Effect of vegetation index choice on soil moisture retrievals via the synergistic use of synthetic aperture radar and optical remote sensing

Jianxiu Qiu[a*], Wade T. Crow[b], Wolfgang Wagner[c], Tianjie Zhao[d]

[a] *Guangdong Provincial Key Laboratory of Urbanization and Geo-simulation, School of Geography and Planning, Sun Yat-sen University, Guangzhou, 510275, China*

[b] *USDA ARS Hydrology and Remote Sensing Laboratory, Beltsville, MD 20705, USA*

[c] *Institute of Photogrammetry and Remote Sensing, Vienna University of Technology, 1040, Vienna, Austria*

[d] *State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, CAS, Beijing 100101, China*

Corresponding author: Jianxiu Qiu

E-mail: qiujianxiu@mail.sysu.edu.cn

Abstract

The recent launch of the Sentinel-1A and Sentinel-1B synthetic aperture radar (SAR) satellite constellation has provided high-quality SAR data with fine spatial and temporal sampling characterizations (6~12 revisit days at 10 m spatial resolution). When combined with high-resolution optical remote sensing, this data can potentially be used for high-resolution soil moisture retrieval over vegetated areas. However, the suitability of different vegetation index (VI) types for the parameterization of vegetation water content in SAR vegetation scattering models requires further investigation. In this study, the widely-used physical-based Advanced Integral Equation Model (AIEM) is coupled with the Water Cloud Model (WCM) for the retrieval of field-scale soil moisture. Three different VIs (NDVI, EVI, and LAI) produced by two different satellite sensors (Moderate Resolution Imaging Spectroradiometer (MODIS) and Landsat) are selected to examine their impact on the parameterization of vegetation opacity, and subsequently, on soil moisture retrieval accuracy. Results indicate that, despite the different sensitivity of estimated surface roughness parameters to various VIs (i.e., this sensitivity is highest when utilizing MODIS EVI and lowest in the LAI-based model), the optimum roughness parameters derived from each VI exhibit no discernible difference. Consequently, the soil moisture retrieval accuracies show no noticeable sensitivity to the choice of a particular VI. Generally, meadow and grassland sites with small differences in VI-derived roughness parameters exhibit good performance in soil moisture estimation. With respect to the relative components in the coupled model, the vegetative contribution to the scattering signal exceeds that of soil at VI about 0.8 [-] in NDVI-based models and 0.6 [-] in EVI-based models. This study provides insight into the proper

selection of vegetation indices during the use of SAR and optical imagery for the retrieval of high-resolution surface soil moisture.

Keywords: Sentinel-1; SAR; surface soil moisture; Advanced Integral Equation Model; Water Cloud Model; vegetation water content; Heihe River Basin.

1. Introduction

Soil moisture is a crucial nexus in the exchange of water, energy and carbon between the land surface and the lower atmosphere (Seneviratne et al., 2010). Water content within the surface and root-zone soil controls the partitioning of precipitation into runoff and infiltration, the partitioning of incoming radiation into latent and sensible heat fluxes, and CO_2 uptake by plants via transpiration. Based on its importance in linking these cycles, soil moisture is recognized as an Essential Climate Variable (GCOS, 2010), and knowledge of its spatial variation over heterogeneous regions is widely considered essential for understanding the effect of climate change on hydrological processes.

A new constellation of synthetic aperture radar (SAR) satellites, Sentinel-1A (launched in April 2014) and Sentinel-1B (launched in April 2016), provide free and publicly open SAR access with high spatial and temporal resolutions (6~12 revisit days at 10 m spatial resolution). As such, the Sentinel-1 constellation represents a major advance in the development of an operational soil moisture mapping capability at the field- to plot-scale level (Lievens et al., 2017; Li et al., 2018; Santi et al., 2018; Bao et al., 2018; Paloscia et al., 2013). In the past, SAR remote sensing has been widely used to estimate surface soil moisture (SSM) over bare soil surfaces using physical models (e.g., the Integral Equation Model (IEM; Fung et al., 1992), the Advanced Integral Equation Model (AIEM; Chen et al., 2003) and the Integral Equation Model for Multiple Scattering (Álvarez-Pérez, 2001)), empirical models (e.g., Dubois et al., 1995 and Oh et al., 1992), and semi-empirical models (e.g., Chen et al., 1995; Oh et al., 2002; Shi et al., 1997). For soils with moderate to dense vegetation cover, the direct scattering of vegetation, as well as the attenuation of upward soil scattering, cannot be neglected. In these circumstances, the accurate retrieval of SSM requires the coupling of vegetation and bare-soil scattering models. Common vegetation scattering models include the Water Cloud Model (WCM, Attema et al., 1978) and the Michigan Microwave Canopy Scattering Model (MIMICS, Ulaby et al., 1990). The latter has been demonstrated to be suitable for use in forests (McDonald et al., 1990).

Based on information from optical imagery, the above-mentioned vegetation scattering processes can be parameterized using various vegetation indices (VIs), such as the Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), or the Leaf Area Index (LAI) – thereby introducing the synergistic use of SAR and optical remote sensing data for the retrieval of surface soil moisture. Multiple studies have focused on differences in SSM estimation accuracy associated with the use of different VIs over a single land cover type using different SAR data sets, including TerraSAR-X and COSMO-SkyMed (Hajj et al., 2016), Radarsat-2 (Bai et al., 2016) and Experimental SAR (Lievens et al., 2011). However, relatively few studies have evaluated the robustness of different VIs for soil moisture retrieval over a wide range of land cover types.

In the present study, we selected the physically-based AIEM and WCM models to derive a coupled (soil/vegetation) microwave scattering model and utilized five different VI products to investigate their performances (as a proxy for vegetation opacity) in the coupled model. The five VIs differ with respect to both index type (NDVI, EVI, and LAI) and satellite source (Moderate Resolution Imaging Spectroradiometer (MODIS) and Landsat). They are utilized to examine if discrepancies in their spatial/temporal resolutions and sensor type will have a discernible impact on the accuracy of soil moisture retrieval results.

This paper is organized as follows. Section 2 introduces all data sets utilized for high-resolution soil moisture retrieval, including Sentinel-1 SAR imagery, optical remote sensing products for VIs derivations, and in-situ observations collected from the Heihe Watershed Allied Telemetry Experimental Research (HiWATER) program. This HiWATER program conducted in Heihe River Basin of Northwestern China is designed to be a comprehensive experiment to improve the observability of hydrological and ecological processes, to build a watershed observing system, and to enhance the applicability of remote sensing in integrated eco-hydrological studies and water recourse management at the basin scale (Li et al., 2017).

The parameterization of the coupled model, as well as metrics for evaluating soil moisture retrievals are also introduced in Section 2. The impact of VI selection on surface roughness parameter estimation during model establishment and its consequent impact on soil moisture retrieval accuracy are presented in Section 3. Following this, Section 4 reports on the sensitivity of roughness parameter to different VIs and the relative contribution of soil scattering within the coupled model when applying different VIs. Finally, major findings are presented in Section 5.

2. Materials and methodology

2.1 Data sets for soil moisture retrieval

2.1.1 Sentinel-1 SAR data

The Sentinel-1 satellites are equipped with C-band SAR instruments and have produced global observations since October 2014. Here, Level 1 ground range detected (GRD) Sentinel-1 interferometric wide (IW) observations with a VV polarization signals were used to retrieve soil moisture estimations, as this polarization has been proven to be less sensitive to volume scattering of vegetation cover than VH (Baghdadi et al., 2017; Patel et al., 2006; Chauhan et al., 2016). VH polarization records are only included for comparative purposes. The incidence angle of Sentinel-1 ranges between 30°~48°, and our study period is October 2014 to December 2017 (constrained by the temporal coverage of available in-situ measurements). All Sentinel-1 data were accessed through the Google Earth Engine (GEE) platform and pre-processed using the Sentinel-1 Toolbox to derive backscatter coefficients ($\sigma°$) in decibels (dB). The five processing steps can be summarized as follows (Hird et al., 2017):

1) **Apply orbit file;** applies the restituted orbit file to update orbital metadata;

2) **GRD border noise removal;** removes low-intensity noise and invalid data on edges of GRD scene;

3) **Thermal noise removal;** removes additive noise in sub-swaths to reduce discontinuities between sub-swaths for scenes in multi-swath acquisition modes (applied to images produced after July 2015);

4) **Radiometric calibration;** computes backscatter intensity using sensor calibration parameters in the GRD metadata;

5) **Terrain correction (orthorectification);** converts data from ground range geometry, which does not take terrain into account, to $\sigma°$ using the

SRTM 30-meter DEM for high latitudes (greater than 60° or less than -60°). A refined Lee speckle filter (Lee et al., 1999) with a 3 × 3 window size was subsequently applied to the time series of backscattering coefficients.

2.1.2 In-situ network observations

The in-situ observations in the present study were collected from the Heihe Watershed Allied Telemetry Experimental Research (HiWATER) program in the Heihe River Basin of Northwestern China. Between 2012 and 2017, HiWATER utilized simultaneous airborne, satellite-borne, and ground-based remote sensing experiments designed to address scaling issues associated with eco-hydrological processes via process study, modelling, and observation (Li et al., 2013; Li et al., 2017). As such, it provides multiscale data sets of meteorological elements and land surface parameters that facilitate the estimation of soil moisture over heterogeneous land surfaces (Liu et al., 2016; Xu et al., 2013). Fig. 1 shows the distribution of HiWATER in-situ sites with the Heihe Basin. The climate of the study area is semi-arid and prominent land cover/uses in the basin include: meadow, grassland, desert, forest, and cropland (see Table 1 for details).

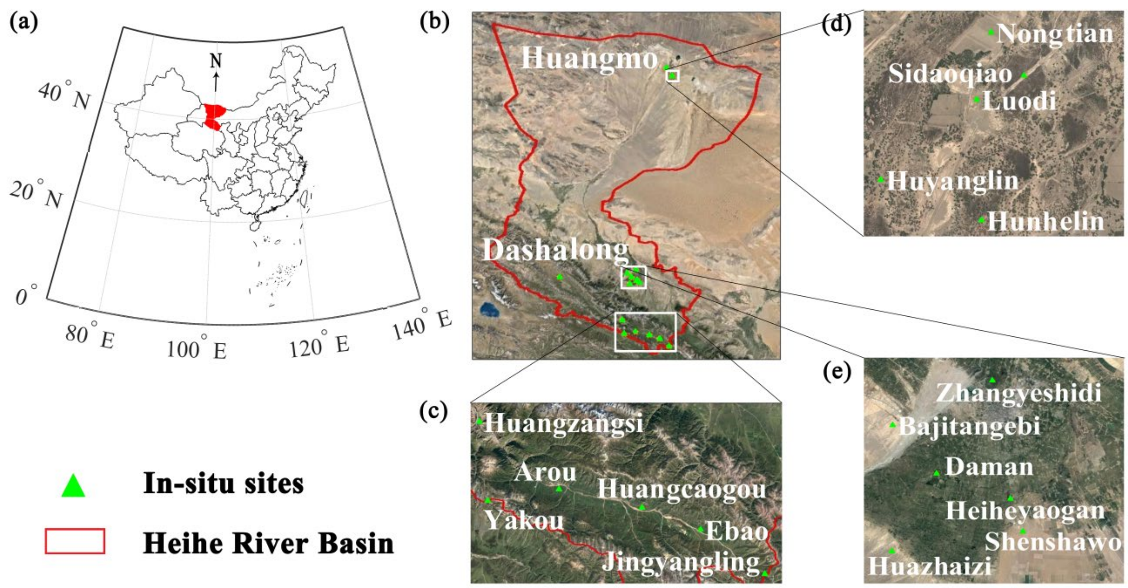


Fig. 1. The location of: (a) the Heihe River Basin in Northwestern China and (b) sub-basins/ground networks of interest within its (c) upper, (d) lower, (e) and middle reaches.

All 19 in-situ sites (see Fig. 1(b, c, d, and e)) were equipped with a set of automatic weather system and measured all components of the surface energy and water balances and associated near-surface atmospheric states. Observed variables include: precipitation, wind speed, air temperature, vapor pressure, net radiation, soil moisture, and temperature of the vertical soil profile (at 2, 4, 10, 20, 40, 80, 120, and 160 cm below the surface) at 10-mininute intervals. To better match the C-band penetration depth of the Sentinel-1 mission, the soil moisture and temperature measurements from the first layer (4-cm observations were used if 2-cm observations were missing) were used in this analysis. Soil moisture sensors included 200 SPADE and 150 Hydra Probe II instruments, which have instrument errors of 0.032 and 0.011 m^3m^{-3}, respectively. Land surface temperature (LST) sensors (SI-111) were calibrated using a BDB blackbody calibrator at a constant temperature of 23 °C and a water-ice mixture at 0 °C. The instrument error of SI-111 was within 0.15 °C. Additionally, soil samples were collected, and soil properties such as texture, bulk density and thermal and hydraulic parameters were analyzed in laboratory. This information was used as input for the Dobson model to estimate the soil dielectric constant (as introduced in Section 2.3). Following careful quality control, data sets collected as part of the HiWATER program have been made publicly available to the scientific community through the official project website (www.heihedata.org) (Li et al., 2017).

Table 1 Attributes of 19 HiWATER in-situ sites

Site name	**Longitude (°E)**	**Latitude (°N)**	**Land use**	**Temporal coverage**	**Sample number (VV+VH)**	**Sample number (VV)**	**SSM range (m^3m^{-3})**
Dashalong	98.9406	38.8399	Meadow	2013-2017	122	89	[0.06, 0.56]
Ebao	100.9151	37.9492	Grassland	2013-2016	71	69	[0.08, 0.32]
Yakou	100.2421	38.0142	Meadow	2015-2017	153	107	[0.07, 0.43]
Heiheyaogan	100.4756	38.827	Grassland	2015-2017	101	100	[0.01, 0.15]
Huazhaizi	100.3201	38.7659	Desert	2013-2017	81	80	[0.00, 0.23]
Huangmo	100.9872	42.1135	Desert	2015-2017	28	14	[0.02, 0.03]
Hunhelin	101.1335	41.9903	Forest	2013-2017	30	15	[0.02, 0.13]
Jinyangling	101.116	37.8384	Meadow	2013-2017	55	53	[0.06, 0.66]
Zhangyeshidi	100.4464	38.9751	Wetland	2013-2017	0	0	
Arou	100.4643	38.0473	Grassland	2013-2017	238	173	[0.07, 0.54]
Daman	100.3722	38.8555	Cropland	2013-2017	245	178	[0.03, 0.50]
Sidaoqiao	101.1374	42.0012	Forest	2013-2017	30	15	[0.08, 0.35]
Bajitangebi	100.3042	38.915	Desert	2013-2015	14	13	[0.04, 0.15]
Huyanglin	101.1239	41.9932	Forest	2013-2015	6	0	[0.01, 0.04]
Huangzangsi	100.1918	38.2254	Cropland	2013-2015	0	0	[0.06, 0.31]
Huangcaogou	100.7312	38.0033	Grassland	2013-2015	12	11	[0.10, 0.29]
Luodi	101.1326	41.9993	Bare land	2013-2015	6	0	[0.00, 0.01]
Nongtian	101.1338	42.0048	Cropland	2013-2015	2	0	[0.06, 0.06]
Shenshawo	100.4933	38.7892	Desert	2013-2015	12	11	[0.02, 0.08]

2.1.3 Vegetation indices

Five different vegetation indices (VIs) were utilized to account for vegetation condition and to investigate their impact on estimating soil moisture in our coupled scattering model. The VIs include products from MODIS (namely MODIS NDVI, http://dx.doi.org/10.5067/MODIS/MOD13Q1.006; MODIS EVI, http://dx.doi.org/10.5067/MODIS/MOD13Q1.006; and MODIS LAI http://dx.doi.org/10.5067/MODIS/MCD15A3H.006) and Landsat 8 (namely Landsat 8 NDVI and Landsat 8 EVI; Vermote et al., 2016). On the other hand, VI from the recently launched Sentinel-2 was not used as its temporal overlap (limited to only 2016-2017) with in-situ observations is not yet sufficient. All VIs were extracted through the GEE platform, and the pixel QA band was used to mask clouds from surface reflectance (SR) data.

To minimize the impact of different temporal interpolation methods on VI dynamics and soil moisture retrievals, we used VI products with temporal resolutions as uniform as possible, i.e., a 16-day product. In addition, the only MODIS LAI products available from the GEE platform are a 4-day and yearly product. Therefore, we used the former dataset in this analysis. Temporal gaps in the VI products were filled using a nearest-neighbor approach. All VI data used in the analysis were based on MODIS version 6 products. The specific characteristics of these data sets, including their product name (or calculation equation), spatial repeat, and temporal resolutions, are given in Table 2.

Table 2 The specific characteristics of the five VI data sets considered

VI	Product name/ Calculation equation	Spatial resolution	Temporal repeat
MODIS NDVI	MOD13Q1	250 m	16 day
Landsat 8 NDVI	$\frac{\rho_{\text{nir}} - \rho_{\text{red}}}{\rho_{\text{nir}} + \rho_{\text{red}}}$	30 m	16 day

MODIS EVI	MOD13Q1	250 m	16 day
Landsat 8 EVI	$2.5(\rho_{\text{nir}} - \frac{\rho_{\text{red}}}{\rho_{\text{nir}} + 6\rho_{\text{red}} - 7.5\rho_{\text{blue}}})$	30 m	16 day
MODIS LAI	MCD15A3H	500 m	4 day

* ρ_{nir}, ρ_{red} and ρ_{blue} denote SR of near-infrared, red and blue bands in Landsat 8.

2.2 Microwave scattering model and soil moisture retrieval

In this study, the first-order radiative transfer model WCM (Attema et al., 1978) was used to simulate the backscattered radar signal over vegetated sites. This semi-empirical model is widely applied in its simplified form due to its efficient performance (Zribi et al., 2011; Gherboudj et al., 2011; Paloscia et al., 2013). For a given polarization, the WCM considers the radar signal as the linear sum of contribution from the vegetation ($\sigma^{\text{o}}_{\text{veg}}$), the soil ($\sigma^{\text{o}}_{\text{soil}}$) – as attenuated by vegetation ($\tau^2\sigma^{\text{o}}_{\text{soil}}$):

$$\sigma^{\text{o}}_{\text{sim}} = \sigma^{\text{o}}_{\text{veg}} + \tau^2\sigma^{\text{o}}_{\text{soil}} \tag{1}$$

$$\sigma^{\text{o}}_{\text{veg}} = \mathrm{A}V_1\cos\theta\,(1 - \tau^2) \tag{2}$$

$$\tau^2 = \exp\,(-\,2\mathrm{B}V_2\,/\,\cos\theta) \tag{3}$$

where V_1 and V_2 are vegetation descriptors that indicate direct canopy backscattering and vegetation attenuation respectively; θ is the radar incidence angle; A and B are the fitted model coefficients which depend on the vegetation descriptor and radar configuration, and τ^2 is the two-way vegetation attenuation.

As commonly assumed in applying (1-3), multiple soil-vegetation scatterings are neglected here and the parameter V_1 is set equal to V_2. This simplifies Eqs. (1-3) to:

$$\sigma^{o}_{\mathrm{sim}} = \mathrm{A}V\cos\theta\,(1-\tau^2) + \tau^2\sigma^{o}_{\mathrm{soil}}\,. \tag{4}$$

The soil contribution $\sigma^{o}_{\mathrm{soil}}$ is simulated using the physically-based microwave scattering model AIEM, which is widely reported to perform well over bare soil surfaces (Wu et al., 2004; He et al., 2017; Zeng et al., 2017). The AIEM forward model requires input parameters describing: 1) sensor configuration: radar frequency (~5.405 GHz for Sentinel-1 C band); incidence angle (range from 30°~48° as specified by Sentinel-1), and polarization mode (VV); 2) surface parameters: soil dielectric constant, root mean square (RMS) height (s), correlation length (cl), and the auto-correlation function ACF.

Here, the Dobson dielectric mixing model was used to determine the relationship between dielectric constant ε_m and soil moisture m_v, in the following form:

$$\varepsilon_m^{\alpha} = 1 + \frac{\rho_b}{\rho_s}(\varepsilon_s^{\alpha} - 1) + m_v^{\beta}\varepsilon_{fw}^{\alpha} - m_v \tag{5}$$

where ρ_b is soil bulk density; ρ_s is characteristic specific density, and ε_s and ε_{fw} are relative permittivities of soil solids and Gouy-layer water. Parameters α and β are optimized constants and the latter is assumed to be soil-texture dependent. In-situ measurements of bulk density, silt, clay, sand percentages, and soil surface temperature/moisture at each site were fed into this model to estimate soil surface dielectric values required as input by the AIEM. Based on previous measurements of surface roughness parameters acquired during simultaneous ASAR observations (Chen et al., 2017), the s and cl in the study area were constrained between [0, 3.0] cm and [0, 20.0] cm, respectively. The increments for these two parameters were set as 0.2 cm and 2 cm, respectively. For each iteration of s and cl combinations, the vegetation parameters A and B were calibrated by minimizing the cost function $\boldsymbol{J}$ constructed by root-mean-square error (RMSE) of the simulated vegetation backscattering coefficients

$\sigma^{o}_{\rm sim}$ (evaluated against observations from Sentinel-1, $\sigma^{o}_{\rm obs}$) in VV and VH polarizations.

$$\boldsymbol{J} = \sqrt{\frac{1}{n}\sum\left(\sigma^{o}_{\rm sim} - \sigma^{o}_{\rm obs}\right)^{2}}\,. \tag{6}$$

Consequently, the optimum surface roughness parameters *s* and *cl* were selected based on the minimization of backscatter RMSE (across all iterations). For each site, we used the *K*-fold (*K*=10) cross validation method that takes the mean of the *K*-fold validation results to estimate model parameters and evaluate algorithm accuracy. The flowchart of this retrieval process is shown in Fig. 2.

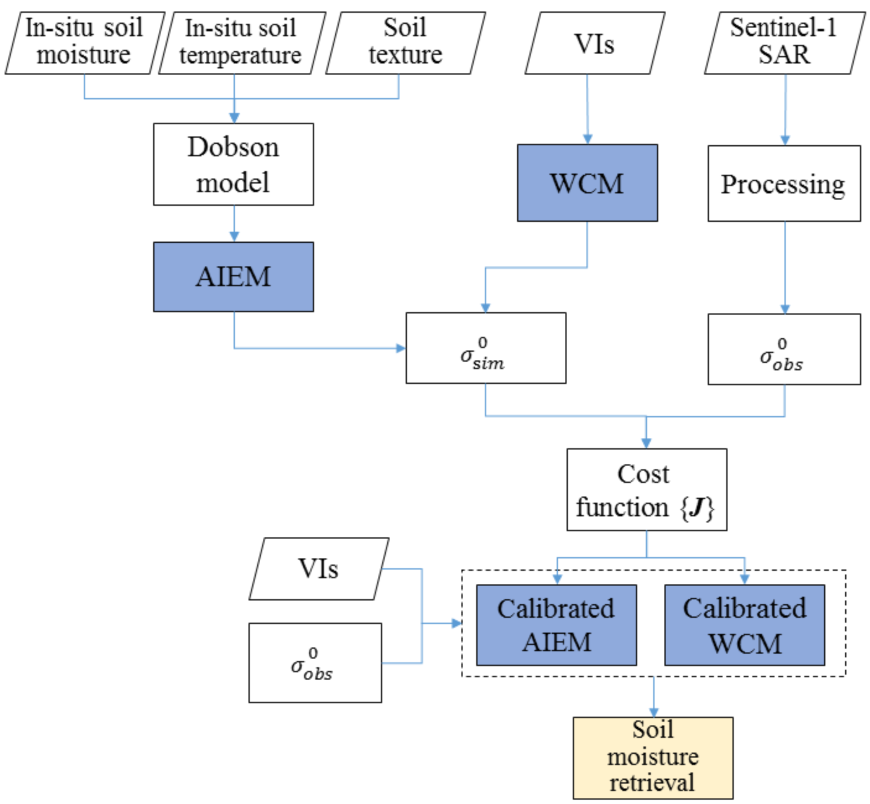


Fig. 2. Flowchart of our soil moisture retrieval process.

2.3 Soil moisture accuracy metrics

In addition to the commonly-used root mean square error (RMSE) and Pearson product-moment correlation coefficient, we also applied mutual information (MI, Cover and Thomas, 1991) to assess the accuracy of soil moisture estimation. MI is a nonparametric measure of correlation (here defined strictly as the lack of independence) between two random variables, and represents the reduction of entropy (uncertainty) in either variable given knowledge of the other. It is a more rigorous measure compared to commonly-used metrics such as Spearman's rank correlation coefficient and Pearson correlation coefficient - the latter being an approximation of MI under certain conditions (Nearing et al., 2015).

Here, we calculate the MI content between retrieved soil moisture and in-situ measurements in each site. Estimated MI is normalized by the entropy of the corresponding in-situ measurements to remove the effect of inter-site variation on the magnitude of difference, and the normalized MI (NMI) represents the fraction of uncertainty in ground observations that is resolvable given knowledge of the soil moisture retrievals or simulations (Nearing et al. 2013). For details on MI estimation, please refer to Qiu et al. (2014; 2016).

3. Results

3.1 Intercomparison of different VIs over in-situ sites

We first compare the five VIs collected from 19 in-situ sites. All were extracted at their original spatial resolution and without any spatial resampling procedures. Seen from the temporal evolution of the extracted VI values at individual site (please see Section 4.1), the dynamics of VIs from dataset with different scales are generally very consistent. Direct VI comparisons are shown in Fig. 3, with point density indicated by color shading. It is clear from Fig.

3(a, h) that, for the same VIs from different instruments (Landsat 8 and MODIS), all points are evenly scattered along the 1:1 line – with no apparent systematic bias. For different VIs acquired from the same instrument (Fig. 3(b, f)), this 1:1 agreement persists for low-vegetation points. However, a sigmoid shape is observed at high levels of vegetation, suggesting that EVI is more responsive to vegetation variations than NDVI during the peak of growing season. This pattern persists for different VIs acquired from different sensors (Fig. 3(c, e)), except that the points are more scattered. This is in line with our prior expectations and justifies the selection of EVI as a VI candidate in our comparison study.

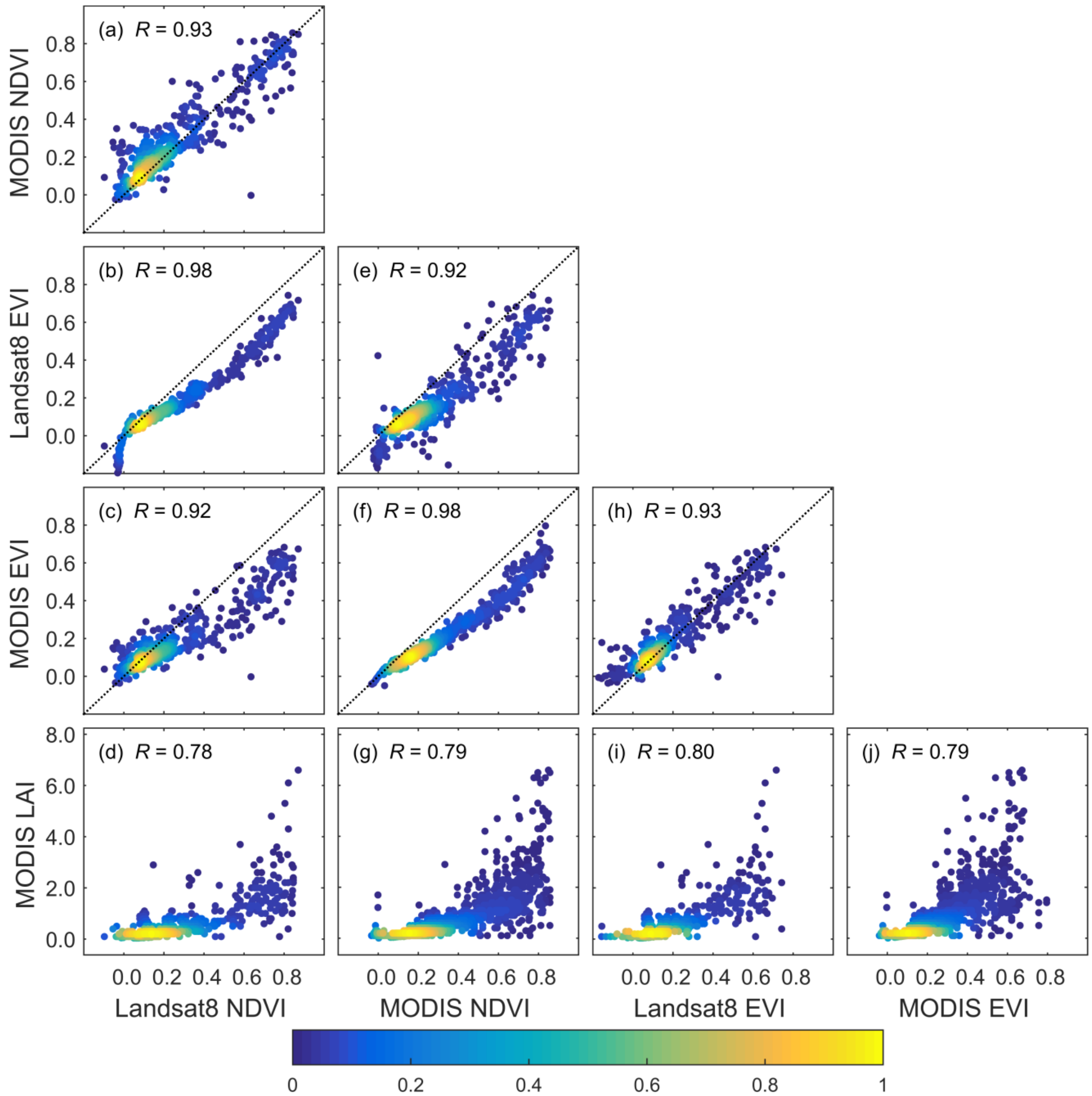


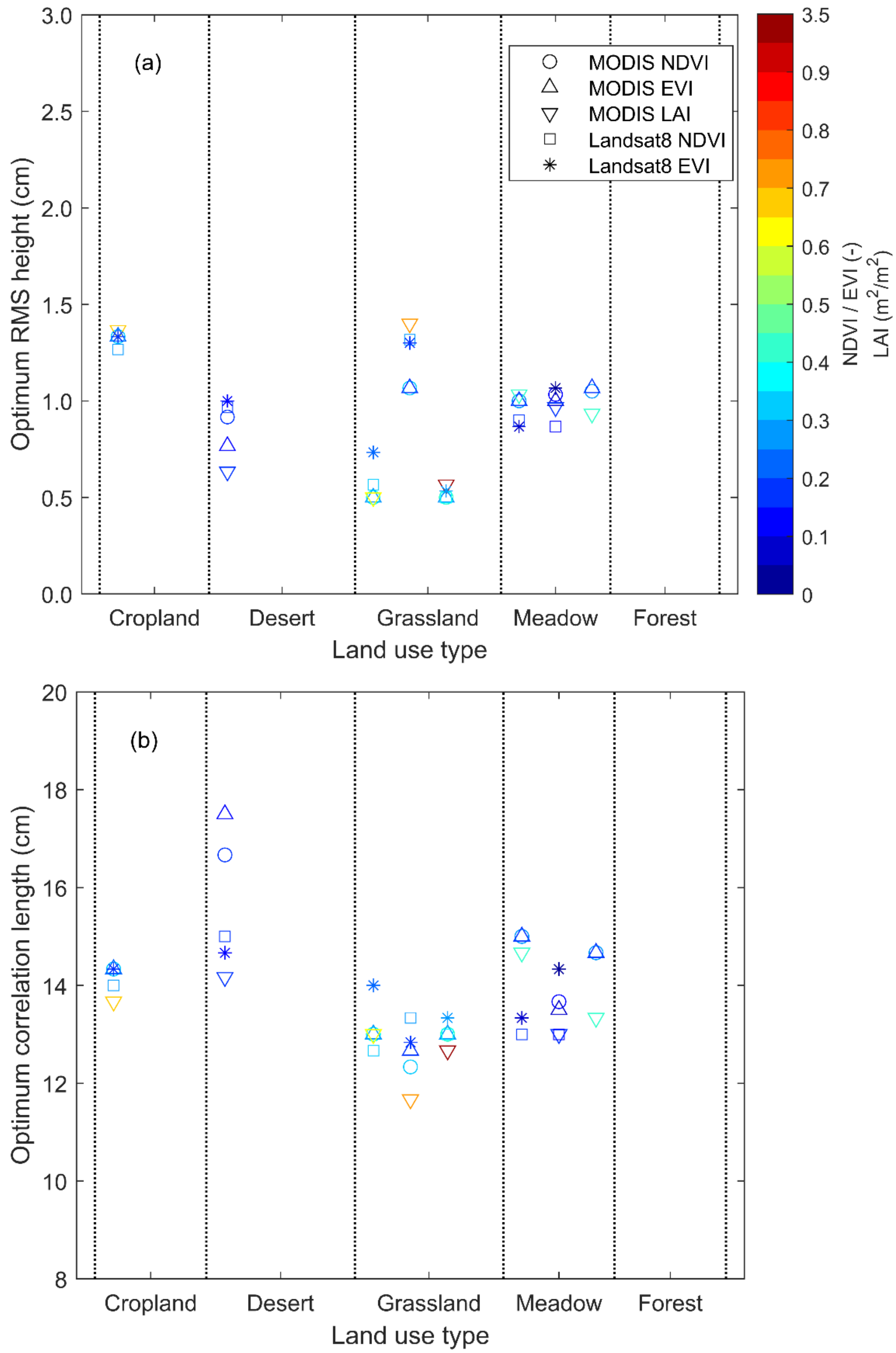
Fig. 3. Scatterplots comparing five VIs collected from 19 in-situ sites, with points density indicated by color shading.

3.2 Impact of VIs selection on surface roughness parameter estimations

As the inclusion of VH polarization in the cost function has slightly reduced the soil moisture retrieval accuracy (compare Fig. 5 in Section 3.3 and Fig. A1 in the Appendix), all analyses were conducted using only the VV polarization record. In addition, the observation sample size could greatly affect the soil moisture retrieval - since insufficient sample size could lead to erroneous model parameter estimation. Therefore, we set the threshold of sample number for each site to be 15 (in VV polarization). Sites with fewer observations generally failed to converge to a unique *A*, *B*, *sig* and *cl* solution. The observation sample size for the MODIS NDVI product at each site are listed in Table 1. Sample sizes for other MODIS VI products (EVI and LAI) are very similar but decrease significantly for Landsat 8.

Using the above-mentioned five VI products in the coupled AIEM and WCM models (introduced in Section 2.3), we estimated the optimum surface roughness parameters for each site by minimizing the cost function in Eq.(5). Final optimized parameters were obtained by taking the mean of the 6 candidate parameter sets (i.e., *s*, and *cl*) achieving the lowest value of the cost function in Eq. (5). Optimized *s*, *cl* and effective roughness (s^3/cl^2) values are summarized in Fig. 4, with different land use types separated by vertical dash lines. Parameters estimated by different VI products are indicated by different marker symbols, and annual mean vegetation cover conditions are captured using color shading.

As Landsat 8 VI observations are much temporally sparser than MODIS observations at some sites, they did not always provide sufficient observation numbers (>=15) for parameter estimation and soil moisture retrieval. Thus, even with identical number of sites, some Landsat 8 results are missing in Fig. 4. For meadow and grassland types (i.e., sites with comparatively higher vegetation cover, generally exhibit lower optimum RMS height, and consequently, lower effective roughness) the variation of optimum correlation length is comparatively less sensitive to variations in VIs (Fig. 4b). Overall, variations in surface roughness parameters between each land use group are more significant than those seen between various VI products.



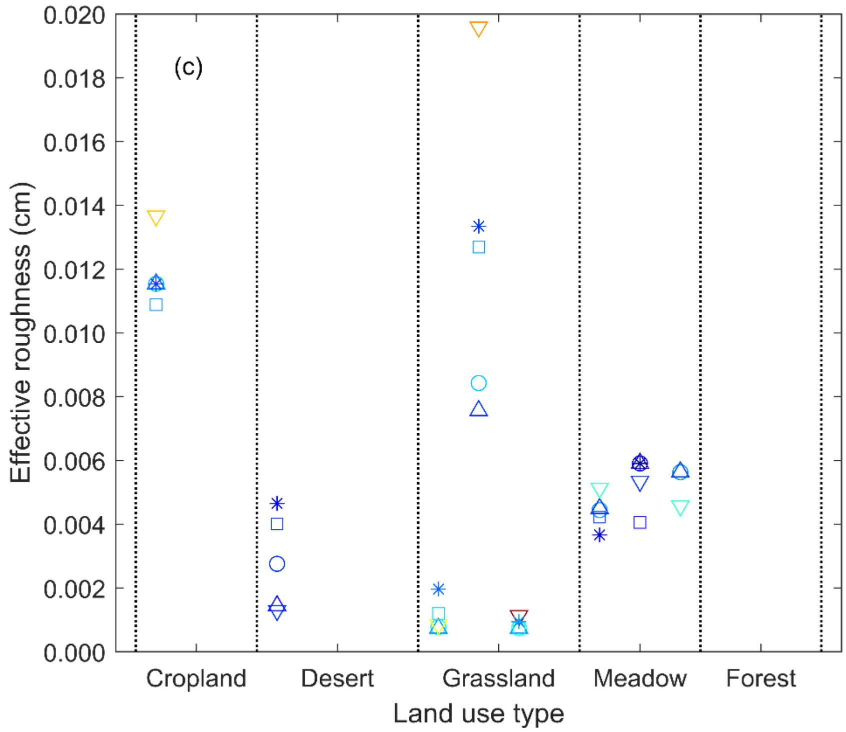


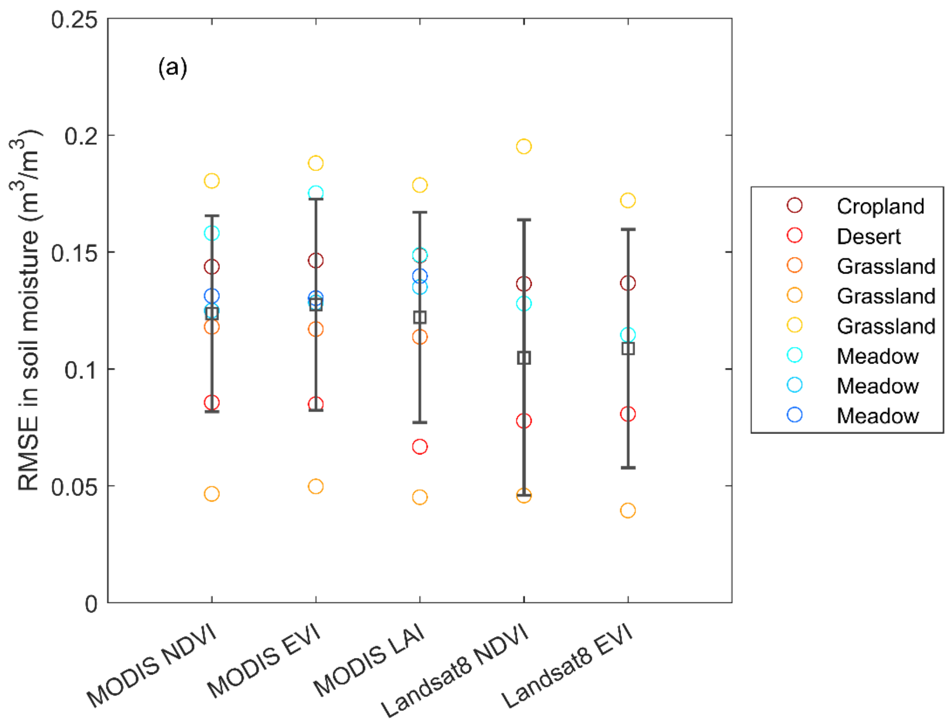
Fig. 4. Surface roughness parameters estimated by each VI product for all sites, with different land use types separated by dashed vertical lines. VI type is indicated by different marker, while annual mean VI values are indicated by color shading. Surface roughness parameters include: (a) optimum RMS height (s, cm), (b) optimum correlation length (cl, cm), and (c) effective roughness (cm)

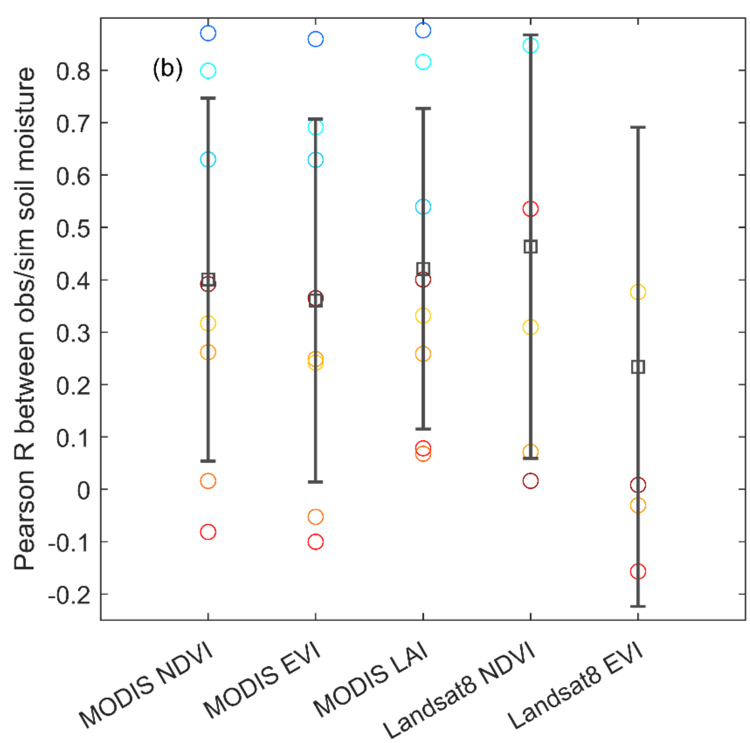
3.3 Impact of VIs selection on soil moisture estimations accuracy

Using the estimated optimum surface roughness parameters, we retrieved soil moisture from the coupled model and evaluated them against in-situ observations in terms of RMSE, Pearson correlation coefficient, and normalized mutual information (NMI). These results are shown in Fig. 5. Each site is marked with an identical color, so that differences in ranking (among all sites) between each VI data set can be clearly observed.

There is no significant difference in soil moisture estimation accuracy associated with different VIs calculated from MODIS products. The RMSE rankings among all sites are quite close between different VIs - indicating barely discernible differences in SM retrieval accuracy (Fig. 5a). SM retrieval performance does differ somewhat between VI products derived from Landsat 8 versus MODIS observations - likely due to the reduced temporal sampling of Landsat 8 VI products. As some sites lacking sufficient observation samples from Landsat 8 VI, the coupled model cannot be established and soil moisture retrievals are missing. It is worth noting retrievals at certain sites (e.g., the desert and the grassland site Heiheyaogan shown as the second site in the grassland column in Fig. 4c) with an observable discrepancy in VI-derived effective roughness parameters, and very limited soil moisture variability (SSM range <0.23 m^3m^{-3} in Table 1), perform poorly for all three evaluation metrics. Specifically, the Heiheyaogan site exhibits a Pearson R of approximately 0.2 and demonstrates the lowest observed NMI.

In addition, soil moisture retrieval accuracy based on microwave scattering model is closely related to vegetation cover conditions. For instance, meadow sites with lower LAI demonstrate generally higher Pearson R and higher NMI (Fig. 5(b, c)) than grassland sites with higher LAI. On the other hand, the small temporal variability of soil moisture at the Ebao and Heiheyaogan grassland sites (SSM ranges are 0.24 to 0.14 m^3m^{-3} respectively, please see Table 1) results in lower RMSE than in the meadow sites (Fig. 5a). Besides this analysis on the original SSM time series, we also conducted evaluations using short-term SSM anomalies (i.e. variations relative to a 32-day moving average window). Relative to our original results, these anomaly-based results reveal a slight decrease in Pearson R. However, overall sensitivity to VI selection remained low.





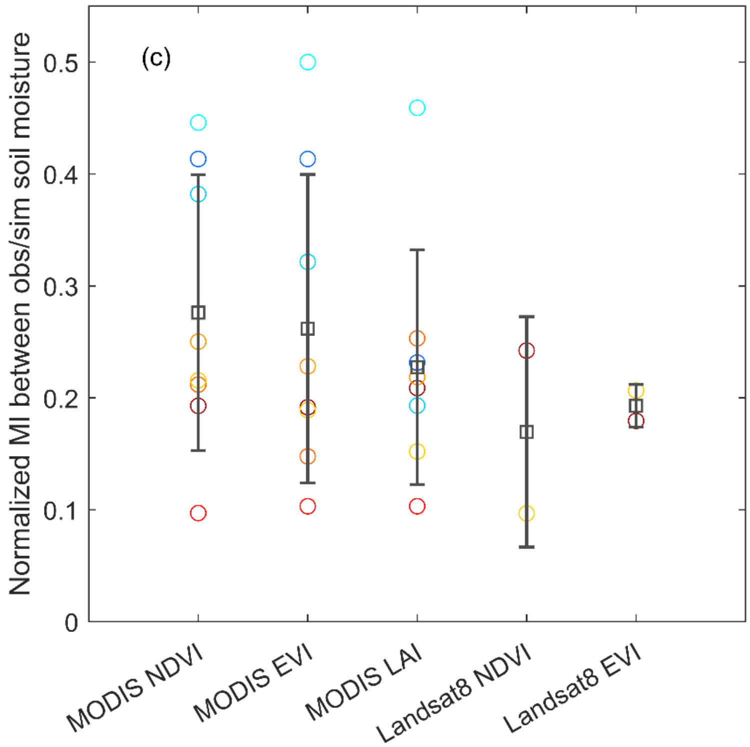
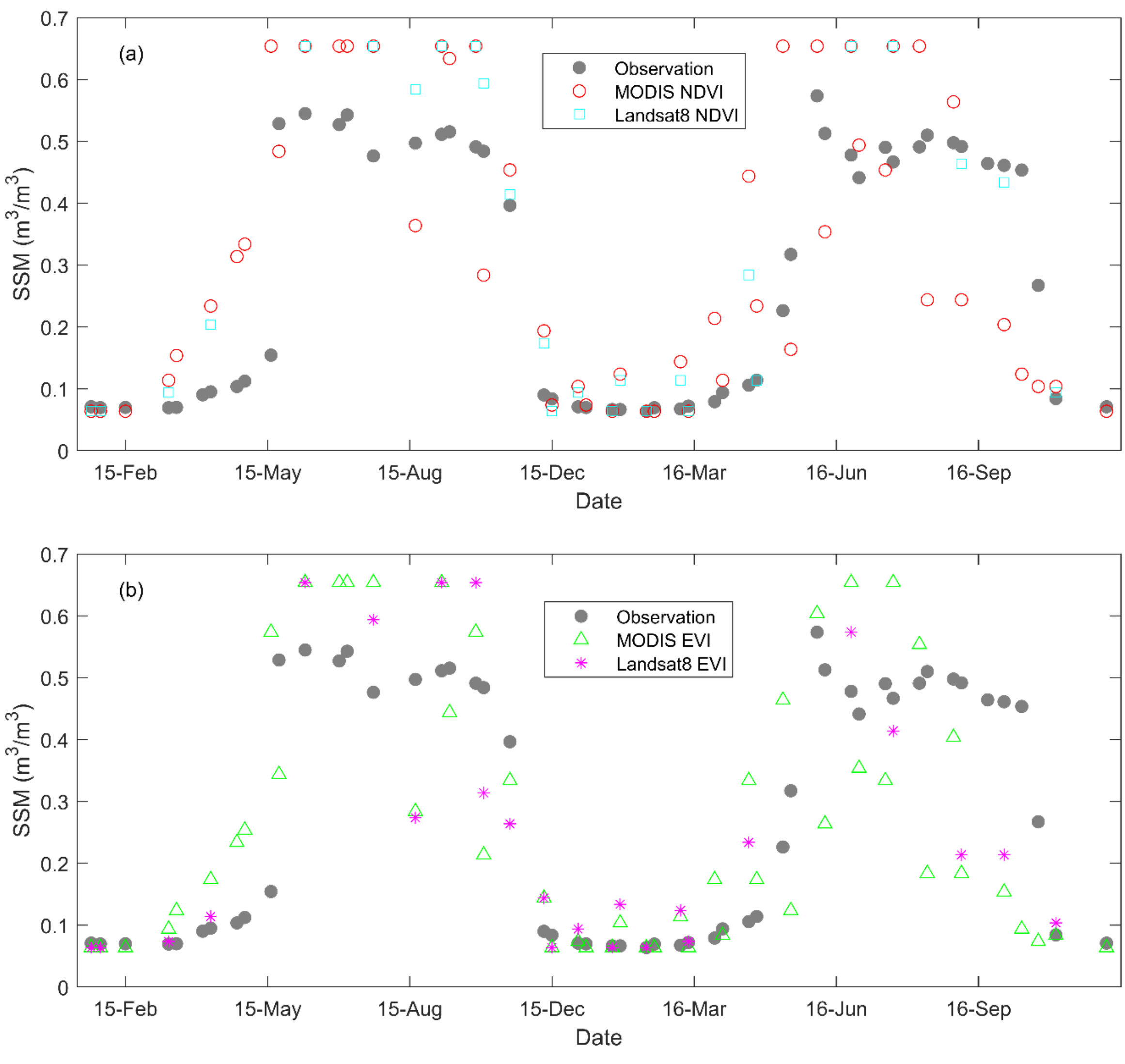


Fig. 5. Accuracy of soil moisture estimation derived from five VI data sets, based on (a) RMSE, (b) Pearson correlation coefficient, and (c) Normalized MI. Each site is marked with a unique symbol color.

Apart from the aggregated performance from all eligible sites, we also examined retrieved SSM dynamics at individual sites. For instance, the temporal evolution of retrieved SSM at the Jinyangling site for all five VIs are shown in Fig. 6. The seasonality of SSM time series from five VIs are similar and, in general, properly captured by all five SSM retrievals. On the other hand, the short-term variabilities of SSM retrievals occasionally deviate from that of observations, as the rapid fluctuation of point-scale SSM cannot be adequately captured by pixel-scale SSM retrievals. The discrepancy observed in Fig. 3 consequently lead to differences in SSM retrievals seen in Fig. 6. Specifically, as opposed to NDVI (Fig. 6a), VI types less prone to saturation at high vegetation levels, such as EVI and LAI, do not result in the levelling off in SSM retrievals (especially during June to July of 2016 in Fig. 6(b, c)). In addition, SSM retrieval differences between different sensors for the same VI (i.e., MODIS NDVI vs. Landsat 8 NDVI and MODIS EVI vs. Landsat 8 EVI) are less substantial than differences between different VI from the same sensor.



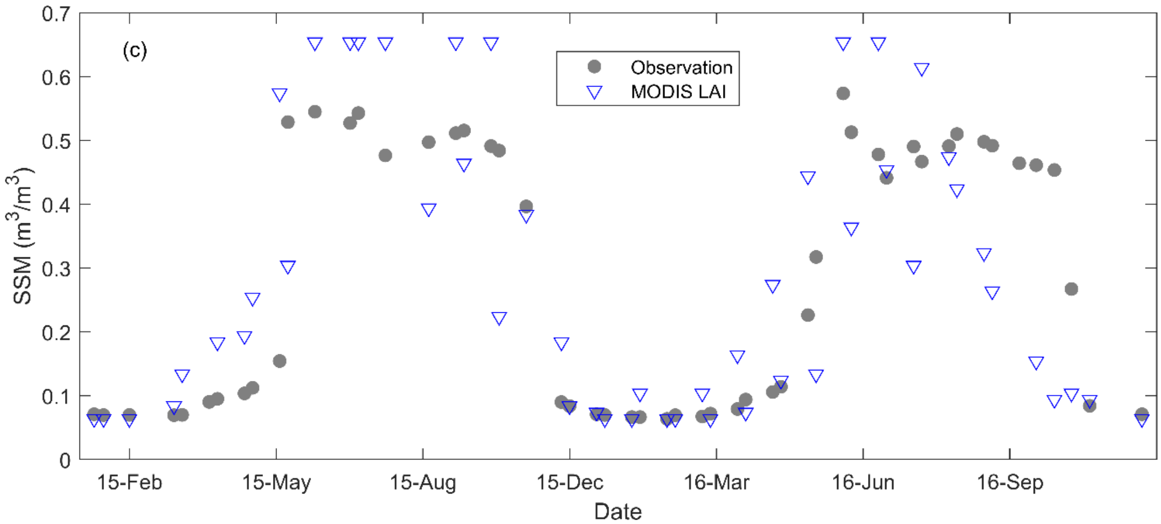


Fig. 6. The temporal evolution of SSM retrieved for the Jinyangling site based on: (a) NDVI, (b) EVI, and (c) LAI acquired from different sensors and in-situ observations.

Therefore, all five coupled models can retrieve similar optimum surface roughness parameters (except for the grassland site Heiheyaogan in the LAI-constructed model, see Fig. 4 for details), and consequently, achieve similar soil moisture estimation accuracy. Unlike previous investigations based on single land use types (Lievens et al., 2011; Bai et al., 2016), this study cannot generally recommend any single VI for soil moisture retrieval in the coupled microwave models – as the optimal VI choices varies across different land cover types. It should also be noted that the overall performance of the coupled model varies from site to site. These variations are related to changes in data sample size and the range of observed SSM at each site.

It should be noted that our overall SSM retrievals accuracies are relatively low (e.g., correlation values tend to be below 0.4 [-]). This suggests that our algorithm actually captures only bulk seasonal patterns, which are likely to be highly correlated across different VIs. It is possible that other, more accurate, approaches could reach slightly different conclusions regrading sensitivity to VI choice. For instance, instead of using a physically-based AIEM, Bao et al. (2018) employed a best-fitting regression method to directly estimate soil moisture measurement using different VIs. They retrieved more accurate soil moisture retrievals and found slightly higher sensitivity to VI choice.

4. Discussion

4.1 Sensitivity of surface roughness parameters to different VIs

Above (Section 3.2), we examined the impact of VI selection on optimum roughness parameter estimations. In this section, we will examine the grassland site Ebao in greater detail to further investigate the sensitivity of calibrated surface roughness parameters to VI dataset choice. First, the temporal evolution of the five VI sets in Ebao is shown in Fig. 7. These time series reflect similar seasonal phasing, although MODIS-based LAI exhibits much less temporal variation during low-biomass seasonal periods.

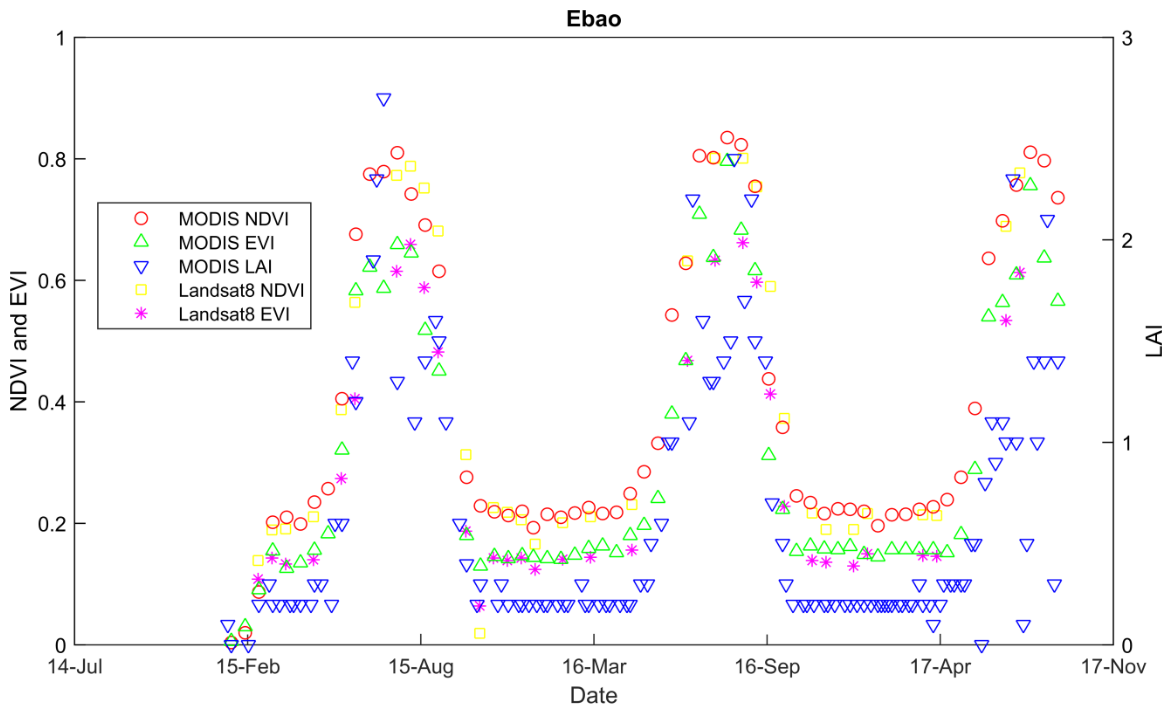


Fig. 7. The temporal variation of all five VI data sets at the Ebao grassland site.

Then, RMSE of the backscattering coefficient in the coupled AIEM and WCM models established by different VI data sets are shown in Fig. 8. The RMSE surface is masked (in white) for cases in which the optimization of the cost function ***J*** could not converge to a unique solution for parameters *A* and *B*. We can see that different combinations of surface roughness parameters (i.e., different combinations of correlation length and RMS height) can result in identical performance for the coupled model. This convincingly demonstrates the ill-posed nature of the soil moisture inversion problem for microwave scattering modeling. In addition, it is seen that errors in the coupled models - associated with different VIs - have different sensitivities to variations in *s* and *cl*. Generally, this sensitivity is highest when the model parameterized by MODIS EVI and the lowest for the LAI-based model.

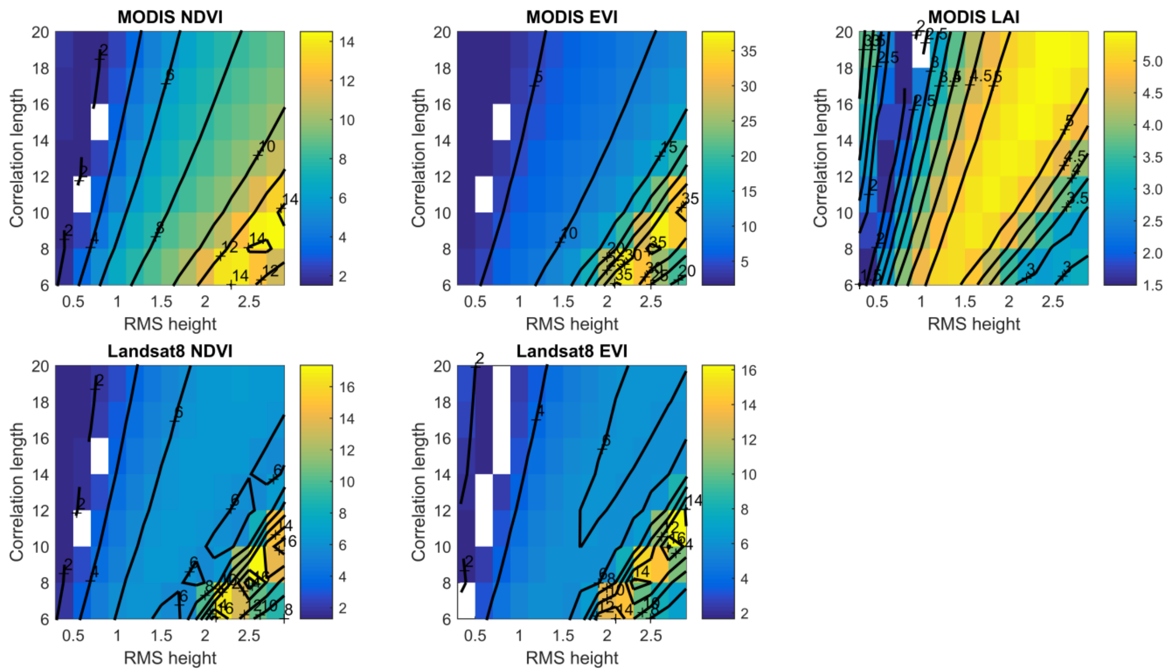


Fig. 8. At the Ebao grassland site, RMSE surface (as a function of *s* and *cl*) for backscattering coefficient estimates (units: dB) provided by the coupled AIEM and WCM models associated with different VI data sets.

4.2 Contribution of bare soil to observed scattering with different VIs

Using five different VI products, the soil and vegetation backscattering contributions (σ^o_{veg}, $\tau^2\sigma^o_{\text{soil}}$) of five-established coupled models are investigated in this section. To better facilitate inter-model comparisons, we plotted the ratio of the soil backscattering contribution ($\tau^2\sigma^o_{\text{soil}}$) to total scattering signal ($\sigma^o_{\text{veg}} + \tau^2\sigma^o_{\text{soil}}$) as a function of VI for all eligible sites in Fig. 9. The relationship at each site are fitted with an exponential regression with high goodness of fit (all $R^2 > 0.8$). To improve the readability of figure after curve fitting, only one-tenth of the data at each site are randomly selected and plotted. Sites with less than 60 sample data are not considered.

Results in Fig. 9 demonstrate that, for VV polarization within the incidence angle of 30°~48° (typical case for Sentinel-1 SAR imagery over the study area), the contribution of soil to the total backscattering coefficient decreases with increasing VI, as expected. In addition, regardless of product type, the demarcation value for vegetation's contribution exceeds soil's contribution (ratio of 0.5 [-]) is approximately 0.8 for NDVI (Fig. 9(a, b). This is in line with numeric simulations of the coupled IEM and WCM model for a grassland site in French (Baghdadi et al., 2017). This threshold value of VI decreases to 0.6 [-] for EVI (both MODIS and Landsat 8 products) and increases to above 2.0 [-] for LAI. The desert site Huazhaizi shows very little variation in vegetation cover and is thus excluded from consideration.

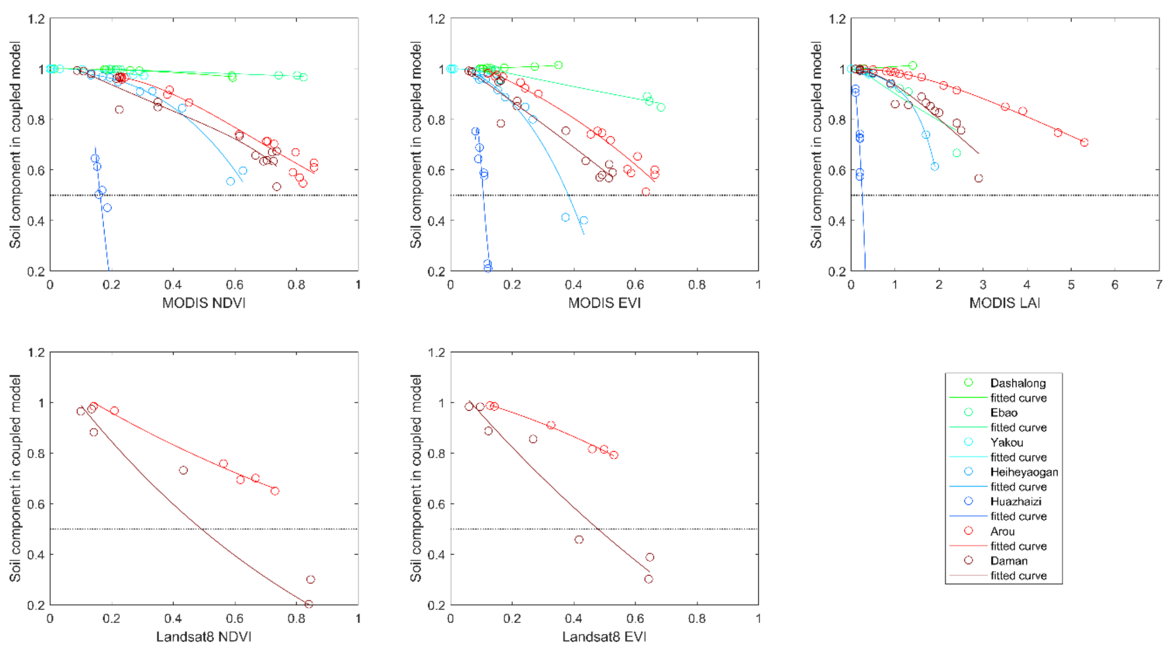


Fig. 9. The ratio of soil contribution to total scattering ($\frac{\tau^2\sigma^0_{\text{soil}}}{\sigma^0_{\text{veg}}+\tau^2\sigma^0_{\text{soil}}}$) for all eligible sites using 5 different VI products. All sites are fitted with a regression

equation in an exponential form [a*exp(bx)+c*exp(dx)] with high goodness of fit (all $R^2 > 0.8$). Sites with less than 6 samples are not plotted.

In addition, the impact of SSM range on soil's contribution to the scattering model can also be observed in Fig. 9. It is noted by comparing the curvatures of different sites, that the sensitivity of soil contribution to VI is higher in sites with a smaller SSM range (e.g., Heiheyaogan) than sites with a larger SSM range (e.g., Dashalong and Arou). This variation in the sensitivity of soil contribution is noticeable across five different VI data sets.

5. Summary and conclusion

In this study, we applied a coupled microwave scattering model (consisting of AIEM and WCM) to retrieve soil moisture from Sentinal-1 SAR images in the Heihe River Basin. Five separate vegetation products (MODIS NDVI, Landsat 8 NDVI, MODIS EVI, Landsat 8 EVI, and MODIS LAI) are used as vegetation descriptor in the model to investigate their effectiveness in retrieving soil moisture.

Comparison of the selected five VIs over all in-situ sites showed no systematic bias in any VI data set, while EVI and LAI are more responsive to vegetation variation in the high VI range, and consequently reduced the levelling off phenomenon observed in soil moisture retrieval based on NDVI during peak growing season. Despite their discrepancies, optimum surface roughness parameters (including RMS height, correlation length and effective roughness) derived from all five VI data sets do not show any noticeable difference except for one grassland site with very limited SSM variability. In terms of retrieved soil moisture accuracy, sites with distinctly different VI-derived roughness parameters showed the lowest accuracy in terms of RMSE, Pearson correlation and NMI.

A detailed comparative study was conducted in site Ebao to examine the sensitivity of surface roughness parameters to different VIs. It is observed that sensitivity is highest in the coupled model established by MODIS EVI, while lowest in the LAI-based model. Furthermore, in different VI-established models, the threshold at which the vegetation contribution dominates the total scattering signal differs significantly. The demarcation value at which point vegetation's contribution exceeds that of the soil is approximately 0.8 [-] for NDVI (regardless of what sensor is used for NDVI) (Fig. 9(a, b)). This value decreases to 0.6 [-] for MODIS EVI and Landsat 8 EVI products.

It should be addressed that this work is based on the assumption of equal V_1 and V_2 parameters in the WCM. A detailed analysis based on utilizing a combination of different VIs as vegetation descriptors for V_1 and V_2 should be considered for future study.

Acknowledgment

This work was supported by National Natural Science Foundation of China (Grant No. 41501450), Natural Science Foundation of Guangdong Province, China (Grant No. 2016A030310154) and the Fundamental Research Funds for the Central Universities (No. 16lgpy06).

We are very thankful to the researchers in HiWATER-MUSOEXE for their efforts on data acquisition and sharing. The data set used in this study is provided by the Heihe Plan Science Data Center, National Natural Science Foundation of China (http://www.heihedata.org).

References

Álvarez-Pérez, J. L, 2001: An extension of the IEM/IEMM surface scattering model. *Wave Random Complex*, **11**(3), 307-329, doi:10.1080/13616670109409787.

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

Baghdadi, N., El Hajj, M., Zribi, M., & Bousbih, S, 2017: Calibration of the water cloud model at C-Band for winter crop fields and grasslands. *Remote Sens-Basel.*, **9**(9), 969, doi:10.3390/rs9090969.

Bai, X., He, B., & Li, X, 2016: Optimum surface roughness to parameterize advanced integral equation model for soil moisture retrieval in prairie area using Radarsat-2 data. *IEEE T. Geosci. Remote.*, **54**(4), 2437-2449, doi:10.1109/TGRS.2015.2501372.

Bao, Y., Lin, L., Wu, S., Deng, K. A. K., and Petropoulos, G. P., 2018: Surface soil moisture retrievals over partially vegetated areas from the synergy of Sentinel-1 and Landsat 8 data using a modified water-cloud model. Int. *J. Appl. Earth. Obs. Geoinf.*, **72**, 76-85.

Chauhan, S., & Srivastava, H. S, 2016: Comparative evaluation of the sensitivity of multi-polarised sar and optical data for various land cover classes. *Int. J. Remote Sens., GIS and Geo.*, 2016, 4, 01-14.

Chen Luwan, Han Ling, Qin Xiaobao, Zhang Wu, A new method for constructing land surface combined roughness parameter in the process of soil moisture retrieval by microwave remote sensing, Geography and Geo-Information Science, 2017, 33(6), 37-43 (in Chinese).

Chen, K.S., Wu, T.D., Tsang, L., Li, Q., Shi, J.C., and Fung, A.K, 2003: Emission of rough surfaces calculated by the Integral Equation Method with comparison to three-dimensional moment method simulations. *IEEE T. Geosci. Remote.*, 2003, **41**(1): 90–101, doi:10.1109/36.134085.

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

Cover, T. M., and J. A. Thomas, 1991: *Elements of information theory*. John Wiley & Sons, New York.

Dubois, P.C., van Zyl, J.J., and Engman, E.T, 1995: Measuring soil moisture with imaging radar. *IEEE T. Geosci. Remote.*, 1995, **33**(4): 915–926, doi:10.1109/36.406677.

El Hajj, M., Baghdadi, N., Zribi, M., Belaud, G., Cheviron, B., Courault, D., & Charron, F, 2016: Soil moisture retrieval over irrigated grassland using X-band SAR data. *Remote Sens. Environ.*, **176**, 202-218, doi:10.1016/j.rse.2016.01.027.

Fung, A.K., Li, Z., and Chen, K.S, 1992: Backscattering from a randomly rough dielectric surface. *IEEE T. Geosci. Remote.*, 1992, 30: 356–369.

GCOS. 2010. Implementation plan for the Global Observing System for Climate in support of the UNFCCC report. World Meteorological Organization, Geneva Switzerland.

Gherboudj, I., Magagi, R., Berg, A. A., & Toth, B, 2011: Soil moisture retrieval over agricultural fields from multi-polarized and multi-angular RADARSAT-2 SAR data. *Remote Sens. Environ.*, **115**(1), 33-43, doi:10.1016/j.rse.2010.07.011.

He, L., Chen, J. M., & Chen, K. S, 2017: Simulation and SMAP observation of sun-glint over the land surface at the L-band. *IEEE T. Geosci. Remote.*, **55**(5), 2589-2604, doi:10.1109/TGRS.2017.2648502.

Hird, J. N., DeLancey, E. R., McDermid, G. J., and Kariyeva, J., 2017: Google Earth Engine, open-access satellite data, and machine learning in support of large-area probabilistic wetland mapping. *Remote Sens*., **9**(12), 1315.

Lee, J. S., Grunes, M. R., and De Grandi, G., 1999: Polarimetric SAR speckle filtering and its implication for classification. *IEEE Trans. Geosci. Remote Sens*., **37**(5), 2363-2373.

Li, J., Wang, S., Gunn, G., Joosse, P., and Russell, H. A., 2018: A model for downscaling SMOS soil moisture using Sentinel-1 SAR data. *Int. J. Appl. Earth. Obs. Geoinf*., **72**, 109-121.

Li, X., Cheng, G., Liu, S., Xiao, Q., Ma, M., Jin, R., Che, T., Liu, Q., Wang, W., Qi, Y, 2013: Heihe watershed allied telemetry experimental research (HIWATER): Scientific objectives and experimental design. *B. Am. Meteorol. Soc.*, 2013, **94**(8), 1145–1160, doi:10.1175/BAMS-D-12-00154.1.

Li, X., Liu, S., Xiao, Q., Ma, M., Jin, R., Che, T., ... & Wang, L, 2017: A multiscale dataset for understanding complex eco-hydrological processes in a heterogeneous oasis system. *Sci. Data.*, **4**, 170083, doi:10.1175/BAMS-D-12-00154.1.

Lievens, H., and Verhoest, N.E.C, 2011: On the retrieval of soil moisture in wheat fields from L-band SAR based on Water Cloud Modeling, the IEM, and effective roughness parameters. *IEEE Geosci. Remote S.*, 2011, **8**(4):740–744, doi:10.1109/LGRS.2011.2106109.

Lievens, H., R.H. Reichle, Q. Liu, G.J.M. De Lannoy, R.S. Dunbar, S.B. Kim, N.N. Das, M. Cosh, J.P. Walker, W. Wagner, 2017: Joint Sentinel-1 and SMAP data assimilation to improve soil moisture estimates. *Geophys. Res. Lett.*, **44**(12), 6145-6153, doi:10.1002/2017GL073904.

Liu, S., Xu, Z., Song, L., Zhao, Q., Ge, Y., Xu, T., Ma, Y., Zhu, Z., Jia, Z., Zhang, F, 2016: Upscaling evapotranspiration measurements from multi-site to the satellite pixel scale over heterogeneous land surfaces. *Agric. For. Meteorol.*, 2016, **230**, 97–113, doi:10.1016/j.agrformet.2016.04.008.

McDonald, K. C., Dobson, M. C., & Ulaby, F. T, 1990: Using MIMICS to model L-band multiangle and multitemporal backscatter from a walnut orchard. *IEEE T. Geosci. Remote.*, **28**(4), 477-491, doi:10.1109/TGRS.1990.572925.

Nearing, G. S., & H. V. Gupta, 2015: The quantity and quality of information in hydrologic models. *Water Resour. Res.*, **51**(1), 524-538, doi: 10.1002/2014WR015895.

Nearing, G. S., H. V. Gupta, W. T. Crow, and W. Gong, 2013: An approach to quantifying the efficiency of a Bayesian filter. *Water Resour. Res.*, **49**(4), 2164-2173, doi: 10.1002/wrcr.20177.

Oh, Y., Sarabandi K, and Ulaby F T, 2002: Semi-empirical model of the ensemble-averaged differential Mueller matrix for microwave backscattering from bare soil surfaces, *IEEE T. Geosci. Remote.*, 2002, **40**(6): 1348–1355, doi:10.1109/TGRS.2002.800232.

Oh, Y., Sarabandi, K., and Ulaby, F.T, 1992: An empirical model and an inversion technique for radar scattering from bare soil surface. *IEEE T. Geosci. Remote.*, 1992, **30**(2): 370–381, doi:10.1109/36.134086.

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. Sens. Environ.*, **134**, 234-248, doi:10.1016/j.rse.2013.02.027.

Patel, P., Srivastava, H. S., Panigrahy, S., & Parihar, J. S, 2006: Comparative evaluation of the sensitivity of multi−polarized multi−frequency SAR backscatter to plant density. *Int. J. Remote Sens.*, **27**(2), 293-305, doi:10.1080/01431160500214050.

Qiu, J., W. T. Crow, G. S. Nearing, X. Mo and S. Liu, 2014: The impact of vertical measurement depth on the informationcontent of soil moisture time series data. *Geophys. Res. Lett.*, **41**(14), 4997-5004, doi: 10.1002/2014GL06001.

Qiu, J., W.T. Crow, and G.S. Nearing, 2016: The impact of vertical measurement depth on the information content of soil moisture for latent heat flux estimation. *Journal of Hydrometeorology.*, **17** (9): 2419-2430, doi: 10.1002/2014GL060017.

Santi, E., Paloscia, S., Pettinato, S., Brocca, L., Ciabatta, L., and Entekhabi, D., 2018: On the synergy of SMAP, AMSR2 AND SENTINEL-1 for retrieving soil moisture. *Int. J. Appl. Earth. Obs. Geoinf.*, **65**, 114-123.

Seneviratne, S. I., T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky, and A. J. Teuling, 2010: Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Sci. Rev.*, **99**(3), 125-161, doi:10.1016/j.earscirev.2010.02.004.

Shi, J., Wang, J., Hsu, A.Y., O'Neill, P.E., and Engman, E.T, 1997: Estimating of bare surface soil moisture and surface roughness parameters using L-band SAR images data. *IEEE T. Geosci. Remote.*, 1997, 35: 1254–1266.

Ulaby, F. T., Sarabandi, K., Mcdonald, K. Y. L. E., Whitt, M., and Dobson, M. C, 1990: Michigan microwave canopy scattering model. *Int. J. Remote Sens.*, 1990, **11**(7): 1223–1253, doi:10.1080/01431169008955090.

Vermote, E., Justice, C., Claverie, M., and Franch, B., 2016: Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sens. Environ.*, **185**, 46-56.

Wu, T. D., & Chen, K. S, 2004: A reappraisal of the validity of the IEM model for backscattering from rough surfaces. *IEEE T. Geosci. Remote.*, **42**(4), 743-753, doi:10.1109/TGRS.2003.815405.

Xu, Z., Liu, S.,Li, X., Shi, S., Wang, J., Zhu, Z., Xu, T., Wang, W., Ma, M, 2013: Intercomparison of surface energy flux measurement systems used during the HIWATER−MUSOEXE. *J. Geophys. Res.*, 2013, **118**(23), 13–140, doi:10.1002/2013JD020260.

Zeng, J., Chen, K. S., Bi, H., Zhao, T., & Yang, X, 2017: A comprehensive analysis of rough soil surface scattering and emission predicted by AIEM with comparison to numerical simulations and experimental measurements. *IEEE T. Geosci. Remote.,* **55**(3), 1696-1708, doi:10.1109/TGRS.2016.2629759.

Zribi, M., Chahbi, A., Shabou, M., Lili-Chabaane, Z., Duchemin, B., Baghdadi, N., ... & Chehbouni, A, 2011: Soil surface moisture estimation over a semi-arid region using ENVISAT ASAR radar data for soil evaporation evaluation. *Hydrol. Earth Syst. Sc.*, **15**(1), 345-358, doi:10.5194/hess-15-345-2011.