0-50 cm depth.

The SWI SMOS is calculated with SMOS data and T_{opt} value in Table 7 used metric R. As shown in Figure 14, results of F6 station T= 1, 20 and 40 at 25cm depth, where T=20 is T_{opt}. It can be

seen that when the value of T is less than T_{opt}, SMOS data distribution is extremely scattered on both sides of in-situ observation data, which is very unfavourable to judge the relationship between two data sets. When the T value is above than T_{opt} value, SMOS data distribution will be highly concentrated, and the curve tends to be flat, but at this time, SMOS data will lose too many data details and comparison points, which is also not conducive to judging the relationship between two data sets. However, after using T_{opt}, the problem of data distribution being too scattered is solved, and the loss of accuracy caused by the curve being too smooth is also avoided.

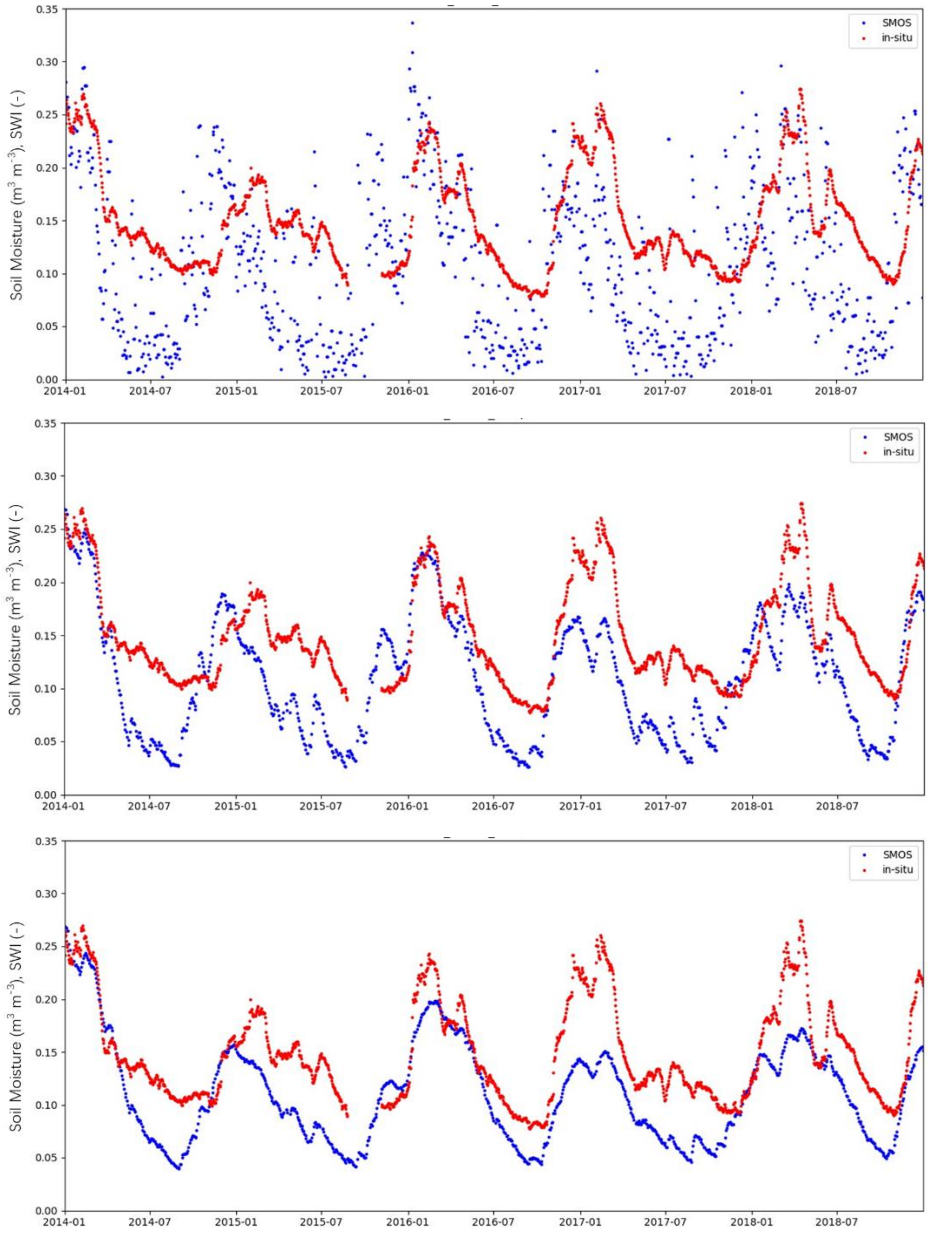


Figure 14 Results of the comparison between the time series of the SWI SMOS (using T obtained with R) with the in-situ measurement stations at F6 station 25 cm depth. Top: T=1; Mid (T_{opt}): T=20; Below: T=50.

4.3. T-optimal Estimation for Soil Moisture Seasons and Soil Textures

The previous section confirmed that soil root zone moisture could be inferred using an exponential filter from surface soil moisture. Correlation coefficient (R) was evaluated as the most suitable decision metric for T_{opt} estimation, so R was taken as the only metric method for optimal T estimation of soil moisture season. Meanwhile, there may be many factors affecting the final result of T_{opt} estimation. So this part will use the same data as the last part to analyze whether different soil moisture seasons and soil textures would impact T_{opt} estimation or not.

4.3.1. Soil moisture season

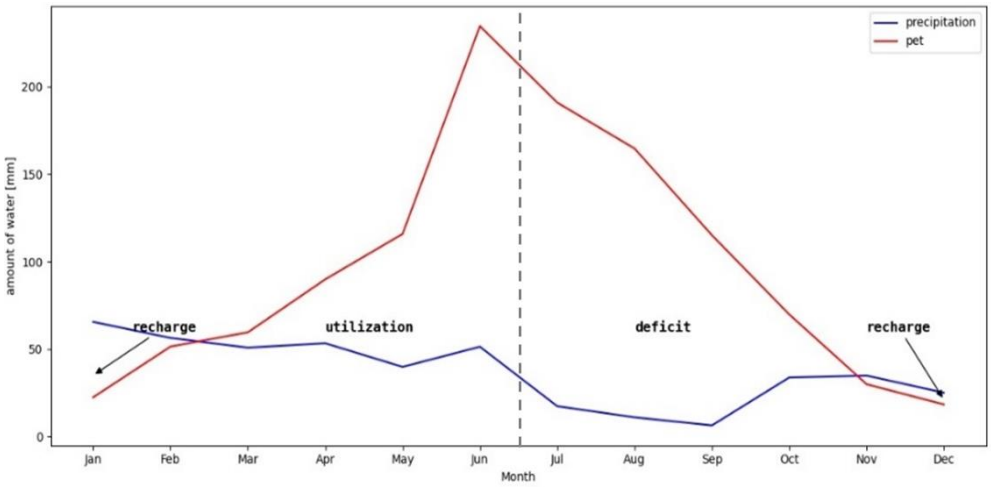


Figure 15 Division of soil moisture seasons in REMEDHUS from 2014 to 2018.

Figure 15 shows the average monthly cumulative precipitation and estimated potential evapotranspiration (PET) of REMEDHUS from 2014 to 2018. Water enters the soil in the form of precipitation and is released into the atmosphere through evapotranspiration. From this relationship (Table 5), the soil moisture season is defined, including recharge (November - February 10th), utilization (February 10th - Mid June) and deficit (Mid June - October).

Figure 16 shows the results of three soil seasons obtained by numerical average. It can be seen that compared with recharge and deficit, the overall results of utilization is better with the optimal R values of all stations are above 0.7, indicating that there is a strong or extremely strong correlation between the root zone and surface soil moisture in utilization. Besides, most stations' best R-value

in recharge is between 0.4-0.7 and between 0.5-0.8 in deficit season, both of which are inferior to utilization.

	E10	F6	H13	J12	L3	M5	M9	CAR
Recharge	12	9	2	34	1	1	15	12
Utilization	30	14	4	13	2	3	10	30
Deficit	17	100	1	1	1	6	41	17

Table 10 T_{opt} obtained by the different soil moisture seasons at 25 cm depth used numerical average method.

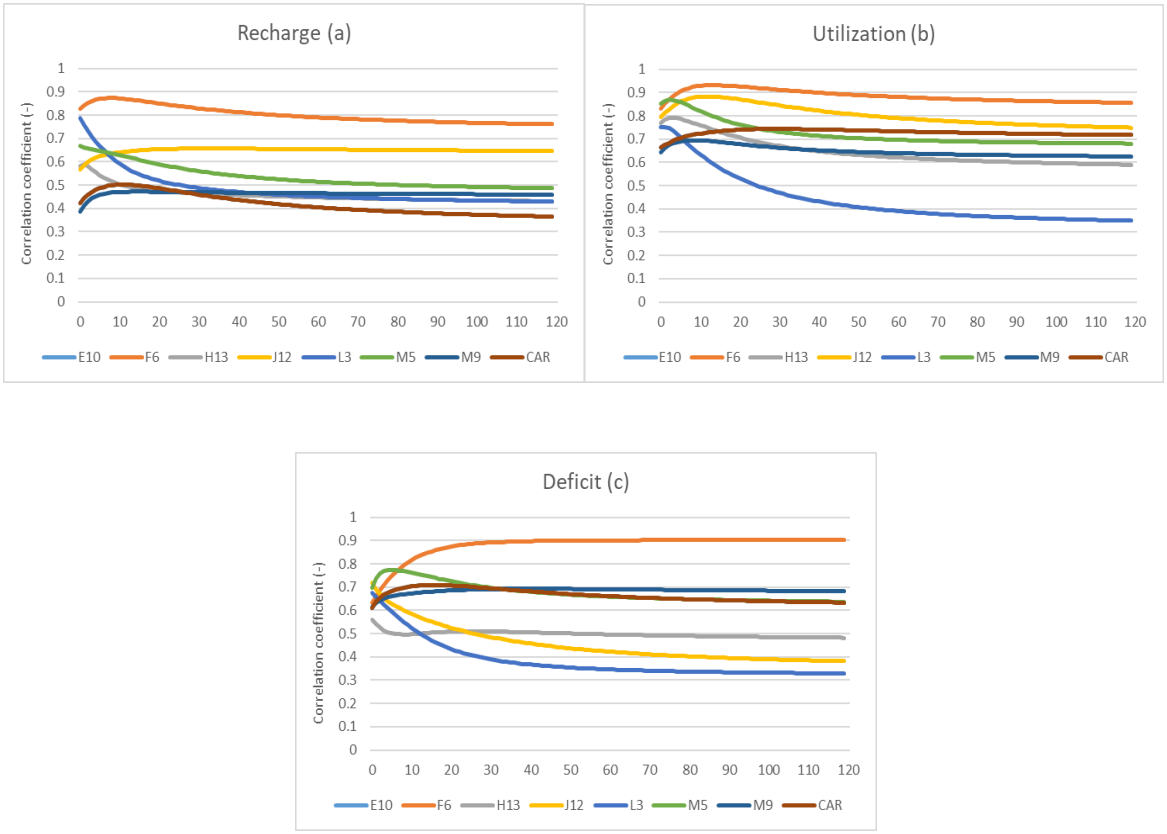


Figure 16 T parameter obtained by the different soil moisture seasons at 25 cm depth used numerical average method.

Figure 17 shows the results of three soil seasons obtained by time series splicing. It can be seen that the optimal R-value of utilization at each site is between 0.5-0.9, which is better than 0.3-0.8 of recharge and the results obtained at each station in deficit.

	E10	F6	H13	J12	L3	M5	M9	CAR
Recharge	1	6	1	17	1	1	4	5
Utilization	1	8	3	11	1	3	15	7
Deficit	1	17	1	1	1	8	3	1

Table 11 T_{opt} obtained by the different soil moisture season at 25 cm depth used time series splicing method.

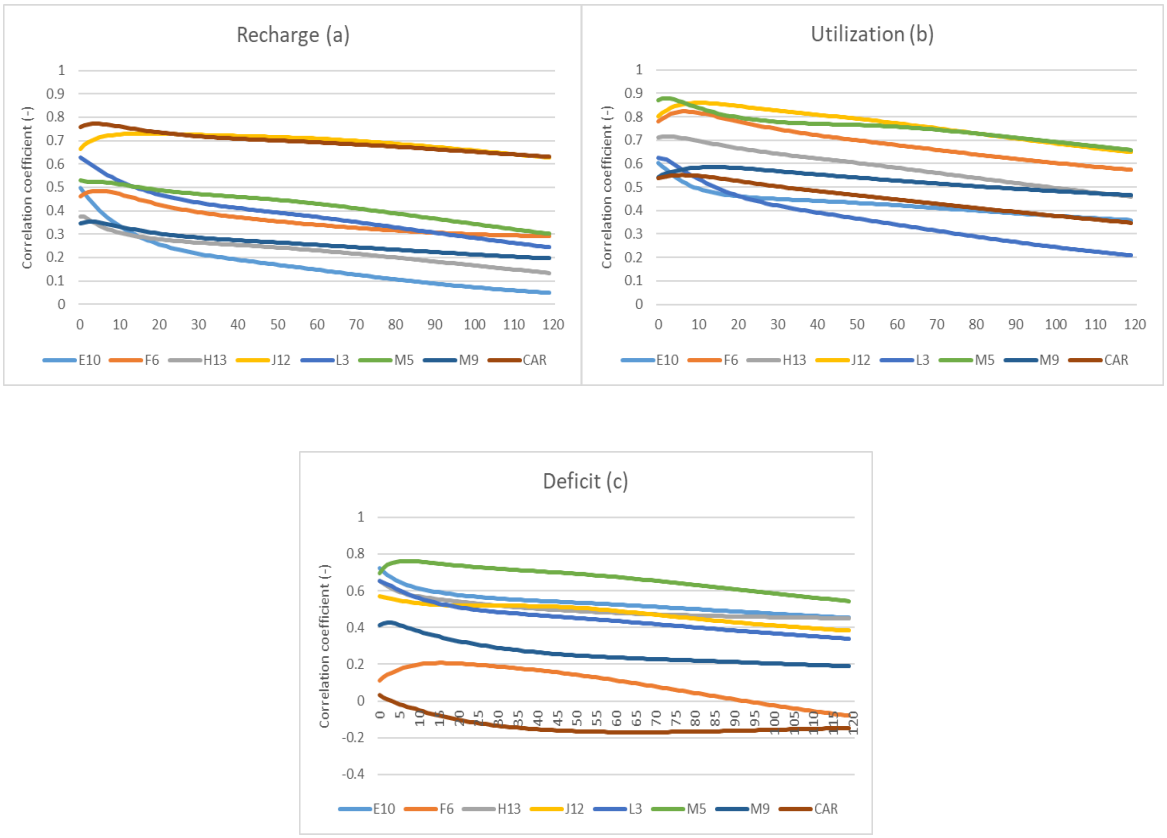


Figure 17 T parameter obtained by the different soil moisture season at 25 cm depth used time series splicing method.

The above analysis proves that soil moisture season affects the optimal R-value, the utilization is significantly better than the other two seasons. It can be seen from Table 7, Table 10 and Table 11 that the T_{opt} results of Utilization at each site are similar to the annual T_{opt}. Utilization is the main growing season of plants when the strong root water absorption and evapotranspiration make the soil water activity reach the highest value in the year. It can be said that the T_{opt} value of Utilization can replace the annual T_{opt} value to some extent. However, recharge and deficit do not have such representation. But after averaging the T_{opt} results of the three seasons, similar results of annual T_{opt} were obtained.

4.3.2. Soil textures

The stations with the most similar seasonal and annual estimates, L3, M5, and H13, have different land uses. As Table 2 shown, M5 is rainfed, L3 is the vineyard, and H13 is forest-pasture. It seems that there is no significant relationship between the stability of T optimal estimation for soil

moisture seasons and land use. However, coincidentally, the soil textures of these stations are homogeneous (L3 and M5: Loamy Sand, H13: Sandy Loam), and the soil textures in each layer are the same. Such stable soil structure may lead to a very stable soil water demand, making it easier to obtain ideal results when estimating T_{opt}. The results of the stations with large fluctuation (F6, M9 and CAR) were significantly higher than those of other stations without exception, which seems to indicate that field capacity (θ_{FC}) can affect the results of T-optimal estimation for soil moisture seasons. However, the value of θ_{FC} is usually affected by soil textures, usually clay< loam sandy< sandy soil, so that the soil textures may affect the results of T_{opt} estimation to a certain extent.

4.4. Estimating Root Zone Soil Moisture Using SMOS

The results show that near-surface soil moisture is medial to strongly coupled with soil moisture in the root zone and that near-surface SWI. The data based on in-situ observations can be used to provide reasonable predictions of root zone SWI. Next, we will use SMOS derived surface soil moisture for each station to infer root zone soil moisture to evaluate the effectiveness of the exponential decay filter. SMOS data were collected every two days between January 2014 and December 2018. M9 station is not included in the following experiments due to data loss.

4.4.1. Station–SMOS surface soil moisture comparison

The corresponding relationship between the SMOS surface inversion and the near-surface in-situ soil moisture observation provides the "baseline" error, which will compare the accuracy of SMOS estimation of soil moisture in the root zone. The station data is directly compared with SMOS, and the same metric (R, RMSD, cRMSD, Bias, NS-Score) as before is used to evaluate the strength of the relationship between the two data sets. Due to the scale differences between the station and satellite footprints, all error metrics are based on SWI rather than the raw volumetric water content. This ensures that the two data sets have the same mean and value range so that the error evaluation is only based on the covariance of the data set.

As Table 12 shown. The R values of the seven stations ranged from 0.51 to 0.64, and the

fluctuation range of each station was minimal, with an overall average of 0.58, although it was in the "Medium correlation" range（0.4 < R < 0.6）, but it is very close to the "Strong correlation" range （0.6 < R < 0.8）. This shows that there is a good correspondence between the two data sets. In the evaluation of RMSD, the average RMSD of the seven stations is 0.27. Except for J12, which is slightly larger (0.39), the results of the remaining six stations are all between 0.2-0.3. The same situation appeared in the results of Bias. Similarly, the value of the J12 station was significantly higher than that of other stations, which may further confirm the particularity of this station. When cRMSD was used for evaluation, the results of each station showed significant consistency with a small value, indicating that the evaluation results were very positive and the consistency between the two data sets was good. Due to the value range of NS-score is $(-\infty, 1]$, the results of the seven stations are scattered, so it is inconvenient to estimate.

		E10	F6	H13	J12	L3	M5	CAR
R	Surface	0.51	0.59	0.61	0.57	0.53	0.64	0.60
	Root Zone	0.66	0.53	0.61	0.76	0.54	0.50	0.60
RMSD	Surface	0.20	0.25	0.26	0.39	0.27	0.26	0.23
	Root Zone	0.20	0.29	0.39	0.30	0.38	0.29	0.26
cRMSD	Surface	0.19	0.18	0.18	0.19	0.20	0.19	0.19
	Root Zone	0.17	0.16	0.16	0.15	0.18	0.18	0.18
Bias	Surface	0.22	0.29	0.31	0.51	0.33	0.31	0.27
	Root Zone	0.23	0.36	0.53	0.37	0.48	0.37	0.30
NS-score	Surface	-0.19	-0.73	-0.98	-3.90	-0.90	-0.32	-0.09
	Root Zone	0.09	-1.58	-4.47	-0.66	-3.50	-1.90	-0.42

Table 12 Statistics showing the correspondence of SMOS and station SWI data sets. Surface values are calculated from the in-situ surface – SMOS surface, while values under the root zone are computed from the in-situ root zone – SMOS root zone. All values are SWI unitless.

4.4.2. Station–SMOS root zone soil moisture comparison

The exponential decay filter described in section 3.3 is used to estimate root zone soil moisture based on SMOS surface data. The T_{opt} value of each station has been obtained in the previous experiment, and the SWI of the SMOS root zone is estimated from surface data using the T_{opt} obtained by R in Table 7. The SMOS root zone SWI estimated value was compared with the in-situ root zone SWI value, and the accuracy was measured by R, RMSD, cRMSD, Bias and NS-Score.

SMOS root zone SWI estimates are compared with 25cm in situ SWI measurements at all stations. The prediction results of the root zone using SMOS were evaluated based on the relationship between SMOS surface inversion results and 5 cm soil moisture observation. It can be used to determine the error in the SWI estimation of the root zone is due to the exponential filtering method or the uncertainty of SMOS. Table 12 shows the evaluation results. The R-value ranges from 0.50 to 0.76, and the overall average value is 0.6, similar to the R result of SMOS data and in-situ surface observation data in the previous section. In the evaluation of RMSD, the minimum value of the seven stations is 0.20, the maximum is 0.39, and the average value is 0.30, which is not much different from 0.27 in the previous experiment. When cRMSD was used to evaluate the results, the results were still highly concentrated, with an average of 0.17. All the above results indicated that there was a good agreement between the in-situ root soil moisture and root zone SMOS SWI.

Figure 18 top shows the R-value for all stations. R is calculated between in-situ surface observations and in-situ root zone (blue triangles), in-situ surface observations and SMOS surface inversion (green squares), and in-situ root zone observations and SMOS root zone estimation (red circles). It can be seen that the red circles and the green squares have a very similar distribution and are lower than the blue triangles. This indicates that the accuracy between in-situ surface observations and in-situ root zone estimates is highest, which is obvious since their values are derived from actual in-situ measurements. However, the remaining two data sets also had good estimation accuracy, with values above 0.5 at each station. This indicates that it is feasible to estimate soil moisture in the root zone from SMOS data. Although the accuracy is not as high as that between in-situ surface observation and in-situ root zone estimation, it is similar to the "baseline" error (the accuracy between in-situ surface observation and SMOS surface inversion). A similar conclusion can be drawn from the RMSD results (Figure 18 below), the blue is significantly better than the green and red, but the green and red also have good accuracy.

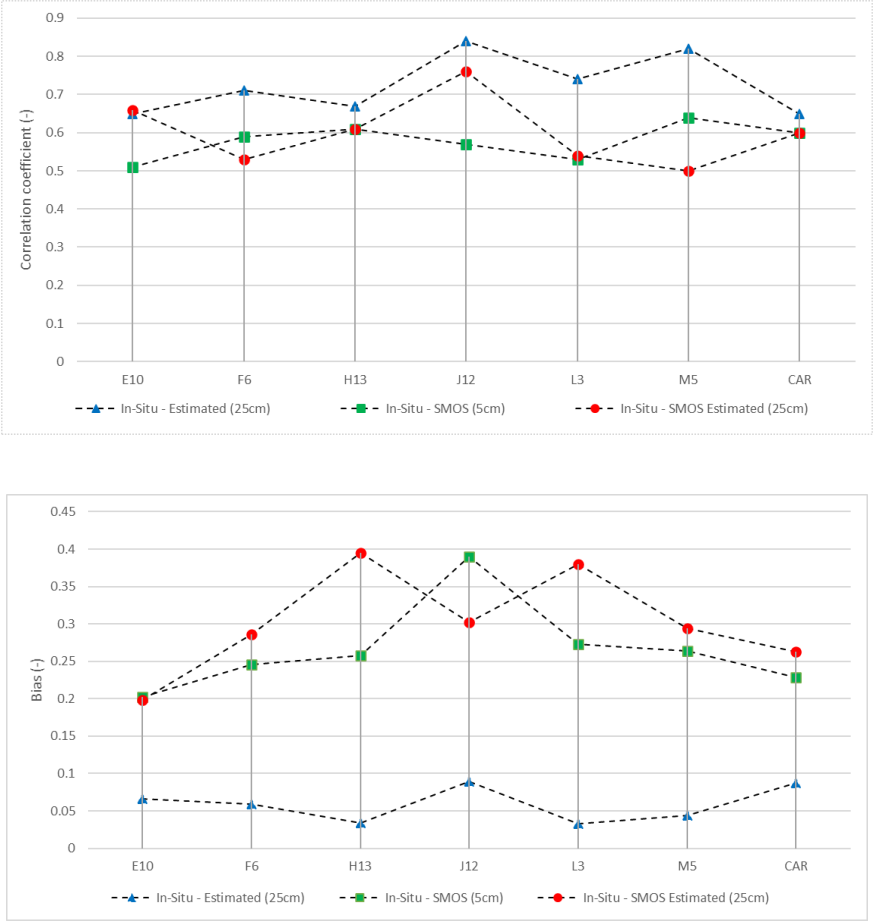


Figure 18 The figure shows R-value (top) and RMSD value (below) at all stations. Blue triangles are calculated with the in-situ surface observations and in-situ root zone, green squares are the in-situ surface observations and SMOS surface inversion, and red circles are the in-situ root zone observations and SMOS root zone estimation.

The results in Section 4.3 indicate that soil moisture season will affect T_{opt} selection. Next, it will explore whether different soil moisture seasons will affect the Station-SMOS root moisture comparison results. Here, the method in Section 4.3.1 and the T_{opt} value in Table 10 are selected. The Correlation coefficient (R) results are shown in Table 13 and Figure 19.

	E10	F6	H13	J12	L3	M5	CAR	MEAN
Annual	0.66	0.53	0.61	0.76	0.54	0.50	0.60	0.60
Recharge	0.59	0.35	0.75	0.53	0.48	0.55	0.71	0.56
Utilization	0.89	0.90	0.77	0.86	0.59	0.69	0.84	0.79
Deficit	0.60	0.32	0.61	0.22	0.34	0.53	0.54	0.45

Table 13 The R-value of station-SMOS root zone soil moisture estimation in different soil moisture seasons.

There is a big difference in the results across the three soil moisture seasons. The average of the annual correlation coefficient is 0.60, but Utilization is the highest, reaching 0.79, with a range of

0.59-0.90. Recharge is slightly better than deficit, with an average of 0.56. The lowest result is the deficit, 0.45 on average, and even the J12 station recorded 0.22. The histogram can very clearly reflect this situation, the green column (utilization) is significantly higher than that of the other two soil moisture seasons at all stations. Thus, it can be said that different soil moisture seasons will significantly affect the accuracy of soil moisture estimation in the root region using SMOS, and the Utilization result is obviously better than the other two soil moisture seasons. Selection of Utilization for soil moisture estimation can improve the accuracy of the results.

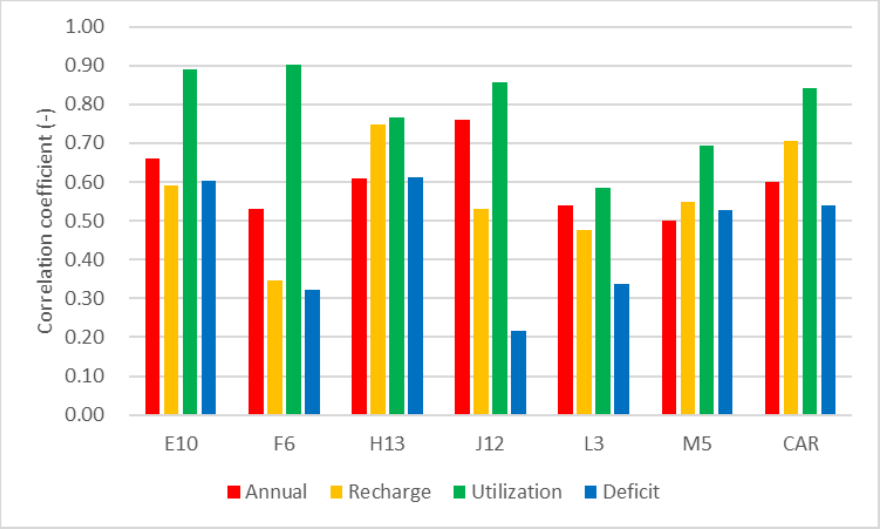


Figure 19 The histogram when estimating station-SMOS root zone soil moisture in different soil moisture seasons used Correlation coefficient (R).

5. Discussion

Satellite soil moisture inversions are valuable because they provide better spatial coverage than in-situ soil moisture observations. SMOS measures water content in a few centimetres of soil at the top, so it cannot directly determine soil moisture status in the root zone. Therefore, it is essential to evaluate the degree of association between near-surface and root-soil moisture when retrieving soil root zone moisture by satellite. This study is located in a specific area. Therefore, any results obtained cannot directly generalize all cases.

5.1. T-optimal Estimation

Researchers used simulated soil moisture data to obtain the T_{opt} value of 30-90 days with ASCAT soil moisture and used correlation coefficient as the estimated metric of T_{opt} for a layer depth of 1–1.5 m (Brocca et al., 2010). Pellarin (2006) used ERS soil moisture data to obtain the T_{opt} is 39 days, which is higher than the T_{opt} t obtained in this study. Other studies use different depths of in-situ soil moisture to estimate T_{opt}. For example, Albergel et al. (2008) and Ford et al. (2014) used NS statistics to find that T_{opt} has a low value. Su (2015) pointed out that since the in-situ data is less noisy than satellite retrieval, the lower T value is not surprising. In particular, Albergel et al. (2008) obtained T_{opt} = 6 days in the 30 cm layer in SMOSMANIA and SMOSREX in France, and Ford et al. (2014) found that T_{opt} = 8 days and nine days in Messonet, Oklahoma and Nebraska Auto Weather. The results obtained in these two different studies are similar to the T_{opt} value obtained in this study.

Paulik et al. (2014) used different T values and ASCAT surface soil moisture data to obtain SWI at different depths, but they did not find a clear T_{opt} value. But they concluded that the observed soil T_{opt} generally increases with soil depth increase. It is also similar to the results obtained in this study.

Alberger et al. (2009) found that the T_{opt} using 30 cm ASCAT soil moisture data is 14 days. Broca et al. (2010) found that T_{opt} = 19.5 days (10cm), T_{opt} = 23 days (20cm), and T_{opt} = 29 days (40cm), also using ASCAT soil moisture. These results of those T_{opt} are higher than the value obtained

using R in this study (9 days for 25 cm). In addition, in a study on the same area using ERS data (Ceballos et al. 2005), T_{opt} values (40 days) at a depth of 0-25 cm were also obtained. The time interval can explain the difference between the two studies in situ data measurements. The latter study is once a day, the former is once every two weeks, and the soil moisture probes are inconsistent.

In terms of the length of the experimental data, Broca et al. (2010) believe that the high variability of T_{opt} may be due to the different lengths of data used. However, the time period used to calculate T_{opt} in our study and that of Albergel et al. (2008) is very different, but T_{opt} is similar. Therefore, questioning the length of the series is not that important.

Regarding the selection of the database used to calculate the T_{opt}, the use of in-situ soil moisture or remote sensing series may result in a significant difference in the estimated value of T, which is more important than the length of the series. Many little-known factors seem to affect T_{opt} 's search. Theoretically, soil texture and climate will be the key factors.

5.2. Effects of Soil Moisture Season and Soil texture

Soil moisture season is a special seasonal division that reflects soil moisture variation, which requires more accurate PET and precipitation data support. In the absence of this data, the same purpose can be achieved to a certain extent by using the natural division of the four seasons of spring, summer, autumn and winter.

The REMEDHUS network used in this study is not very large, and the data of four climate collecting stations can reflect the climate change of this network. But from the analysis of the actual situation, different stations should have different soil moisture season periods, if the soil moisture season can be divided according to different stations, the result will be more accurate in theory. In addition, the five years PET and precipitation data were integrated in this study to obtain a uniform seasonal period of soil moisture and apply it to each year. This facilitates the experiment to some extent, but it also ignores the unique climate differences of different years.

Previous studies have shown that soil texture does not affect T_{opt} (Paulik et al., 2014). However,

there are very few research materials related to this issue, and further exploration is needed in the follow-up, and the results need to be further verified. In this study, whether the soil structure of the station is uniform or not seems to indicate the results of the experiment. When the soil characteristics vary greatly with depth, the performance of the exponential filter will also fluctuate. But this is just a conjecture. This study only discovered the possibility of this influencing factor, and a lot of research must be combined with geological knowledge to verify it further. Whether θ_{FC} or soil textures will affect the T_{opt} results also needs further investigation.

It can be seen from this study that the T_{opt} of the exponential filter is very sensitive to soil moisture and varies greatly according to climate conditions. Detailed and thorough research of this problem will certainly be conducive to the study of T_{opt} in the future.

5.3. Numerical Average and Time Series Splicing

The two methods used in seasonal T_{opt} estimation of soil moisture, numerical average and time series splicing, are both trying to solve the problem that time series of the same season in different years are not connected. In essence, there are only different data processing methods, so it is temporarily impossible to find appropriate indicators to judge the merits of the two schemes. In principle, numerical averaging is more reasonable because splicing time-discontinuous series as a continuous series to calculate SWI is inconsistent with the initial conditions of SWI. The method of numerical average will be adopted to calculate the average value, which will reduce the impact of random situations on the results. To analyze the influence of the two methods on experimental results, it is necessary to set up more complete practical steps for verification. However, this issue is of limited value, and the scientific nature and merits of the two methods will not be explored here.

5.4. Future Work

The coupling relationship between surface soil moisture and root zone soil moisture was emphatically verified in this study. This laid a foundation for the inversion of zone soil soil moisture using SMOS satellite data. Future work can be divided into two main directions.

The first option is to continue exploring the factors that influence the correlation between surface soil moisture and zone soil soil moisture, including climatic conditions, seasonal factors, soil textures, water content, etc. Although this study reached some preliminary conclusions, it did not establish a significant control group throughout the study. Future work can select one or multiple of the potential influencing factors for targeted research so that the relationship between surface soil moisture and root zone soil moisture becomes predictable.

The second option is to use SMOS satellite data to inverse zone soil soil moisture. This direction can obtain zone soil soil moisture from satellite missions and generate many branches from it. The method can calculate plant water use, predict crop growth, monitor regional rainfall and even respond to floods and droughts caused by unusual weather conditions.

6. Conclusion

Satellite soil moisture inversions are valuable because it provide better spatial coverage than in-situ soil moisture observations. SMOS measures the moisture content of several centimetres on the soil surface, so it cannot directly determine the soil moisture status in the root zone. Therefore, when using satellites to retrieve soil moisture in the root zone, it is essential to assess the degree of correlation between near-surface and root zone soil moisture. This study quantified the coupling strength between the surface soil moisture and root zone soil moisture. It also evaluated whether the exponential filtering method can predict soil moisture in the root zone. Finally, the in-situ observations are used to evaluate SMOS root zone soil moisture estimation accuracy based on exponential filtering.

The study analyzed the correlation between the near-surface and root zone soil moisture at eight stations of the REMEDHUS network. Through the sample cross-correlation, it proved that the coupling strength decreases with the soil depth. The exponential filter's main factor (T_{opt}) is very sensitive to soil moisture and varies greatly depending on the climate. T selection based on RMSD and cRMSD metrics should be ignored because the range between the maximum and minimum values of these metrics is too small, making selection difficult or even impossible. The resulting curve is almost a straight line when Bias is used as the metric, it cannot give a clear T_{opt}. The shape of the metric R curve is similar in all stations, and the R range of each station is also similar, which is significantly better than the NS score result.

Therefore, when using exponential filtering to infer root zone soil moisture, R is considered to be the most suitable decision metric for T_{opt} estimation. SMOS soil moisture data and in-situ soil moisture have obvious seasonality. Different soil moisture seasons will affect the selection result of T. In the study of soil moisture season, the result of utilization is better than the other two seasons. The study did not find that land use is a factor that affects the T_{opt}, but choosing the station with the homogeneous soil texture is conducive to the development of the experiment. Because when the soil characteristics vary greatly with depth, the exponential filtering method also performs poorly. Different soil textures may make the final experimental results different from

expectations. Therefore, it is better to choose the stations with uniform and homogeneous soil textures when making preliminary inferences on the network.

After establishing a strong correlation between near-surface and root-zone soil moisture, the exponential filtering method of Albergel et al. (2008) was used to estimate root-zone soil moisture based on near-surface observation data. This method is proved effective. Based on SMOS surface inversion, the exponential filtering method is used to estimate the root zone soil moisture. The estimated value of the root zone was compared with the observed value of 25cm soil moisture. The SMOS root zone soil moisture correlates with the observed in-situ root zone values, and the correlation of all stations has reached the standard of "medium correlation" or "strong correlation". But different soil moisture seasons can significantly affect this result.

In summary, the main conclusions of this study are: (1) The near-surface and root zone soil moisture is strongly coupled; (2) The exponential filter method can provide an accurate estimation of root zone soil moisture, but it is necessary to pay attention to the influence of soil moisture season and soil texture on parameter T; (3) SMOS surface soil moisture can be used to estimate root-zone soil moisture using exponential filter method however the accuracy of prediction varies greatly in different soil moisture seasons.

References

Albergel C, Rüdiger C, Pellarin T, et al. 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[J]. Hydrology and Earth System Sciences, 2008, 12(6): 1323-1337.

Albergel C, Rüdiger C, Carrer D, et al. An evaluation of ASCAT surface soil moisture products with in-situ observations in Southwestern France[J]. Hydrology and Earth System Sciences, 2009, 13(2): 115-124.

Basara J B, Crawford K C. Linear relationships between root-zone soil moisture and atmospheric processes in the planetary boundary layer[J]. Journal of Geophysical Research: Atmospheres, 2002, 107(D15): ACL 10-1-ACL 10-18.

Brocca L, Melone F, Moramarco T, et al. ASCAT soil wetness index validation through in situ and modeled soil moisture data in central Italy[J]. Remote Sensing of Environment, 2010, 114(11): 2745-2755.

Brocca L, Melone F, Moramarco T, et al. Improving runoff prediction through the assimilation of the ASCAT soil moisture product[J]. Hydrology and Earth System Sciences, 2010, 14(10): 1881-1893.

Calvet J C, Fritz N, Froissard F, et al. In situ soil moisture observations for the CAL/VAL of SMOS: The SMOSMANIA network[C]//2007 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2007: 1196-1199.

Carranza C, Nolet C, Pezij M, et al. Root zone soil moisture estimation with Random Forest[J]. Journal of Hydrology, 2021, 593: 125840.

Ceballos A, Scipal K, Wagner W, et al. Validation of ERS scatterometer-derived soil moisture data in the central part of the Duero Basin, Spain[J]. Hydrological Processes: An International Journal, 2005, 19(8): 1549-1566.

Ford T W, Harris E, Quiring S M. Estimating root zone soil moisture using near-surface observations from SMOS[J]. Hydrology and Earth System Sciences, 2014, 18(1): 139-154.

Frye J D, Mote T L. Convection initiation along soil moisture boundaries in the southern Great Plains[J]. Monthly Weather Review, 2010, 138(4): 1140-1151.

Georgakakos K P, Bae D H. Climatic variability of soil water in the American Midwest: Part 2. Spatio-temporal analysis[J]. Journal of hydrology, 1994, 162(3-4): 379-390.

Gill M K, Asefa T, Kemblowski M W, et al. Soil moisture prediction using support vector machines 1[J]. JAWRA Journal of the American Water Resources Association, 2006, 42(4): 1033-1046.

González-Zamora Á, Sánchez N, Martínez-Fernández J, et al. Long-term SMOS soil moisture products: A comprehensive evaluation across scales and methods in the Duero Basin (Spain)[J].

Physics and Chemistry of the Earth, Parts A/B/C, 2015, 83: 123-136.

Illston B G, Basara J B, Crawford K C. Seasonal to interannual variations of soil moisture measured in Oklahoma[J]. International Journal of Climatology: A Journal of the Royal Meteorological Society, 2004, 24(15): 1883-1896.

Jackson T J. Remote sensing of soil moisture: implications for groundwater recharge[J]. Hydrogeology journal, 2002, 10(1): 40-51.

Jackson T J, Bindlish R, Cosh M H, et al. Validation of Soil Moisture and Ocean Salinity (SMOS) soil moisture over watershed networks in the US[J]. IEEE Transactions on Geoscience and Remote Sensing, 2011, 50(5): 1530-1543.

Kerr Y H, Waldteufel P, Wigneron J P, et al. The SMOS mission: New tool for monitoring key elements ofthe global water cycle[J]. Proceedings of the IEEE, 2010, 98(5): 666-687.

Koster R D, Dirmeyer P A, Guo Z, et al. Regions of strong coupling between soil moisture and precipitation[J]. Science, 2004, 305(5687): 1138-1140.

Mahmood R, Hubbard K G. Relationship between soil moisture of near surface and multiple depths of the root zone under heterogeneous land uses and varying hydroclimatic conditions[J]. Hydrological Processes: An International Journal, 2007, 21(25): 3449-3462.

Mahmood R, Littell A, Hubbard K G, et al. Observed data-based assessment of relationships among soil moisture at various depths, precipitation, and temperature[J]. Applied Geography, 2012, 34: 255-264.

Martínez-Fernández J, González-Zamora A, Sánchez N, et al. A soil water based index as a suitable agricultural drought indicator[J]. Journal of Hydrology, 2015, 522: 265-273.

Paulik C, Dorigo W, Wagner W, et al. Validation of the ASCAT Soil Water Index using in situ data from the International Soil Moisture Network[J]. International journal of applied earth observation and geoinformation, 2014, 30: 1-8.

Rüdiger C, Hancock G, Hemakumara H M, et al. Goulburn River experimental catchment data set[J]. Water Resources Research, 2007, 43(10).

Rüdiger C, Calvet J C, Gruhier C, et al. An intercomparison of ERS-Scat and AMSR-E soil moisture observations with model simulations over France[J]. Journal of Hydrometeorology, 2009, 10(2): 431-447.

Sanchez N, Martínez-Fernández J, Scaini A, et al. Validation of the SMOS L2 soil moisture data in the REMEDHUS network (Spain)[J]. IEEE Transactions on Geoscience and Remote Sensing, 2012, 50(5): 1602-1611.

Schmugge T, O'Neill P E, Wang J R. Passive microwave soil moisture research[J]. IEEE Transactions on Geoscience and Remote Sensing, 1986 (1): 12-22.

Su C H, Narsey S Y, Gruber A, et al. Evaluation of post-retrieval de-noising of active and passive microwave satellite soil moisture[J]. Remote Sensing of Environment, 2015, 163: 127-139.

Taylor C M, Parker D J, Harris P P. An observational case study of mesoscale atmospheric circulations induced by soil moisture[J]. Geophysical Research Letters, 2007, 34(15).

Taylor C M, de Jeu R A M, Guichard F, et al. Afternoon rain more likely over drier soils[J]. Nature, 2012, 489(7416): 423-426.

Van Genuchten M T. A closed‐form equation for predicting the hydraulic conductivity of unsaturated soils[J]. Soil science society of America journal, 1980, 44(5): 892-898.

Wagner W, Lemoine G, Rott H. A method for estimating soil moisture from ERS scatterometer and soil data[J]. Remote sensing of environment, 1999, 70(2): 191-207.

Wagner W, Noll J, Borgeaud M, et al. Monitoring soil moisture over the Canadian Prairies with the ERS scatterometer[J]. IEEE Transactions on Geoscience and Remote Sensing, 1999, 37(1): 206-216.

Wu W, Dickinson R E. Time scales of layered soil moisture memory in the context ofland–atmosphere interaction[J]. Journal of Climate, 2004, 17(14): 2752-2764.

Wu W, Geller M A, Dickinson R E. The response of soil moisture to long-term variability of precipitation[J]. Journal of Hydrometeorology, 2002, 3(5): 604-613.