ISPRS Journal of Photogrammetry and Remote Sensing 185 (2022) 111–128

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Identifying determinants of spatio-temporal disparities in soil moisture of the Northern Hemisphere using a geographically optimal zones-based heterogeneity model

Peng Luo [a], Yongze Song [b,*], Xin Huang [c,d], Hongliang Ma [c], Jin Liu [e], Yao Yao [f], Liqiu Meng [a]

[a] Chair of Cartography and Visual Analytics, Department of Aerospace and Geodesy, Technical University of Munich, Munich, Germany
[b] School of Design and the Built Environment, Curtin University, Perth, Australia
[c] State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China
[d] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China
[e] State Key Laboratory of Earth Surface Processes and Resource Ecology, Faculty of Geographical Science, Beijing Normal University, Beijing, China
[f] School of Geography and Information Engineering, China University of Geoscience, Wuhan, China

ARTICLE INFO

Keywords:
Spatial heterogeneity
Soil moisture
Spatial determinants
Geographically optimal zones-based heterogeneity model
Google Earth Engine (GEE)
Spatial statistics

ABSTRACT

Soil moisture is a fundamental ecological component for climate and hydrological studies. However, the distribution patterns of soil moisture are spatially heterogenous and influenced by multiple environmental factors. The knowledge is still limited in assessing the large-scale spatial heterogeneity of soil moisture in in situ data modelling, in situ network design, spatial down-scaling, and remote sensing-based soil moisture retrieval. Heterogeneity models are effective in characterizing spatial disparities, but they are not capable of examining the maximum regional disparities. To address this bottleneck, the authors of this study developed a geographically optimal zones-based heterogeneity (GOZH) model. By progressively optimizing geographical zones of soil moisture and quantifying the heterogeneity among zones, GOZH may help identify individual and interactive determinants of soil moisture across a large study area. It was applied to identify spatial determinants of in situ soil moisture data collected at 653 monitoring stations in the Northern Hemisphere in unfrozen and frozen seasons from April 2015 to December 2017, with only thawed data considered in both seasons. Correspondingly, a series of variables were derived from Google Earth Engine (GEE) remote sensing data. The results demonstrated the significant regional disparities of soil moisture, and the combinations of determinants are critically different among geographical zones and between unfrozen and frozen seasons. At a global scale, the combinations of determinants can explain about 48% of the spatial pattern of soil moisture. Spatial heterogeneity of soil moisture in frozen seasons is much more complex than that in unfrozen seasons regarding geographical zones and explanatory variables. The variability of soil moisture during unfrozen seasons can be more explainable than that during frozen seasons, which was a convincing evidence for previous studies that soil moisture predictions were mostly performed during unfrozen seasons. Primary variables that determine spatial patterns of soil moisture are changed from climate variables during the unfrozen season to geographical variables during the frozen season. Results show that GOZH model can effectively explore spatial determinants of soil moisture through avoiding the underestimation of individual variables, overestimation of multiple variables, and finely divide zones. The research findings from this study provide an in-depth understanding of the spatial heterogeneity of soil moisture and can be implemented in more effective in situ sampling network design, spatial down-scaling of soil moisture, and accurate inversion of surface parameters from the satellite data of soil moisture.

1. Introduction

Soil moisture is an essential component of an ecosystem (Green et al., 2019; Li et al., 2020), and plays a fundamental role in plant growth (Lei et al., 2018), food security (McColl et al., 2017), carbon and water cycles (Wang et al., 2017), soil productivity, and projection of the global climate change (Berg et al., 2017). Monitoring of soil moisture is required for agricultural production (Lei et al., 2018), drought

* Corresponding author.
E-mail address: yongze.song@curtin.edu.au (Y. Song).

https://doi.org/10.1016/j.isprsjprs.2022.01.009
Received 15 August 2021; Received in revised form 5 December 2021; Accepted 19 January 2022
Available online 2 February 2022

monitoring (Babaeian et al., 2018), on-farm moisture management, water resources management, hydrological simulation, and forecasting (Kumar et al., 2018). Nevertheless, the strong and complex spatial heterogeneity of soil moisture challenges large-scale and regional studies in hydrology and climate (Chaney et al., 2015; Molero et al., 2018). The spatial heterogeneity of soil moisture is closely associated with precipitation, atmospheric variability, evapotranspiration (ET), runoff etc (Quinn et al., 1995; Peters-Lidard et al., 1997; Famiglietti et al., 2008; Chaney et al., 2015). It also critically affects the in situ sampling network design (Chaney et al., 2015; Dari et al., 2019), spatial downscaling of soil moisture, hydrologic modelling, and agricultural management (Vereecken et al., 2014).

In-situ observations have been the fundamentally essential data for understanding soil moisture, climate variability, and the validation and refinements of remote sensing products (Albergel et al., 2009; Ma et al., 2019; Zappa et al., 2019; Dari et al., 2019). The in situ soil moisture observations have a number of advantages. First, the in situ data of soil moisture has high accuracy and continuous observations (Albergel et al., 2012). The in situ data are generally continuously observed by seconds at sampling sites and the accuracy can reach up to 0.05 m^3/m^3 (Albergel et al., 2010; Gruber et al., 2013). In addition, compared with remote sensing products with spatial resolutions between 250 m and 1 km, the in situ data are collected at precise locations of monitoring stations or GPS locations in the field (Wu and Liu, 2012). Finally, the regional, national, or global sampling networks assure the overall reliability and quality of in situ observations (Dorigo et al., 2011; Molero et al., 2018). The increasing number of open data libraries have brought about new opportunities for global data sharing and collaborative studies. For these reasons, the in situ observations provide the most essential data source for understanding soil moisture, climate variability, and the validation and refinement of remote sensing products.

The spatial variability of soil moisture is influenced by complex geographical and environmental factors, such as temperature, vegetation, topography, soil properties, depth to water table, freeze–thaw states, and scales (Brocca et al., 2010; Chaney et al., 2015; Ochsner et al., 2019; Li et al., 2021). These factors fall into four categories: (1) Climate conditions, for instance, precipitation may directly influence the water balance and the hydrological cycle (Kusangaya et al., 2016; Wang et al., 2018), and temperature may influence the energy balance principle and water circulation (Tao et al., 2021); (2) Geographical and terrain conditions, they may affect the storage and evaporation of soil moisture, and influence the direction and amount of water flow (Silva et al., 2014). The altitude may change the soil properties by influencing the environmental factor such as temperature (Niu et al., 2017); (3) Soil properties and freeze–thaw states, they are related to the forms and amount of water stored in soil. Fine-textured soils can store water more readily than coarse soils, resulting in high soil moisture. Soil texture can also affect the heat fluxes and thus soil moisture (Albergel et al., 2008; Shellito et al., 2018); (4) Surface coverage, for example, vegetation can influence the vertical drinage and ET flues, and closely associated with soil moisture (Baroni et al., 2013; Vereecken et al., 2014). In addition to individual variables, studies have demonstrated variables usually have interactive effects in affecting soil moisture patterns (Famiglietti and Wood, 1994; Western et al., 2004; Wilson et al., 2005; Konare et al., 2008; Chaney et al., 2015). Characteristics of spatial heterogeneity is also affected by the spatial scale (Han et al., 2018). The spatial variability of soil moisture is essentially different at the field scale (Nielsen et al., 1973; Bell et al., 1980; Vereecken et al., 2014), catchment scale (Western et al., 2004; Rosenbaum et al., 2012), regional scale (Romshoo, 2004; Zhao et al., 2013), and continental scale (Entin et al., 2000).

Given the considerable differences among regional environmental impacts on soil moisture, effective and reliable geographical zones are critically important for regional soil moisture inversions from remote sensing data, downscaling with the supports of local terrain and environmental variables (Zhuo et al., 2020), and network design (Vereecken et al., 2014). For example, the valid reference of regional difference of the soil moisture determinants is increasingly needed at a global scale when calibrating the ground roughness parameterization scheme with ground observation data (Verhoest et al., 2008). The limited knowledge in the regional disparity of soil moisture and its controls have been a challenge for the interpretability and transferability of the parameters. In addition, the geographical zones considering the spatial heterogeneity of soil moisture can support the network design (Zhuo et al., 2020). The network can capture spatial variability of soil moisture at the lowest possible cost by improving the representation of the soil moisture samples (Vereecken et al., 2014; Chaney et al., 2015).

A wide range of methods have been developed to understand the spatial heterogeneity of soil moisture. The commonly use methods can be classified into three categories, geostatistical analysis, wavelet analysis, and empirical orthogonal function (EOF) (Vereecken et al., 2014). Geostatistical models are effectively applied to identify static mapping patterns in soil properties (Ochsner et al., 2019). The spatial pattern of soil moisture at the field scale determined by multiple factors was observed through geostatistical analysis (Brocca et al., 2010). Wavelet analysis was originally used to analyze time series and has been applied to characterize the spatial variability of soil data patterns (Song et al., 2021; Vereecken et al., 2014). For example, the spatial pattern of soil moisture and temperature in the southern interior of British Columbia was characterized based on wavelet analysis (Redding, 2003). The difference of spatial scales in soil moisture variability was revealed using the wavelet analysis (Das and Mohanty, 2008). Empirical orthogonal function (EOF) methods were developed in terms of spatial modes and signal processing of soil moisture data (Wang et al., 2017). For instance, studies based on the EOF methods demonstrate that soil characteristics and topography were the two most critical factors to soil moisture (Perry and Niemann, 2007), and soil texture explained 61% of the variation in soil moisture (Jawson and Niemann, 2007).

Spatial stratified heterogeneity (SSH) models are effective approaches to investigate determinants of spatial variability of geographical variables (Wang et al., 2016). The basic assumption of SSH models is to compare the zonal spatial distribution patterns of dependent and independent variables. The zones are determined by categories of categorical variables or the spatial discretization of continuous explanatory variables (Song et al., 2020; Wang et al., 2010). As such, the spatial discretization is essential for identifying spatial determinants, and the process of spatial discretization is presented in the next paragraph. The power of determinants (PD) is calculated as a ratio between the sum of zonal variance and the variance of data across the whole space. This means that a higher PD value is associated with higher zonal variance. The commonly used SSH models include Geodetector (Song et al., 2018; Wang et al., 2010, 2016), optimal parameter-based geographical detector (OPGD) (Song et al., 2020; Luo et al., 2021), interactive detector of spatial associations (Song and Wu, 2021), etc. The SSH models have been increasingly implemented to characterize the spatial variability of soil properties. For example, the spatial difference of tillage factors of the China soil loss equation was characterized using the SSH model (Chen, 2021). The driving forces of soil erosion were explored using the GD model (Liang and Fang, 2021). The spatiaotemporal variability of soil organic matter was also revealed based on the heterogeneity using the GD model (Hu et al., 2021). In addition, some soil properties, such as soil organic carbon, were mapped using the GD–based kriging model (Liu et al., 2021). Overall, existing research demonstrated the effectiveness and viability of using the spatial stratified heterogeneity model to reveal the variability of soil variables.

However, regarding the complex spatial heterogeneity of in situ soil moisture in large regions, there are still difficulties in addressing following issues using current SSH models. First, spatial discretization is an essential step to identify geographical zones based on spatial patterns of explanatory variables (Song and Wu, 2021). In current studies, the general procedure of spatial discretization is performed using a two-step approach. The individual geographical variables are first discretized using supervised or unsupervised approaches, such as equal, quantile, standard deviation, and geometric breaks, to determine spatial zones based on an individual variable, and then combine the spatial zones through a spatial overlay (Cang and Luo, 2018; Song et al., 2020). In this method, distribution characteristics of the response variables are not fully explained in the discretization process, leading to the incomplete exploration of the influence of explanatory variables on the response variable. Thus, it is necessary to identify the geographically optimal zones which can maximize the difference among zones and minimize the similarities within zones. In addition, the reliability of estimating the power of interactive determinants needs to be improved due to the massive finely divided zones from the spatial overlay of zone layers of multiple explanatory variables (Song and Wu, 2021). In most of the previous studies, the power of interactive determinants is only estimated for the interaction of only two or three explanatory variables as for the problem of massive finely divided zones. Therefore, it is essential to develop reliable models to identify geographical optimal zones and more accurately estimate the power of interactive determinants of spatial heterogeneity of the soil moisture in large regions.

In this study, we developed a geographically optimal zones-based heterogeneity (GOZH) model to characterize the spatial heterogeneity and examine determinants of large-scale soil moisture. In the GOZH model, an optimal power of determinant (OPD) indicator was developed to reveal the contributions of variables on spatial patterns of soil moisture, where the spatial discretization was conducted heuristically in a step-wise process. The GOZH model was used to identify the geographically optimal zones during the unfrozen and frozen season and estimate the determinants of spatio-temporal disparities in soil moisture of the Northern Hemisphere. Soil moisture in situ data were collected at 653 monitoring stations in the Northern Hemisphere from April 2015 to December 2017 to present the soil moisture with high accuracy and in precise locations. Only soil data at thawed status were included to ensure the modelling reliability. Correspondingly, remote sensing-based explanatory variables were derived from Google Earth Engine (GEE), and classified into four categories, geography, climate, soil, environment ecology. First, impacts of individual variables on soil moisture and temporal variations during 33 months were characterised. Second, the geographically optimal zones of seasonal soil moisture were identified using the GOZH model. Third, the determinants of spatial patterns were demonstrated during two seasons according to the geographically optimal zones derived in the last step. Finally, the performance of GOZH model was evaluated and compared with the OPGD model.

The remainder of this paper is structured as follows: Section 2 introduces the in situ soil moisture data and explanatory variables used in this study. Section 3 describes the objective, definition, and derivation of the GOZH model. Section 4 covers the methodologies to explore the soil moisture variability in the Northern Hemisphere using the GOZH model. Section 5 presents results of this study, including impacts of individual variables and temporal variations, geographically optimal zones, and determinants of spatial disparities and seasonal effects. Findings and research contributions are discussed in Section 6, and the study is concluded in Section 7.

2. Data

2.1. In-situ soil moisture data

In this study, monthly in situ soil moisture data in 762 observation locations from 653 monitoring stations across the Northern Hemisphere were selected to reveal the heterogeneity and determinants of soil moisture (Table 1). All stations belong to 12 networks in the International Soil Moisture Network (ISMN) (Dorigo et al., 2011; Dorigo et al., 2021). ISMN is a widely used soil moisture network that collects soil moisture and soil temperature data sets from global networks, including 1,400 stations from 40 global networks of soil monitoring (Dorigo et al., 2015; Ma et al., 2019; Ma et al., 2021). As a data hosting facility of soil moisture data, ISMN has been widely used for validations of satellite-derived soil moisture products (Ma et al., 2019; Dorigo et al., 2021; Ma et al., 2021). In each station, soil moisture data ranging from April 2015 to December 2017 were collected and analyzed to characterize the spatial and temporal patterns of soil moisture in the Northern Hemisphere.

Table 1
A brief description and sources of soil moisture in situ data used in this study.

Network name	Country	Number of stations	Depth (cm)	Reference
USCRN	America	97	5	(Bell et al., 2013)
SNOTEL	America	208	5	https://www.wcc.nrcs.usda.gov/
SCAN	America	157	5	(Schaefer et al., 2007)
CTP_SMTMN	China	53	0–5	(Yang et al., 2013)
RISMA	Canada	14	0–5	(McNairn et al., 2014)
HOBE	Denmark	28	0–5	(Jensen and Illangasekare, 2011)
FMI	Finland	19	5	(Zeng et al., 2016)
SMOSMANIA	France	15	5	(Albergel et al., 2008)
TERENO	Germany	4	5	(Zacharias et al., 2011)
BIEBRZA-S-1	Poland	18	5	http://www.igik.edu.pl/en
REMEDHUS	Spain	20	0–5	(Martínez-Fernández and Ceballos, 2005)
RSMN	Romania	20	0–5	(Ma et al., 2019)

2.2. Explanatory variables

Four categories of explanatory variables have been collected to explain the spatial disparities of soil moisture. They include geographical, climate, soil, and environmental variables derived from remote sensing data (Table 2). All the remote sensing data were derived and processed using the Google Earth Engine (GEE). Climate and environmental variables, with the temporal resolution from 8 days to one month, were collected from April 2015 to December 2017, consistent with the in situ soil moisture data.

(A) Geographical variables

Four terrain explanatory variables, including elevation, slope, aspect, and hill shade, were included in this study to demonstrate the local geographical conditions. The terrain variables were derived from the Shuttle Radar Topography Mission (SRTM) data. The SRTM provides the digital elevation model (DEM) data with the resolution of about 30 m (Elkhrachy, 2018). Slope, aspect, and hill shade variables were

Table 2
Explanatory variables of the spatial disparities of soil moisture.

Category	Variable	Product	Temporal resolution
Geography	Elevation	SRTM DEM	-
	Slope	SRTM DEM	-
	Aspect	SRTM DEM	-
	Hill shade	SRTM DEM	-
Climate	Precipitation	GPM	Monthly
	Temperature	MOD11	8 days
Soil	Soil texture	OpenLandMap	-
	Soil pH	OpenLandMap	-
	Soil bunk density	OpenLandMap	-
Environment	Normalized difference vegetation index (NDVI)	MOD13Q1	16 days
	Enhanced Vegetation Index (EVI)	MOD13Q1	16 days
	Leaf Area Index (LAI)	MOD15A2H	8 days
	Evapotranspiration	MOD16A2	8 days

calculated based on the DEM data using GEE spatial analysis, where local gradients were computed using the 4-connected neighbors of each pixel for the calculation of slope and aspect.

(B) Climate variables

Precipitation and temperature are two essential climate controls on soil moisture at a large spatial scale. In this study, monthly precipitation data was derived from the Global Precipitation Measurement (GPM) IMERG Final Precipitation L3 1 month 0.1 degree x 0.1 degree V06 (Joyce and Xie, 2011; Hou et al., 2014). The GPM is an international satellite mission to provide precipitation data at a 0.1-degree resolution. The temperature data were taken from the land surface temperature (LST) product of the 8-days MOD11 composition with a spatial resolution of 1.2 km (Hashimoto et al., 2008). LST has been used as an effective data source for assessing soil conditions and forecasting the soil moisture (Holzman et al., 2014; Jiang and Weng, 2017).

(C) Soil properties

The soil properties used in this study include soil texture, soil pH, and soil bulk density extracted from the Soil Moisture Active Passive (SMAP) product (Entekhabi et al., 2010). The SMAP provides a series of soil properties data at different depths between 10 cm and 200 cm, and a resolution of 250 m. Corresponding to the soil moisture data that were collected at depths of 0 to 5 cm from soil monitoring networks (Table 1), soil texture, pH and bulk density data were collected at a 10 cm depth to represent soil properties related controls of soil moisture.

(D) Environmental variables

Local environmental and ecological conditions surrounding soil moisture monitoring stations were characterized using Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Evapotranspiration (ET), and Leaf Area Index (LAI) data driven from MODIS products using GEE. The NDVI and EVI data were extracted from the 16-days Terra MODIS products (MOD13Q1) with a spatial resolution of 250 m (Didan et al., 2015). The ET data were collected from the global 8-day MOD16A2 product with a 1 km spatial resolution (Mu et al., 2013). The ET variable is used to present the water cycle of the Earth's climate system, especially the evaporation and transpiration processes that are critically associated with soil moisture (Purdy et al., 2018). LAI data, used as an ecology indicator to explore the soil moisture variability, was extracted from the MOD15A2H product, which was a 8-days composite data with a 500 m resolution. LAI is defined as the one-sided green leaf area per unit ground area in broadleaf canopies and as one-half the total needle surface area per unit ground area in coniferous canopies. LAI is an essential indicator of vegetation structure for revealing the interaction between soil and vegetation (Fang et al., 2019). The interaction of LAI and soil moisture has a significant impact on drought, vegetation growth, vegetation senescence, and drought forest (Sawada, 2018; Liu et al., 2017).

The explanatory variables are derived from the pixels at the location of the soil moisture monitoring stations. In this study, spatial heterogeneity of soil moisture at stations in the Northern Hemisphere is much higher than that of data within grids of explanatory variables, e.g. 90 m or 250 m. Therefore, spatial analysis in the study will not be affected by the scale effects of explanatory variables derived from remote sensing or grid data. Similar processing of deriving explanatory variables of soil moisture from grid data can be found in (Peng et al., 2015; Qu et al., 2021).

2.3. SMAP freeze/thaw product

To ensure the reliability of soil moisture analysis, only the in situ data of soil moisture at thaw-status landscape were used in the study. To select the monthly in situ data at the unfrozen situation, the SMAP L3 Freeze/Thaw product (SPL3FTP) with a 36 km resolution in the Northern Hemisphere were collected (Xu et al., 2018). The SMAP is a NASA satellite mission launched in 2015 to monitor the surface (about 5 cm) global soil moisture and landscape freeze/thaw status (Entekhabi et al., 2010; Al-Yaari et al., 2019). Missing data of the SMAP product was filled through the comparison of data at neighbouring locations and periods. The thaw or freeze status of the landscape at the soil moisture stations were derived through spatial overlay.

3. Geographically optimal zones-based heterogeneity (GOZH) model

3.1. Power of determinants (PD) of spatial stratified heterogeneity (SSH) model

As introduced above, in SSH models, a higher PD value of an explanatory variable indicates that the spatial distribution pattern of this variable tends to be more similar to the spatial pattern of response variable, i.e., soil moisture in this study. The process of estimating PD values of explanatory variables generally includes three steps. First, continuous explanatory variables should be converted to stratified variables using spatial discretization methods. The stratified variables can determine a series of geographical zones of soil moisture. Second, if multiple variables are used to identify the interactive impacts on soil moisture, geographical zones determined by explanatory variables need to be overlapped to generate a new layer of geographical zones, which contain geoinformation of all the variables. Finally, the PD value for the comparison of spatial patterns between response variable and explanatory variables are calculated as a ratio of the variance of soil moisture within geographical zones determined by one or multiple explanatory variables and the variance across the whole study area. The PD is computed as:

$$\mathrm{PD} = 1 - \frac{SSW}{SST} = 1 - \frac{\sum_{z=1}^{h} N_z \sigma_z^2}{N\sigma^2} \tag{1}$$

where SSW is the Sum of Squares Within geographical zones determined by explanatory variables, SST is the Sum of Squares Total of soil moisture in the whole study area, N_z and σ_z are the number and standard deviation of soil moisture within geographical zone $z(z = 1,\dots,h)$, and N and σ are the number and standard deviation of soil moisture across the study area. PD value ranges from 0 to 1, where a high PD value indicates a high spatial association between response variable and the explanatory variable.

From this equation and recent studies, we can find that the PD value is sensitive to the geographical zones determined by the spatial discretization of explanatory variables. As such, a more effective and reliable spatial discretization approach is required to maximize the variance values among zones and minimize variance within zones. In addition, as explained in the introduction section, reliable geographical zones are also essential for regional soil moisture inversions from remote sensing data and downscaling with the supports of local terrain and environmental variables.

3.2. PD of GOZH model

In this study, we define the PD as a function of explanatory variables and geographical zones, which are determined by stratified variables from certain spatial discretization processes:

$$\gamma\left(X, D\right) = 1 - \frac{SSW_{X,D}}{SST} \tag{2}$$

where X is one or multiple explanatory variables, D is the stratified variable for describing geographical zones, and $SSW_{X,D}$ is the sum of squares within geographical zones that are recorded as D and determined by explanatory variable X.

In GOZH model, the optimal PD (OPD) value can demonstrate the maximum explanatory power of variables in terms of geographically optimal zones. As such, the OPD of explanatory variables, expressed as Ω value, is the maximum value of the PD function γ:

$$\Omega = \max\left(\gamma\right) = 1 - \frac{\min\left(SSW_{X,D}\right)}{SST} \tag{3}$$

The geographically optimal zones have the minimum intra-area variance and the maximum inter-area variance. To calculate the Ω value, an optimization process is performed as:

$$min\left(SSW_{X,D}\right) = min\left\{\sum_{z=1}^{h}\sum_{j=1}^{N_z}\left(y_{z,j} - \bar{c}_z\right)^2\right\} \tag{4}$$

where $y_{z,j}$ and $\bar{c}_z$ are the jth observation and mean values of soil moisture in zone z, respectively.

This equation is a nondeterministic polynomial-time complete (NP-complete) problem, which is difficult to derive a global optimum. To solve this equation, a step-wise spatial discretization of soil moisture is performed using a heuristic method with spatial explanatory variables. First, all possible two-zone solutions of spatial discretization are derived for explanatory variables, and the optimal one is selected as the cutoff point according to the squared error minimization criterion. An iteration process is performed for each variable X_k to identify the optimal cutoff point s, and the parameters in the iteration can be presented as (k, s). Accordingly, the input space is divided into two regions. Second, the iteration process is performed for multiple variables. The kth explanatory variable X_k and its fetching value s_k are used as cut-off variables and cut-off points, respectively. Two regions in each iteration are defined as $R_1(k,s) = \{x|x^{(k)} \leqslant s\}$ and $R_2(k,s) = \{x|x^{(k)} > s\}$. In each split, the variable that allows the maximum explanation of the variance of the dependent variable is selected. Thus, the optimization process is converted to a process to identify the optimal variable X_k and the cuttoff point s of variable X_k, which can be expressed as:

$$min_{k,s}\left\{\sum_{x_i \in R_1(k,s)}\left(y_i - \bar{d}_1\right)^2 + \sum_{x_i \in R_2(k,s)}\left(y_i - \bar{d}_2\right)^2\right\} \tag{5}$$

where $\bar{d}_1$ and $\bar{d}_1$ are the average values of soil moisture in group R_1 and R_2, respectively. Thus, the above discretization process is repeated within each group until the data volume of the group less than a certain number, which is called minsplit. During the step-wise spatial discretization, when the data volume in one group is less than the minsplit, this group would not be subdivided further and automatically becomes a final spatial zone. This process is similar to the classification and regression tree (CART) algorithm (Breiman et al., 2017). The whole spatial discretization can be visualized as the binary tree structure.

Fig. 1 shows an example of the spatial discretization process of the GOZH model. In this example, the response variable is D and explanatory variables include A, B, and C. To conduct the step-wise spatial discretization, explanatory variables are processed one by one. For each variable, a series of cutoff points are selected to split the study area into

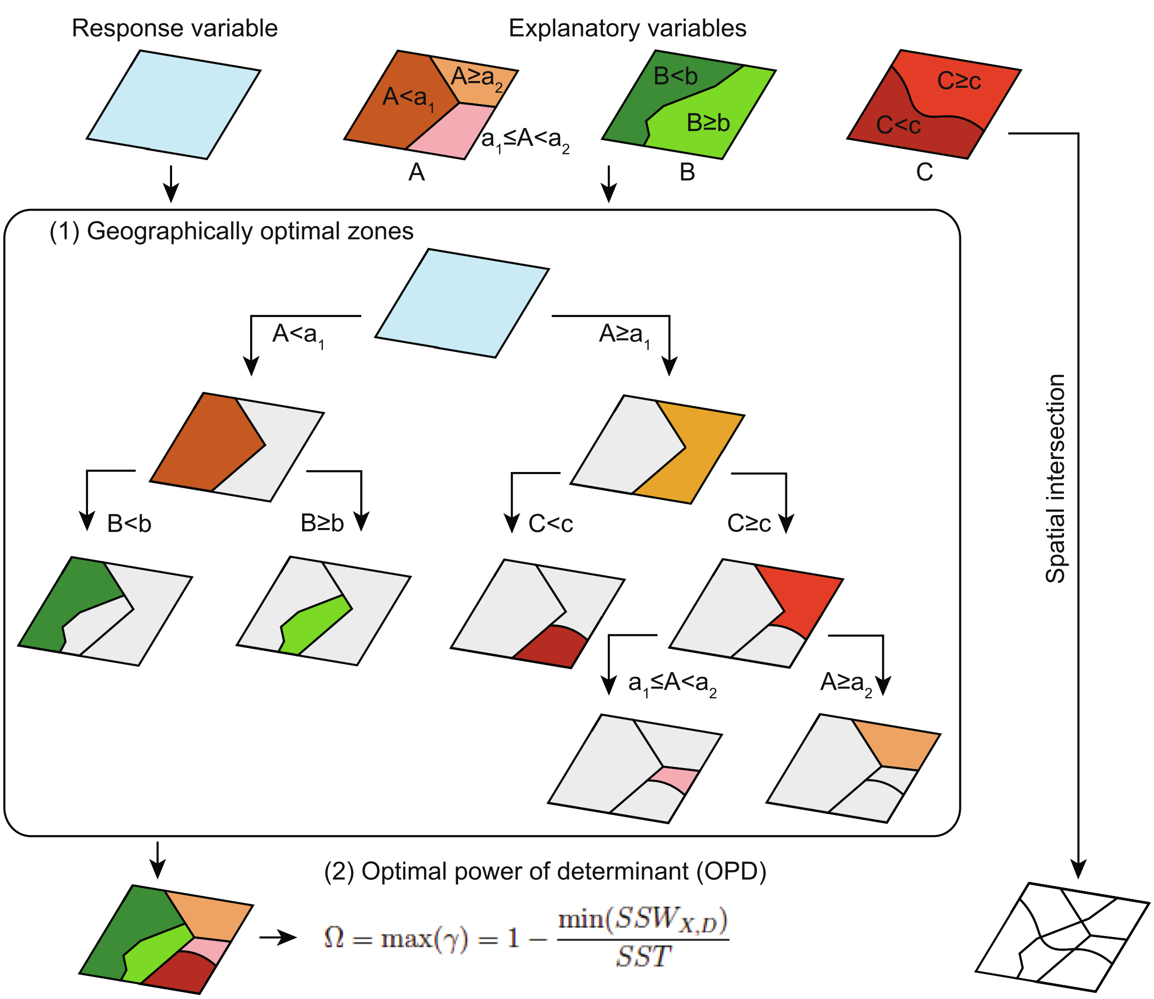

Fig. 1. Process and principle of the geographically optimal zones-based heterogeneity (GOZH) model.

two zones, and the *SSW* of soil moisture is calculated. Among all possible two-zone solutions, the one with the highest *SSW* is regarded as the optimal spatial discretization result, i.e., the optimal geographical zones, in this step. Repeat this process for each variable and data within zones, and finally, the optimal geographical zones with the highest overall *SSW* are regarded as the optimal discretization variable.

Compared with PD values in SSH models, the Ω values of GOZH model can identify the optimal geographical zones of data and demonstrate the maximum PD of explanatory variables. The GOZH model also can more effectively reveal the interaction effects of variables compared with SSH models.

4. GOZH-based spatiotemporal determinants and heterogeneity of soil moisture

Fig. 2 shows a flowchart of the GOZH-based spatio-temporal determinants and heterogeneity analysis of soil moisture in the Northern Hemisphere. The methods include five steps. The first step was the data pre-processing of in situ soil moisture data and explanatory variables data. Second, the monthly Ω values of individual explanatory variables were calculated to assess contributions of variables to spatial patterns of soil moisture. Third, the geographically optimal zones of soil moisture in unfrozen and frozen seasons were identified respectively using the GOZH model. The fourth step was to calculate determinants of spatial heterogeneity in soil moisture in unfrozen and frozen seasons with the support of the geographically optimal zones identified in the previous step. Finally, model validation was performed to assess the reliability and effectiveness of the model.

4.1. Data pre-processing

The in situ soil moisture data and explanatory variables data were first processed before modelling. The data pre-processing consists of following three parts. First, monthly in situ soil moisture at thaw landscape were selected using the corresponding SMAP Freeze/Thaw data. Second, the temporal periods soil moisture data were divided into unfrozen and frozen seasons to characterize respective spatial and temporal variation patterns of soil moisture. In this study, since most of the stations are located in the mid-latitudes of the Northern Hemisphere, April to September was regarded as unfrozen seasons and remaining months were frozen seasons. Finally, explanatory variables were processed to corresponding locations and time periods of soil moisture monitoring stations. For instance, the 8-day composite products LST, LAI, and Evapotranspiration, and the 16-day composite data product NDVI and EVI were processed to monthly data using GEE. A small amount of missing data in a few variables wes interpolated using an inverse distance weighting (IDW) spatial interpolation approach.

4.2. Impacts of individual variables and their temporal variations

In the study, the GOZH model is first implemented in investigating impacts of individual explanatory variables on spatial patterns of soil moisture in each month from April 2015 to December 2017. In the monthly GOZH model, the geographical optimal zones determined by an individual variable were identified through an iteration process to maximize the variance among zones and minimize variance within zones. The minsplit was selected as 30 according to the data volume (i.e.

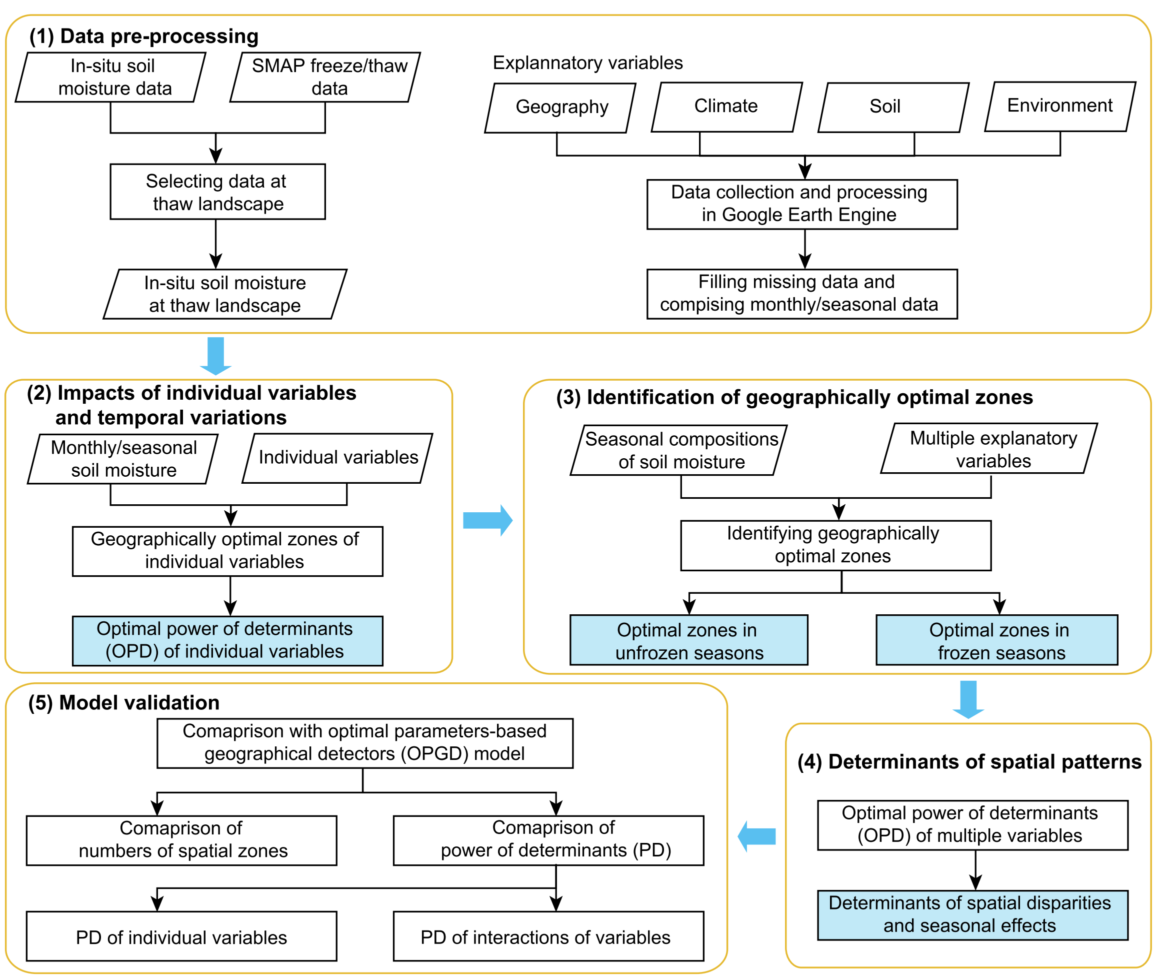

Fig. 2. Schematic overview of GOZH-based identification of determinants for spatiotemporal heterogeneity of soil moisture in the Northern Hemisphere.

653 stations) to avoid finely divided zones. During the spatial discretization process, the group with less than 30 stations would not be further subdivided and regarded as final spatial zones. Then, an Ω value is calculated to present the PD of this variable on soil moisture. Since explanatory variables were classified into four aforementioned categories, Ω value were computed for each variable in the four categories and in each month to indicate the temporal variations of impacts of individual variables on spatial patterns of soil moisture.

4.3. Identification of geographically optimal zones

Reliable and effective geographical zones of soil moisture are essential for parameter inversions from remote sensing data and an accurate downscaling. In the study, geographically optimal zones of soil moisture based on multiple variables were identified using the GOZH model, indicating the highest homogeneity within zones and the highest heterogeneity between zones. The optimal zones were then applied to assess the overall impacts of multiple variables on spatial patterns of soil moisture in the next step.

The identification of geographically optimal zones for unfrozen and frozen seasons takes place in three stages. First, values of monthly soil moisture and explanatory variables were merged for each season. Second, the optimal interaction was explored for each season, and the corresponding stratified variable from the spatial discretization was used to identify the geographically optimal zones. According to the stratified variable, soil monitoring stations were grouped into zones from the optimal interaction. Finally, geographical, climate, soil, and environmental characteristics at the locations of soil monitoring stations were summarized and analyzed according to the geographically optimal zones to reveal the regional spatial variability of soil moisture at a global scale.

4.4. Determinants of spatial disparities and seasonal effects

This step aims at quantifying the overall Ω value of the spatial interaction of explanatory variables on spatial patterns of soil moisture in unfrozen and frozen seasons. To assess the Ω value, an optimal interaction variable was created using the geographically optimal zones identified in the previous step. The optimal interaction variable was a categorical variable that involved the inter-dependencies of different explanatory variables and could control the spatial variability of soil moisture. Assuming the total of number of variables was n, the total number of possible spatial interactions (i.e., combinations) of variables was M ($M = 2^n - n - 1$). The optimal interaction variable demonstrated the highest Ω value among all potential spatial interactions of variables.

In addition to the overall Ω value of the spatial interaction of multiple explanatory variables, contributions of each variable within the overall Ω value was calculated using a variable removal method. The reduction of the Ω value due to the removal of this variable was calculated by removing each explanatory variable one by one in the optimal combination. The percentage of the Ω reduction of a given variable among the sum of the Ω reduction of all variables indicated the relative importance of this variable. Finally, the contribution of a given variable to spatial patterns of soil moisture was defined as the overall Ω value multiplied by its relative importance. This variable removal method has been widely applied in identifying contributions of variables within a total contribution in nonlinear models, such as generalized additive models (Song et al., 2015).

4.5. Model evaluation

To evaluate the effectiveness and reliability of the proposed GOZH model, a set of indicators were developed for comparing model performance between GOZH and the commonly used OPGD model. The indicators include PD values of individual variables, PD values of interactions of variables, and the number of geographical zones for examining interactive effects of variables. The OPGD is an improved geographical detector model, which can be used to estimate PD values of both individual variables and interactions of variables by optimizing the spatial discretization process using unsupervised or supervised approaches (Song et al., 2020). In the OPGD model in this study, the discretization method is quantile breaks and the optional numbers of discretization are consecutive integers from 3 to 22. For each optional number of discretization, PD values were computed for all explanatory variables. Then, a local estimated scatter plot smoothing (LOESS) function was applied to model the trend of the 75% quantile values of PD values and calculate the change rates of the trend, where the span for fitting the LOESS function was 0.75 (Luo et al., 2021; Song and Wu, 2021). Finally, the break number enabled the change rate lower than 5% is selected as the optimal break number. All these parameters are selected based on the parameter selection approaches in previous studies (Song and Wu, 2021). The OPGD model was performed using the "GD" package in R (Song et al., 2020).

5. Results

5.1. Spatial and temporal patterns of soil moisture

Fig. 3 shows spatial distributions of monthly mean in situ soil moisture in the Northern Hemisphere in unfrozen (April-September) and frozen (October-March) seasons from 2015 to 2017. In general, soil moisture monitoring stations used in the study are densely distributed in North America (635 observations), and other stations are distributed in Europe (181 observations) and in China (53 observations). Along the longitude, locations of soil moisture monitoring stations can be classified into four areas as illustrated in Figs. 3 b and d. The spatial disparities of in situ soil moisture in Europe tend to be higher than those in other regions. In addition, the small figure on the right side of the Fig. 3 b shows the seasonal effects of monthly soil moisture at both all monitoring stations and stations at the thawed landscape. The seasonal effects show that the monthly mean soil moisture generally peaks in March and has the lowest values in July. The soil moisture in thawed locations tends to be higher than that in frozen locations. For instance, in March 2017, the mean soil moisture at all stations was 0.27, but at thawed locations was 0.29.

5.2. Impacts of individual variables and their temporal variations

The GOZH model first identified the primary variables of soil moisture. Fig. 4 shows the Ω values of different categories of explanatory variables on spatial patterns of soil moisture and their temporal variations in the study period. The monthly variations of Ω values indicate that spatial associations between patterns of soil moisture and explanatory variables have similar temporal trends to the spatial variability of soil moisture, which is marked with black lines in Fig. 4 a - d. This consistent trends demonstrate the effectiveness of the Ω values in examining spatial disparities of soil moisture. Fig. 4 e shows monthly average Ω values from April 2015 to December 2017. Among the four categories of variables, climate variables have the highest spatial associations with soil moisture, followed by geographical and environmental variables. For instance, from the perspective of individual variables, precipitation, elevation, and temperature have the highest Ω values among 13 variables and across the 33 months. The maximum Ω value is the impact of precipitation (58%) in November 2016. In this month, elevation and temperature can explain 55% and 46% of the spatial variability of soil moisture, respectively. Compared with climate, geographical, and environmental variables, soil property variables tend to have lower spatial associations with soil moisture, where soil texture has the lowest Ω values, ranging from 0% to 12%.

From the perspective of monthly variations, in the transitional months from frozen to unfrozen seasons, i.e., March and April, spatial associations between patterns of soil moisture and explanatory variables

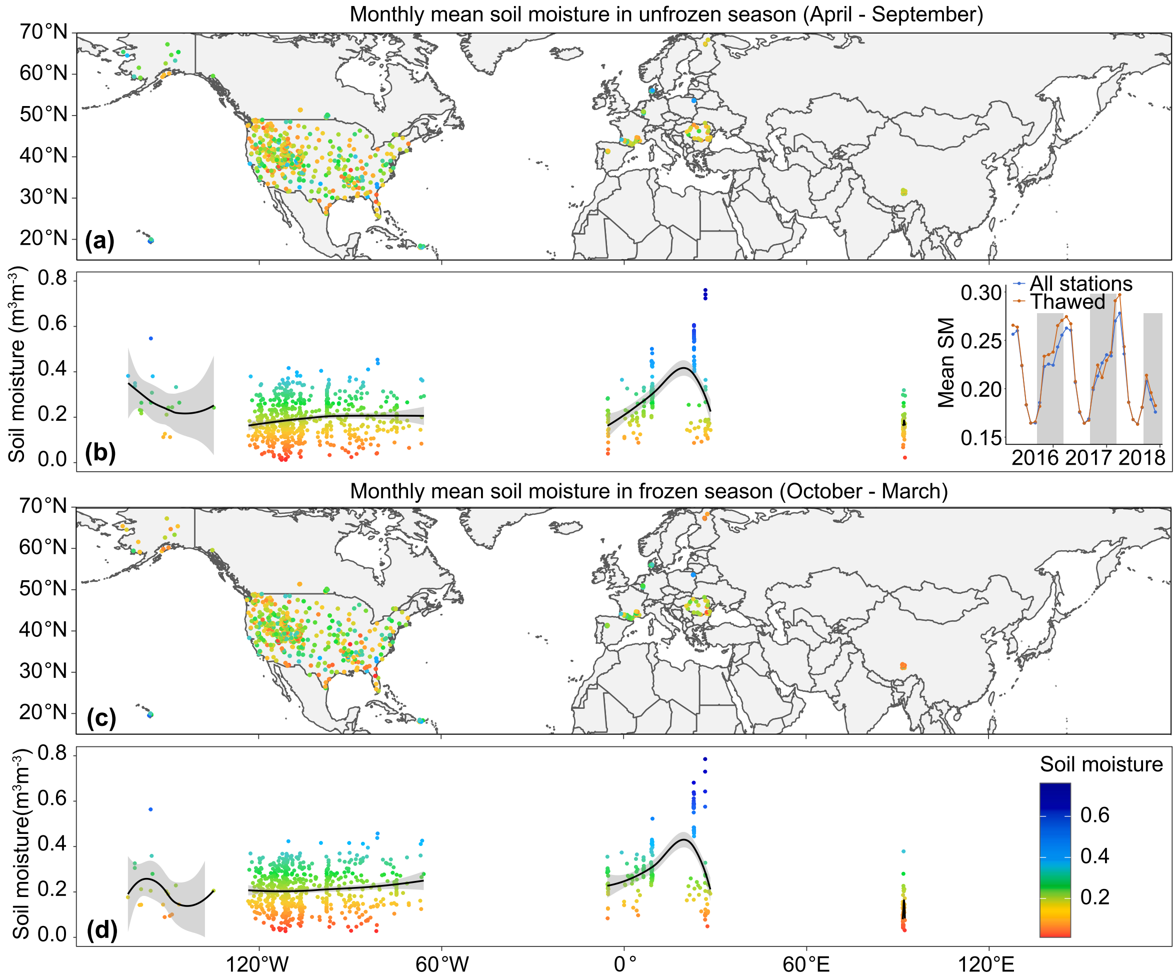

Fig. 3. Spatial distributions of monthly mean in situ soil moisture in the Northern Hemisphere in unfrozen (a and b) and frozen (c and d) seasons. In the small figure inside B, white background shows unfrozen seasons, and gray background shows frozen season.

generally have the highest Ω values. This means that spatial patterns of soil moisture in frozen-unfrozen transitional months can be more explained by geographical, climate, soil, and environmental variables. For instance, impacts of climate variables on patterns of soil moisture have been high in March ranging from 42% to 48% and with an average Ω value 47%. In March, the average Ω values of geographical, environmental, and soil variables are 24%, 22%, and 9%, respectively, which all are the highest monthly average Ω values in each category. We also can find that the least interpretable period of spatial patterns of soil moisture is the middle frozen seasons, i.e. December and January.

From the perspective of seasonal variations, results in Fig. 5 also reveal that the spatial pattern of soil moisture is more interpretable during the unfrozen season compared with that during the frozen season. The spatial associations in unfrozen and frozen seasons can be explained in a number of aspects. First, precipitation, elevation, and temperature have been the variables with the highest spatial associations with soil moisture. Ω values of precipitation, elevation, and temperature are 37.1%, 35.9%, and 32.1% in the unfrozen season, respectively, and 31.3%, 37.3%, and 34.5% in the frozen season, respectively. Ω values of environmental variables, including NDVI, LAI, EI and EVI, are lower than those of climate variables and elevation. Their contributions to spatial patterns of soil moisture are 15.8%-28.4% during the unfrozen season, and 9.5%-23.9% during the frozen season. This means that environmental variables also make important contributions to spatial variability of soil moisture. Soil property variables have the lowest Ω values in both unfrozen and frozen seasons. In addition, Ω values of most variables in the four categories have been reduced from unfrozen to frozen seasons. The average Ω value of individual variables during the unfrozen season (20.0%) is 12.4% higher than that during the frozen season (17.8%). Ω values of precipitation, environmental variables, hill shade, and soil property variables during the unfrozen season are 5.8%, 3.4%-6.4%, 1.3%, and 0.4%-2.5% lower than that during the frozen season. Third, different from most variables, Ω values of temperature and geographical variables, including elevation and aspect, are increased from unfrozen to frozen seasons. Compared with unfrozen season, the Ω value of temperature, elevation, and aspect during the unfrozen season are 2.4%, 1.5%, and 1.3% lower than that during the frozen season. Finally, the above results also indicate that the spatial variability of soil moisture is complex and it is difficult to be explained by individual variables. The maximum interpretability of spatial patterns of soil moisture is only around 37% using individual variables.

5.3. Geographically optimal zones

5.3.1. Unfrozen seasons

Fig. 6 shows the geographically optimal zones of soil moisture during the unfrozen season identified using the GOZH model. The geographically optimal zones during the unfrozen season were identified using four explanatory variables, including precipitation, NDVI, temperature, and soil pH, and included nine zones. Fig. 6 b shows that precipitation was the primary variable that controlled spatial patterns of soil moisture during the unfrozen season. According to parameters of precipitation in the top two layers, the nine zones can be classified into three groups: the first group (precipitation < 0.082 mm/hr) contained zone A, the second

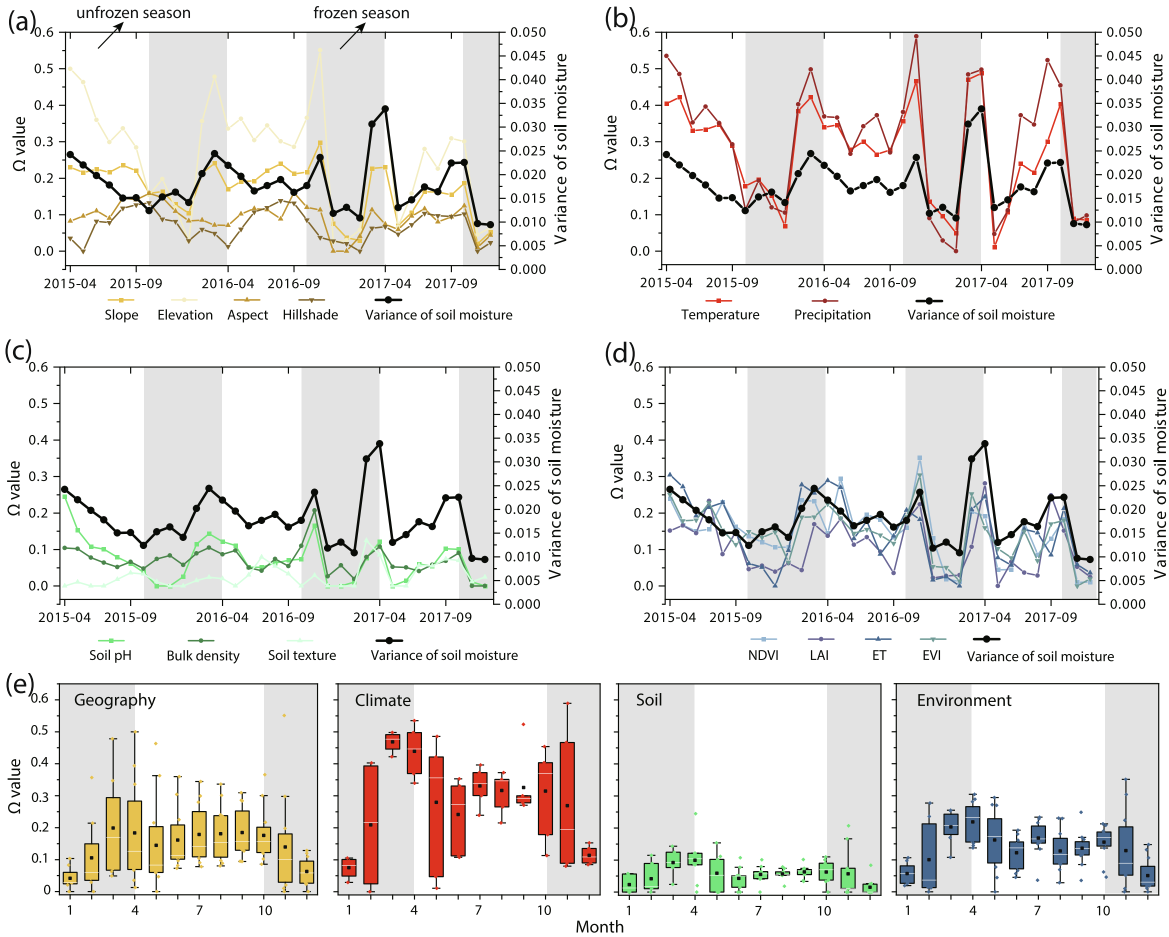

Fig. 4. Optimal power of determinants (OPD or Ω value) of explanatory variables on spatial patterns of monthly soil moisture. a: Ω values of geographical variables; b: Ω values of climate variables; c: Ω values of soil property variables; d: Ω values of environmental variables; and e: monthly summaries of Ω values from 2015 to 2017. White background shows unfrozen seasons (April-September), and gray background shows frozen seasons (October-March).

group (precipitation >= 0.094) contained zones B, C, and D, and the third group (0.082 <= precipitation < 0.094) contained zones E, F, G, H, and I. The characteristics of soil moisture and explanatory variables in the three groups of zones are explained as follows.

The first group, including zone A, was primarily distributed in the western contiguous United States, Alaska, western Spain, southern France, and eastern Romania. The average precipitation in this group is only 0.047 mm/hr, which was much lower than the average precipitation in other groups, such as zone B (0.155 mm/hr) and zone F (0.087 mm/hr). The contiguous United States, southern France, and Romania were typical regions that zones were primarily divided by precipitation. In the contiguous United States, the western part was drought or desert areas and the precipitation was low, and the precipitation was gradually increased from the middle to eastern areas. In the southern France, the precipitation was low in the Mediterranean coast areas, but it was relatively high in the Massif Central areas (Planchon, 2000). The east of Romania was drought and most stations were distributed in zone A, but the western Romania was more humid than other areas and most of the stations were located in zone B.

The second group, including zones B, C, and D, was generally distributed in the eastern contiguous United States, Alaska, southern France, Denmark, western Germany, western Romania, northern Finland, and eastern Tibetan Plateau, China. Zones B, C, and D were divided by temperature, where the temperature in zone B was high, in zone C was low, and in zone D was moderate. For instance, soil moisture monitoring stations in western Germany were located in zones B and D. A typical characteristic of zone B was the high precipitation (> 0.094 mm/hr) and high temperature (> 20 °C), but the average temperature in zone D is 19.1 °C, which was 30% lower than zone B (27.3 °C).

The third group, including zones E, F, G, H, and I, was mainly located at Alaska, southwestern France, eastern Poland, and northeastern Finland. A few monitoring stations in this group were sparsely located at the contiguous United States, Hawaii, and southern Romania. We can find that zones E and F are usually distributed in neighbouring locations, such as central United States, southwestern France and southern Romania. The variable for differentiating zones E and F was NDVI, where the average NDVI in zone E and zone F was 0.49 and 0.65, respectively. For instance, in the southern Romania, the precipitation was moderate compared with eastern and western areas, and the stations were further divided into zones E and F by NDVI, where NDVI in zone F was 28% higher than that in zone E. Zones H and I had much higher average soil moisture than other zones, and they were divided by temperature with a threshold 22. For instance, stations in northeastern Poland were divided into zone H and I, and their average soil moisture were 0.58 and 0.42, respectively.

5.3.2. Frozen seasons

Fig. 7 shows the geographically optimal zones of soil moisture during the frozen season. The soil moisture monitoring stations were grouped into eleven spatial zones based on five variables using the GOZH model. The slope was the most important variable for determining the optimal zones. According to the slope value higher or lower than 0.33, soil moisture stations can be divided into two parts. Area with a slope value higher than 0.33 were generally mountainous areas. Compared with the

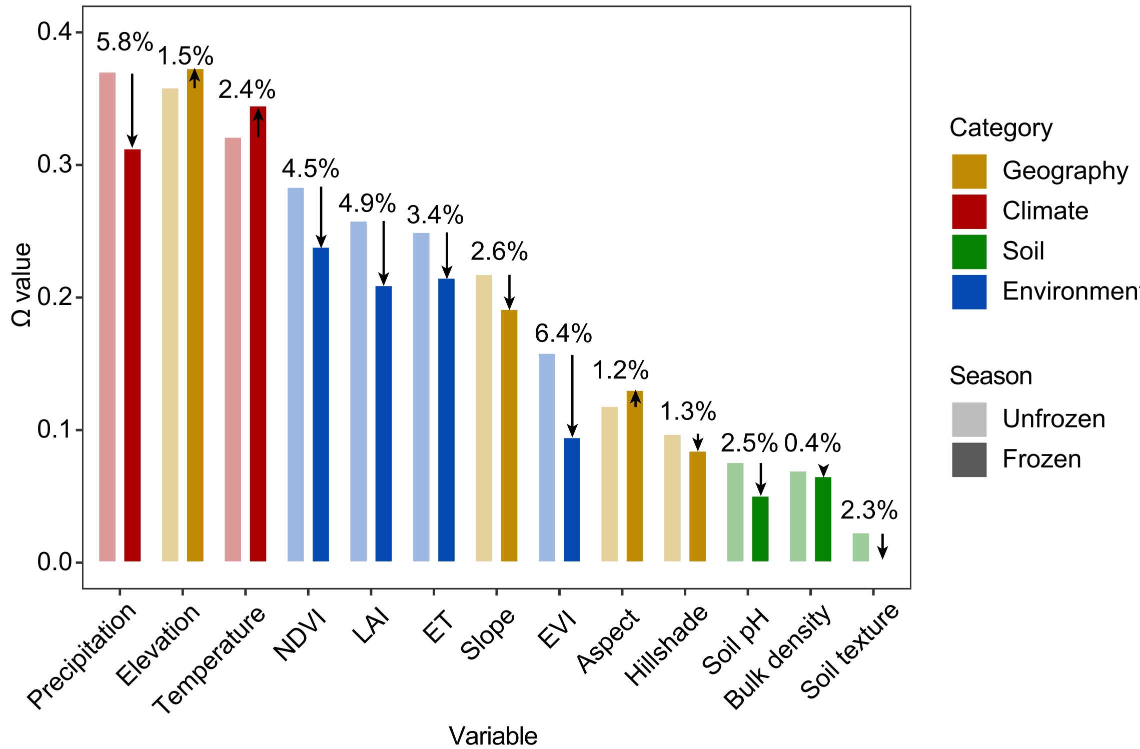

Fig. 5. Ω values of explanatory variables on spatial patterns of seasonal soil moisture and the comparison between unfrozen and frozen seasons.

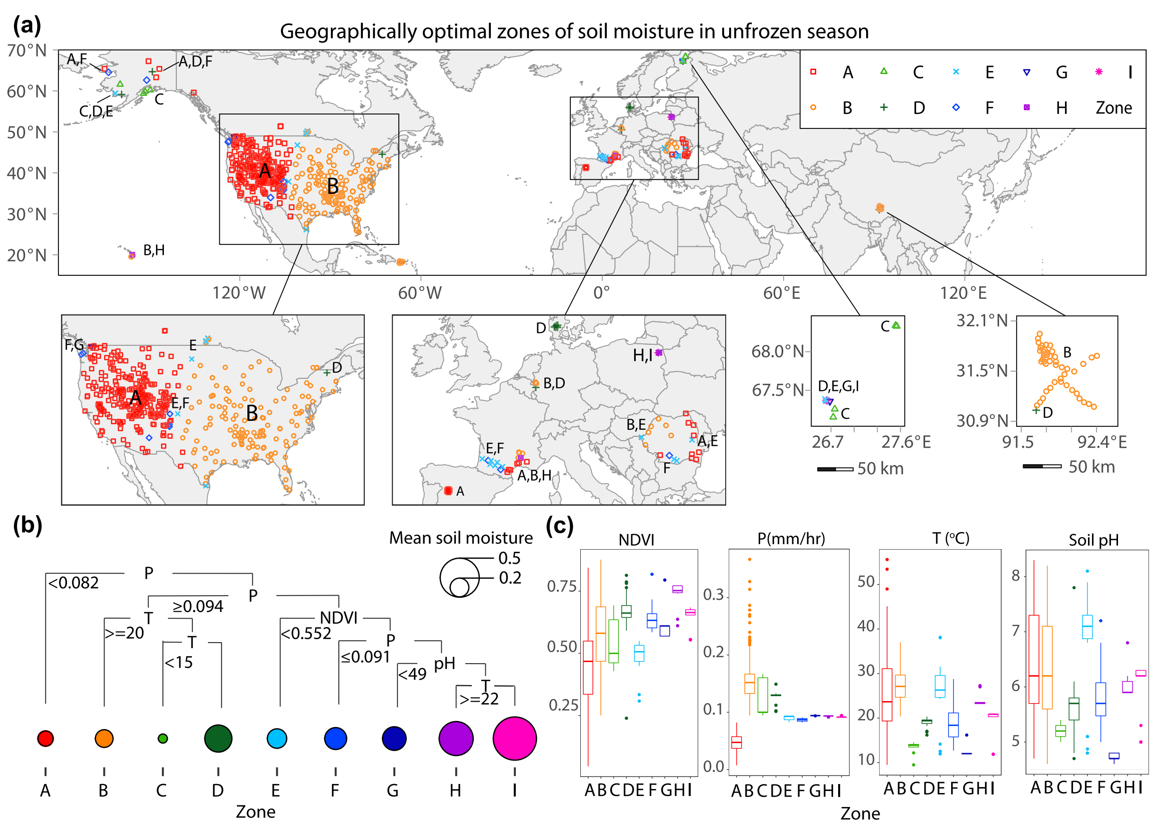

Fig. 6. Geographically optimal zones of soil moisture in the unfrozen season identified using the GOZH model (a), the process of identifying optimal zones (b), and statistical summaries of explanatory variables within zones for explaining characteristics of zones (c).

unfrozen season, terrain conditions had higher impacts on soil moisture disparities during the frozen season. In addition, ET and soil bulk density were variables that further divided zones in the second layers (Fig. 7 b).

According to variables in top two layers, including slope, ET, and soil bulk density, the eleven zones could be classified into four groups.

The first group, consisting of zones A and B, was characterized in

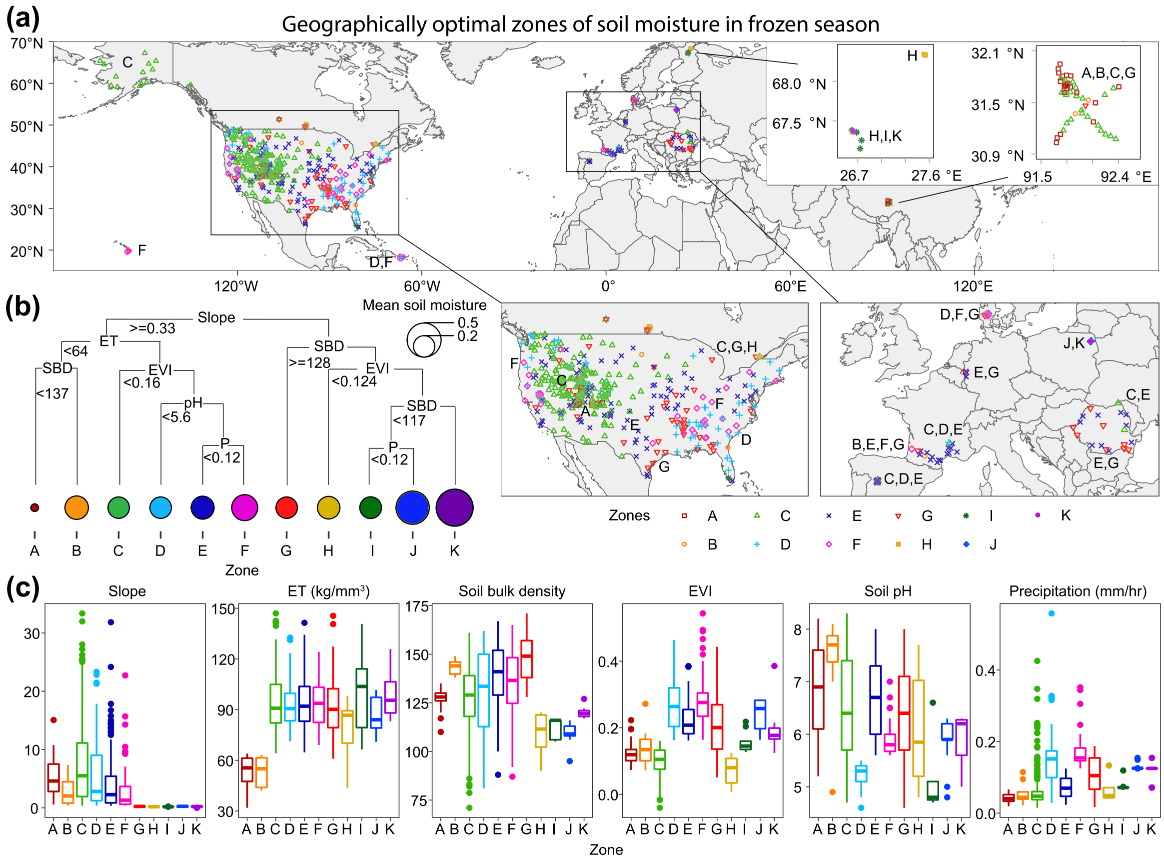

Fig. 7. Geographically optimal zones of soil moisture in the frozen season identified using the GPZH model (A), the process of identifying optimal zones (B), and statistical summaries of explanatory variables within zones for explaining characteristics of zones (C).

high slope and low ET. Zones A and B were divided by soil bulk density, where zone A had low soil bulk density and the lowest average soil moisture among all zones during the frozen season. Most of the stations in this group were located in the Rocky Mountains area in the United States and eastern Tibetan Plateau, China.

The second group, including zones C, D, E, and F, were also located in mountainous areas (slop > 0.33), but had relatively high ET (ET > 64 kg/mm^3). For instance, in the mountainous areas of the western United States, most stations were located in zones C and E, where the average slope were 7.38 and 4.27, respectively. EVI was the variable dividing zones C and E. The average EVI in zone C and E was 0.098 and 0.30, respectively. This means that in addition to slope and ET, vegetation was an essential variable controlling the spatial variability of soil moisture in this region. In addition to the mountainous areas of the western United States, zone C was also distributed in Alaska, eastern Romania, and eastern Tibetan Plateau, China, and zone E was also located in the western Spain, southern France, western Germany, and Romania. In the western and eastern coastal areas of the United States, stations were located in zones D, E and zone F. They were divided by soil pH and precipitation. In zone D, soil pH was lower than 5.6 and the average pH was 0.53, but it was higher than 5.6 in zones E and F. Zones E and F were divided by the 0.12 mm/hr of the precipitation. The average precipitation in zone E and F was 0.107 and 0.17, respectively.

The third group, containing zone G, had low slope and high soil bulk density. Stations in this group were primarily located in the southern United States and Romania. The average slope in zone G is 0.20, which was much lower than its neighbouring zones, such as zone E (slope = 4.28).

The last group, including zones H, I, J, and K, had low slope and soil bulk density. Stations in this group were generally distributed in the northeastern Poland and northern Finland. For instance, stations in northeastern Poland were divided into zones J and K. The soil bulk density controlled the spatial disparities in these two zones. The average soil bulk density in zones J and K were 111.55 and 109.44, respectively. In stations in the northern Finland, EVI and soil bulk density controlled spatial patterns of soil moisture.

5.4. *Determinants of spatial disparities and seasonal effects*

Table 3 shows overall Ω values of explanatory variables on spatial patterns of soil moisture investigated using the GOZH model and contributions of variables to the overall Ω values during unfrozen and frozen seasons. In general, overall Ω values were 47.62% and 47.69% during unfrozen and frozen seasons, respectively. This means that variables tended to have similar total contributions to spatial patterns of soil moisture during both seasons.

During the unfrozen season, climate variables had higher contributions to the overall Ω value, where contributions of precipitation and temperature were 20.99% and 7.90%, respectively. The contribution of precipitation accounted for 44.08% to the overall Ω value. In addition, NDVI and soil pH contributed 11.26% and 7.47%, respectively. All these contributions were lower than impacts of individual variables on spatial patterns of soil moisture. This means that explanatory variables had high interactive impacts on affecting patterns of soil moisture.

During the frozen season, spatial patterns of soil moisture were affected by slope, soil bulk density, ET, EVI, precipitation, and soil pH. The slope was closely associated with local terrain conditions, and it contributed 13.72% to patterns of soil moisture. Similar with the assessment of individual variables, geographical variables controlled the spatial variability of soil moisture during the frozen season. In addition to slope, soil bulk density, ET, EVI, precipitation, and soil pH contributed 12.62%, 6.58%, 6.31%, 4.04%, ad 4.01% to spatial patterns of soil

Table 3
Contributions of explanatory variables on spatial patterns of soil moisture and contributions to dividing optimal zones in unfrozen and frozen seasons.

Unfrozen season			Frozen season		
Variable	Contribution to spatial patterns	Contribution to dividing zones	Variable	Contribution to spatial patterns	Contribution to dividing zones
Precipitation	20.99%	72.22%	Slope	13.72%	54.55%
NDVI	11.26%	11.11%	Soil bulk density	12.62%	12.73%
Temperature	7.90%	11.11%	ET	6.58%	14.55%
Soil pH	7.47%	5.56%	EVI	6.31%	10.91%
Overall Ω	47.62%	/	Precipitation	4.04%	3.64%
			Soil pH	4.01%	3.64%
			Overall Ω	47.69%	/

moisture during the frozen season, respectively.

In addition, Table 3 and Fig. 8 demonstrate contributions of variables to dividing optimal zones in unfrozen and frozen seasons. They had similar trends with contributions of variables to overall Ω values. For instance, precipitation contributed 72.22% to dividing zones during the unfrozen season, and slope contributed 54.55% to the dicision of zones during the frozen season.

5.5. Model evaluation

The performance of the GOZH model in investigating spatial heterogeneity in the large-scale soil moisture was evaluated from four aspects: exploring individual variables, assessing multiple spatial variables with interactive effects, dealing with finely divided zones during spatial overlay, and the reliability of models. These aspects of the GOZH model were evaluated by comparing with the commonly used OPGD model during the unfrozen and frozen seasons. Fig. 9 a-d shows the spatial discretization process of the OPGD model for 12 variables, in addition to soil texture, which was a categorical variable containing six classes of texture. With the break number increase from 1 to 16, the PD, i.e., Q value, of all variables increased gradually. The optimal break numbers were selected when the increase rate was lower than 0.05. In this study, 9 and 12 were selected as the optimal break numbers of continuous variables during unfrozen and frozen seasons, respectively. Fig. 9 e and f shows OPGD-based PD values of individual variables to spatial patterns of soil moisture.

First, the GOZH model supports the derivation of the maximum spatial associations between response and explanatory variables through the identification of geographically optimal zones. The maximum spatial associations can accurately reveal the spatial heterogeneity of soil moisture. Therefore, the GOZH model is a reliable approach for examining spatial heterogeneity and exploring OPD of explanatory variables on spatial patterns of soil moisture.

In addition, the GOZH model can help reduce the underestimation of the PD values by the OPGD model as demonstrated by explorations of individual variables. Ranks of PD values explored by OPGD models during both unfrozen and frozen seasons were similar to those of GOZH models. For instance, precipitation and elevation were variables with both the highest Q (PD) and Ω (OPD) values during unfrozen and frozen seasons, respectively. However, the power of explanatory variables revealed by the GOZH model had a significant enhancement than the OPGD model. The average Ω values of individual variables were 80.9% and 68.2% higher than the average Q values during the unfrozen and frozen seasons, respectively.

Third, the GOZH model can effectively avoid the overestimation of the interactive impacts of multiple spatial variables on patterns of soil moisture compared with the OPGD model. Fig. 10 shows a model performance comparison between the GOZH and OPGD models in terms of the OPD/PD of variables and numbers of zones with the increased number of explanatory variables during the unfrozen and frozen seasons. In Fig. 10 a and c, average GOZH-based Ω values of individual variables were 0.20 and 0.18, and they were gradually increased to 0.48. In GOZH models, numbers of zones were not critically increased (Fig. 10 b and d), which indicated the robustness of GOZH models in the analysis of spatial heterogeneity. However, in Fig. 10 e and g, average OPGD-based Q values of individual variables were both 0.11, respectively, and they were rapidly increased to 0.99. Simultaneously, numbers of zones were also critically increased from 13 to 750 when the number of variables was higher than 1. The critically increased number of zones caused the very limited observations within zones and made the

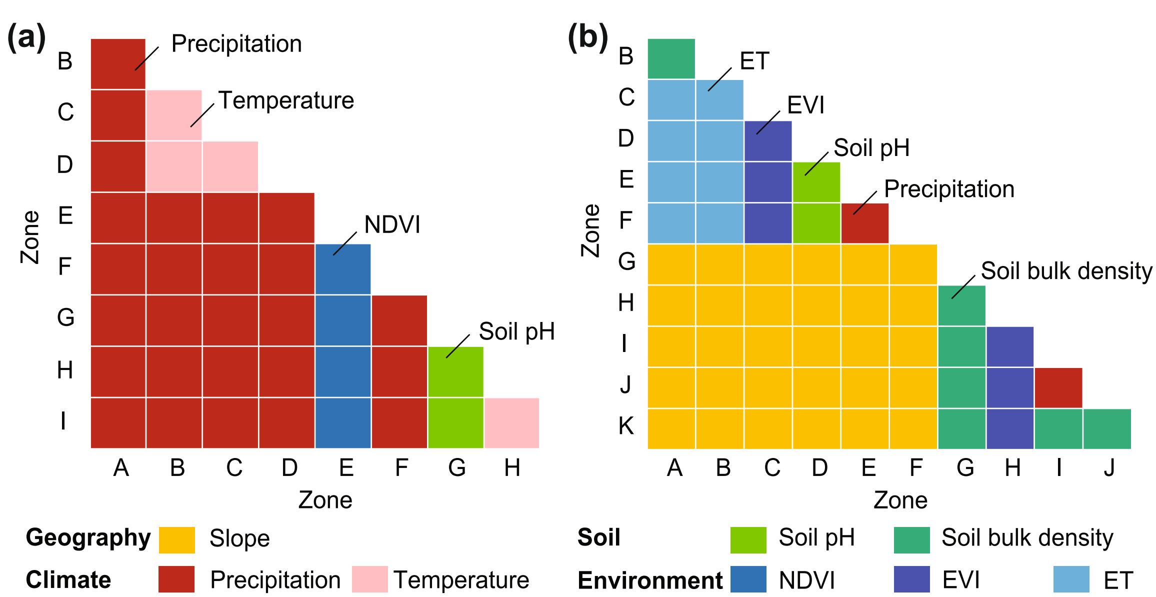

Fig. 8. Summary of explanatory variables used for dividing each pair of geographical optimal zones of soil moisture in unfrozen (a) and frozen (b) seasons.

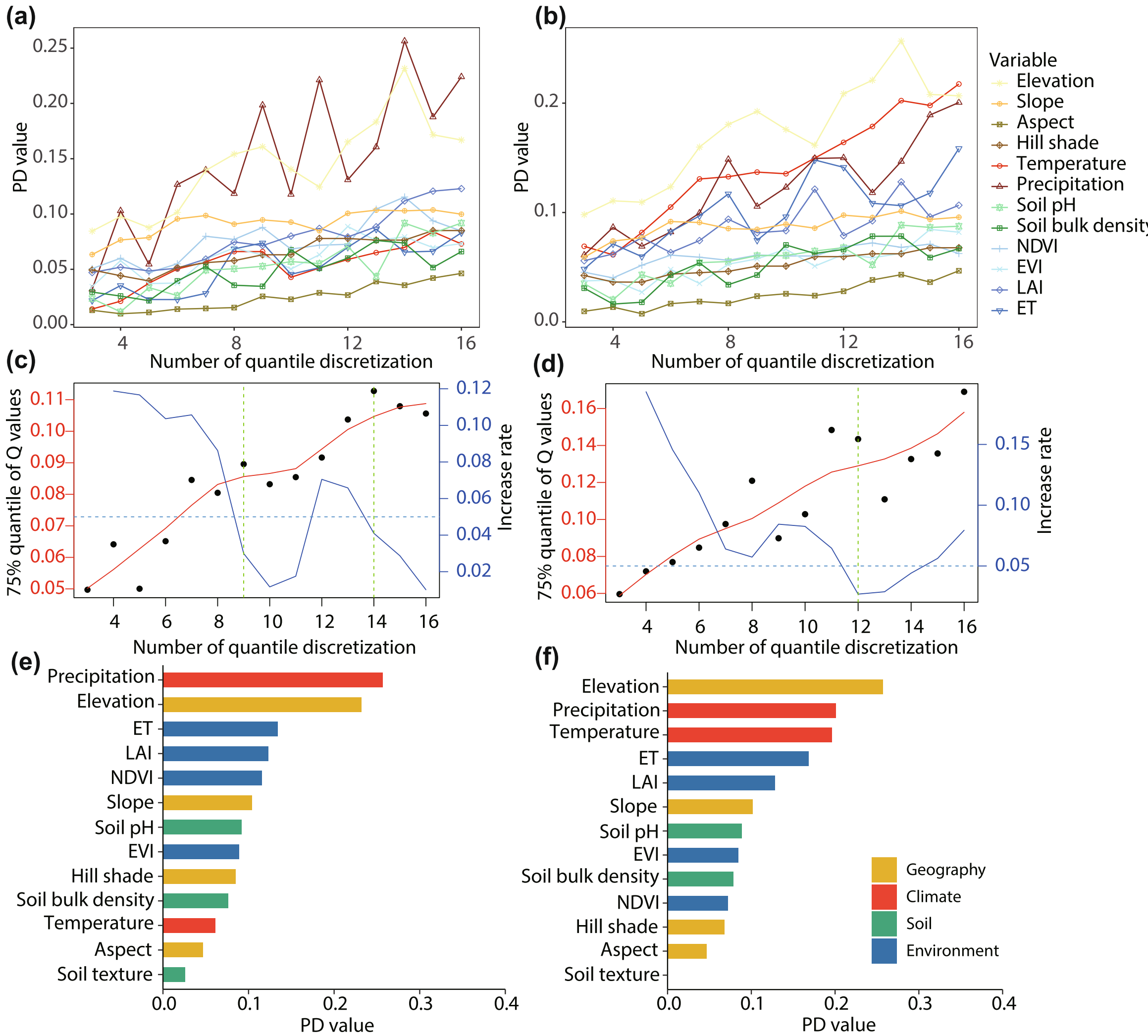

Fig. 9. Processes and results of the optimal parameters-based geographical detectors (OPGD) model for assessing power of determinants (PD) of soil moisture. PD of variables with different numbers of spatial discretization (a and b), processes of selecting optimal numbers of discretization (c and d), and PD of individual variables using optimal parameters (e and f) in unfrozen and frozen seasons, respectively.

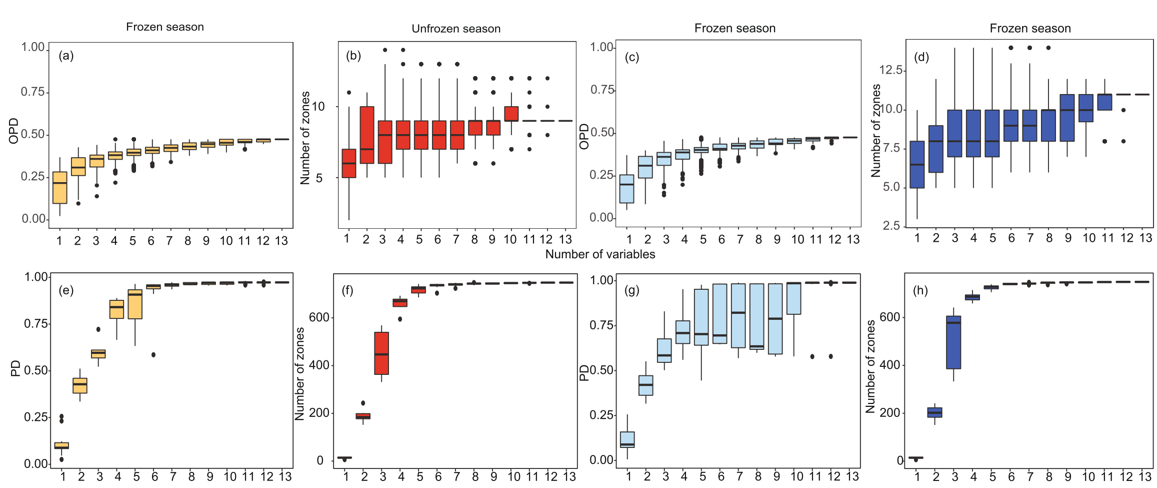

Fig. 10. Model performance comparison between GOZH and OPGD models: power of determinants (PD) and number of zones in the unfrozen and frozen seasons investigated using the GOZH model (a-d), and that investigated using the OPGD model (e-h).

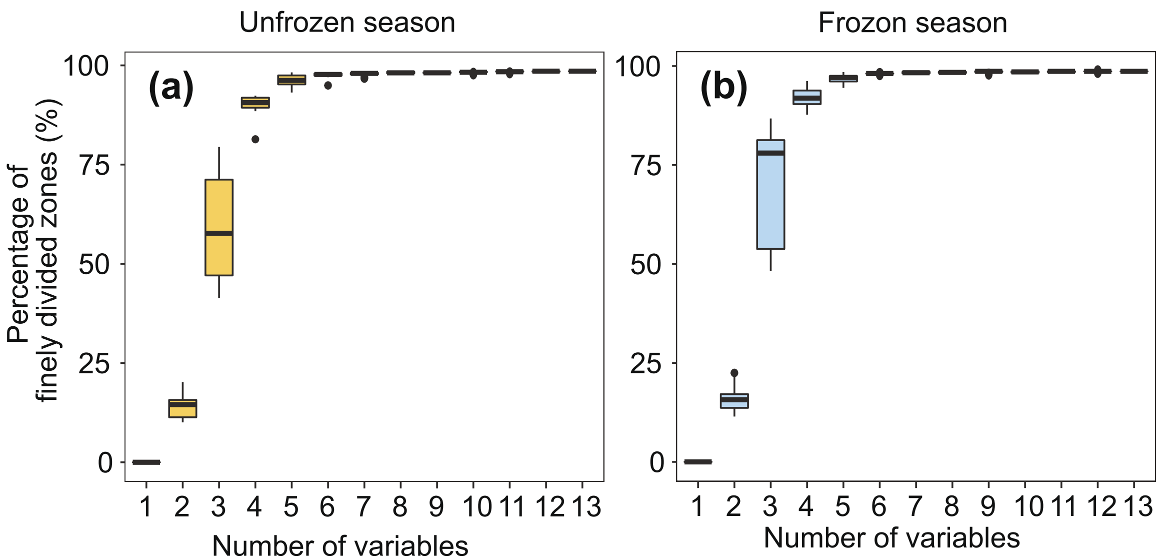

Fig. 11. Percentages of finely divided zones (FDZ), which are zones contain only an observation, are critically increased in the OPGD models for unfrozen (A) and frozen (B) seasons, when the number of variables is higher than two. On the contrary, percentages of FDZ in GOZH-based results are zero for both seasons.

increased Q values unreliable.

Finally, the GOZH model can eliminate finely divided zones (FDZs) that are common in the interactive variable assessment of SSH models. Fig. 11 shows the percentages of FDZs in OPGD models. In OPGD models, when examining interactive impacts of any two variables on soil moisture patterns, about 20% of zones would be FDZs, where there was only one observation in each FDZs. When the number of variables was higher than 5, nearly all zones would be FDZs. The extremely high percentage of FDZs cannot explain the real PD of variables. Thus, the analysis of FDZs demonstrates that the rapidly increased Q values of the interaction of multiple variables in OPGD models were not reliable when three or more variables were used in models. The analysis also explains why only two variables were considered in the interaction analysis in previous OPGD-based studies. When the variable number exceeds 2, most of the zones only have one or a few observations, which makes it difficult to reveal the real PD of the interaction of variables. However, in GOZH models, no FDZs existed no matter how many explanatory variables were used. The analysis of FDZs can further confirm the reliability and robustness of the GOZH model in the analysis of spatial heterogeneity.

6. Discussion

6.1. Methodological contributions

This study proposed a GOZH model to explore the spatial variability of soil moisture in the Northern Hemisphere. The GOZH model has following advantages in spatial determinant explorations. First, in the GOZH model, an OPD indicator was developed reveal the maximum spatial associations between soil moisture variability and determinants. Second, the optimal geographical zones can be derived from the explanatory variables. Third, no statistical assumptions are required in the GOZH model. Finally, the model validation in the study has demonstrated that the GOZH model can effectively explore spatial determinants of soil moisture through avoiding the underestimation of individual variables, overestimation of multiple variables, and finely divide zones.

6.2. Complex spatial heterogeneity of soil moisture patterns

This study revealed the complex spatial heterogeneity of soil moisture during unfrozen and frozen season at a global scale. The complexity of spatial heterogeneity of soil moisture patterns can be explained in following aspects. First, spatial patterns of soil moisture had significant regional disparities that were closely associated with regional geographical, climate, soil, and environmental conditions. The soil moisture monitoring stations in the Northern Hemisphere can be divided into nine and eleven zones during the unfrozen and frozen seasons, respectively. Explanatory variables tended to be similar within zones, and significantly varied among different zones.

In addition, determinants of spatial patterns of large-scale soil moisture have seasonal characteristics. On one hand, the spatial heterogeneity during unfrozen and frozen seasons has similarities. The overall Ω values that examined the maximum PD of four categories of explanatory variables in both seasons were approximate 48%, and they were affected by the interaction of multiple spatial variables. These results revealed the complexity of spatial heterogeneity in soil moisture, that only half of the heterogeneity could be explained by geographical, climate, soil, and environmental variables.

On the other hand, the spatial heterogeneity of soil moisture during the frozen season was more complex than that during the unfrozen season. First, spatial distributions of geographically optimal zones during the frozen season was much more complex than those during the unfrozen season, which appeared in most monitoring stations in the North America, Europe, and China. An exception was Alaska, where stations were divided into five zones during the unfrozen season, but they were located in a zone during the frozen season. Second, more explanatory variables are required to identify geographical optimal zones and estimate Ω values, where numbers of required explanatory variables during the unfrozen and frozen seasons were four and six, respectively. Finally, climate variables, including precipitation and temperature, are predominant variables of spatial patterns of soil moisture during the unfrozen season, accounting for 60.7% of the overall Ω value,but spatial patterns of soil moisture were affected by all four categories of variables during the unfrozen season, where the primary variable, slope, only accounted for 28.8% of the overall Ω value. Precipitation contributed 72.2% to dividing zones. During the frozen season, geographical variables controlled the spatial variability of soil moisture, and slope contributed 54.55% to dividing zones. Soil properties, including soil bulk density and soil pH also determines the geographical optimal zones during the frozen season. The slope can explain 19.16% of soil moisture during the frozen season, which is lower than its contribution to the geographically optimal zones combining with other variables. This result is consistent with previous works, which geographical variables can not be the single determinant of soil moisture variability (Wilson et al., 2005).

Finally, it is more difficult to characterize and explain the spatial heterogeneity than temporal heterogeneity in soil moisture. For instance, previous studies have demonstrated that the fitness of temporal prediction of soil moisture could reach to 96% using deep learning models (Cai et al., 2019; Ahmed et al., 2021). However, the accuracy of spatial prediction was much lower than that in temporal predictions and the accuracy tended to decrease with the increased spatial scale. For instance, the fitness of spatial prediction at local scales were 0.41–0.84 in multiple studies (Badewa et al., 2018; Peng et al., 2017), and that at a global scale can only reached to 0.63 (Montzka et al., 2018), even models have been improved with the consideration of more soil information and characteristics, such as soil texture (Montzka et al., 2018). According to this study, the primary reason of the lower accuracy of spatial prediction is the complex spatial heterogeneity that only about a half of the spatial heterogeneity of soil moisture can be explained by explanatory variables.

6.3. Temporal variations of soil moisture determinants

The temporal phase of soil moisture determinants was investigated from following three aspects. First, the spatial heterogeneity and determinants was varied from April 2015 to December 2017. Generally, climate variables, i.e. precipitation and temperature, have the highest spatial associations with soil moisture during most periods but the associations were varied in different months, which is consistent with previous findings (Wang et al., 2017). Explanatory variables had the highest explanatory power to soil moisture in November 2016, where precipitation can impact 58% of soil moisture, and elevation and temperature can explain 55% and 46% of the spatial variability of soil moisture. However, variables can only explain up to 5% of soil moisture in February 2017. Therefore, a relatively long time period, such as a half year, is recommended for reliable explorations of variability and determinants of spatial patterns of soil moisture.

Second, soil moisture spatial variability has a strong monthly pattern. In the frozen-unfrozen season-changing months, i.e., March and April, spatial associations between patterns of soil moisture and explanatory variables generally have the highest Ω value, and the Ω value of climate variables have the highest improvement in this period. Previous studies have concluded that during the transition phrase, climate variables become more important to soil moisture variability likely because the alternation of cold and warm days controlled by weather variability (Kang et al., 2010; Wei et al., 2019). The least interpretable period of spatial patterns of soil moisture is the middle frozen seasons.

Third, seasonal effects were identified in the spatial heterogeneity of soil moisture, results show that the spatial pattern of soil moisture is more interpretable during the unfrozen season than during the frozen season. The average Ω value of individual variables during the unfrozen season (20.0%) is 12.4% higher than that during the frozen season (17.8%). Different from most variables, Ω values of temperature and two geographically variables, elevation and aspect, are increased from unfrozen to frozen season. This finding is consistent with previous research that in cold weather, soil moisture variability is strongly associated with global warming, and the impacts of temperature can be more significant (Kang et al., 2010; Wei et al., 2019). Studies also found geographical variables is the main driver of soil moisture during the winter when soil is frozen which particular because its association to water table (Rosenbaum et al., 2012). During the unfrozen season, the impacts of geographical variables are negligible due to the low water table (Chaney et al., 2015). During the frozen season, with the low ET, the water table increases and closes to the surface, which enables higher impacts of the groundwater and subsurface flow on the soil moisture variability (Western et al., 1998; Rosenbaum et al., 2012).

6.4. Contributions of heterogeneity and geographical zones to soil moisture studies

Findings about the spatial heterogeneity of soil moisture in this study can help optimize the design of soil moisture monitoring network, spatial down-scaling of soil moisture data, and accurate inversion of surface parameters from soil moisture.

First, the network design of soil moisture can be optimized with the improved understanding of the spatial heterogeneity and determinants of regional disparities of soil moisture identified in this study. Due to the complex heterogeneity of spatial soil moisture, most existing in situ observation networks rarely provide sufficient coverage to capture soil moisture variability at a watershed scale. Thus, it is critically required to develop a systematic approach to soil moisture network design in order to accurately capture soil moisture information in the watershed space with a minimum number of sensors. It was found that the current (simulated and observed) network of soil moisture detectors underestimates the average spatial heterogeneity (Zhuo et al., 2020). The analysis of the determinants of soil moisture heterogeneity and the spatial partitioning results from the GOZH model can be used to inform the development of new techniques for ground-based measurement network design. The intended network design can take into account the spatial variability of soil moisture.

Second, the spatial down-scaling of the soil moisture data requires the spatial heterogeneity information of large-scale soil moisture monitoring data. The coarse resolution of soil moisture remote sensing products limits its application at fine scales, which introduces the need for their spatial down-scaling (Chaney et al., 2015). A series of downscaling methods had been developed to improve resolutions of soil moisture products using multi-source auxiliary data and various methods, such as statistical models, geospatial models, machine learning, deep learning, and hybrid models (Peng et al., 2017). However, due to the existence of spatial heterogeneity of soil moisture, the accuracy of spatial prediction has been lower than that of temporal prediction (Badewa et al., 2018; Peng et al., 2017; Montzka et al., 2018). It is also a challenge to quantitatively assess the large differences in soil moisture determinants in different regions (Molero et al., 2018). The geographically optimal zones of soil moisture obtained using the GOZH model, and the control factors in different regions can effectively guide the spatial down-scaling process. Our study shows that geographical variables are the most important factors to soil moisture in the frozen season. For example, soil moisture heterogeneity in the east and west of North America is controlled by the slope. Therefore, greater weight should be given to geographical variables during spatial down-scaling. In the unfrozen season, environmental and climate variables are essential to soil moisture. Precipitation determines the soil moisture in the western United States, while NDVI determines soil moisture in the central United States.

Finally, understanding soil moisture heterogeneity over different geographical zones can also support the accurate inversion of surface parameters from soil moisture satellite data. The limited knowledge of regional differences in soil moisture and its determinants poses a challenge to calibrate ground roughness parameterization schemes with ground observation data. Obtaining information on soil moisture heterogeneity can improve the accuracy and the geographical transferability of the parameterization scheme. (Verhoest et al., 2008).

There are still limitations of this study. First, the scale effect between soil moisture in situ data and remote sensing images were not considered in this study. The explanatory variables are derived from the pixels in the position of the soil moisture monitor stations. Spatial heterogeneity of soil moisture at stations in the Northern Hemisphere is much higher than that of data within grids of explanatory variables, e.g., 90 m or 250 m. Therefore, we assume spatial analysis in the study will not be affected by the scale effects of explanatory variables derived from remote sensing or grid data. In addition, some explanatory variables, for example, elevation and slope, may be represented by a zone with an area larger than the size of grid in the images (Jasiewicz and Stepinski, 2013). In this case, data at surrounding grids need to be considered for deriving explanatory variables at stations. From the perspective of spatial heterogeneity models, approaches can be developed for more effective use of continuous variables in spatial heterogeneity models. For instance, the spatial association detector (SPADE) (Cang and Luo, 2018) and the interactive detector for spatial association (IDSA) (Song and Wu, 2021) models were developed to compare zonal and global spatial dependence, i.e., spatial autocorrelation of data, instead of zonal and global variance, for computing the PD values. The K-means (Likas et al., 2003; Hartigan and Wong, 1979) and hierarchical clustering (Johnson, 1967) methods also can be used to derive spatial zones with the continuous explanatory variables. Finally, the division of the frozen and unfrozen seasons in this study may introduce uncertainty. The study aims to explore the soil moisture variability in the Northern Hemisphere. The frozen/ unfrozen months were unified in the whole study area since most stations are located in the mid-latitude area and only thawed soil moisture data were selected and analyzed. However, some stations are located in the high latitude area like Alaska, where the unfrozen/frozen season of soil moisture may be different from other areas. Thus, further studies may explore the soil moisture variability in different climate zones.

7. Conclusion

This study developed a geographically optimal zones-based heterogeneity (GOZH) model to explore the spatial variability of soil moisture in the Northern Hemisphere. In the GOZH model, the optimal power of determinant (OPD) indicator can reveal the maximum spatial associations, and the spatial determinants can be effectively explored through avoiding the underestimation of individual variables, overestimation of multiple variables, and finely divide zones.

The GOZH model was implemented to explore the spatial and temporal patterns of soil moisture variability. Results shows that in the frozen-unfrozen season-changing months, spatial associations between patterns of soil moisture and explanatory variables generally have the highest OPD value especially for climate variables. The average OPD value of individual variables during the unfrozen season (20.0%) is higher than that during the frozen season (17.8%). In addition, geographically optimal zones and corresponding determinants of soil moisture were revealed by the interactive of explanatory variables. Variables have similar contributions to spatial pattern of soil moisture during two seasons. At a global scale, the combinations of determinants can explain about 48% of the spatial pattern of soil moisture. During the unfrozen season, climate variables, including precipitation and temperature, have the highest contributions to the overall OPD value. During the frozen season, geographical variables (e.g., slope) controlled the spatial variability of soil moisture.

This study can provide a deep understanding of variability and determinants of soil moisture at a global scale. The knowledge of soil moisture determinants can be better used in situ network design, spatial down-scaling of soil moisture. In addition, the results can also be applied to the evaluate soil moisture in satellite imagery and the accurate inversion of surface parameters from satellite data on soil moisture.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by funding from the Curtin University and the China Scholarship Council.

References

Ahmed, A., Deo, R.C., Raj, N., Ghahramani, A., Feng, Q., Yin, Z., Yang, L., 2021. Deep learning forecasts of soil moisture: Convolutional neural network and gated recurrent unit models coupled with satellite-derived modis, observations and synoptic-scale climate index data. Remote Sens. 13 (4), 554.
Al-Yaari, A., Wigneron, J.P., Dorigo, W., Colliander, A., Pellarin, T., Hahn, S., Mialon, A., Richaume, P., Fernandez-Moran, R., Fan, L., Kerr, Y.H., De Lannoy, G., 2019. Assessment and inter-comparison of recently developed/reprocessed microwave satellite soil moisture products using ISMN ground-based measurements. Remote Sens. Environ. 224 (February), 289–303.
Albergel, C., Calvet, J.-C., Rosnay, P. d., Balsamo, G., Wagner, W., Hasenauer, S., Naeimi, V., Martin, E., Bazile, E., Bouyssel, F., et al., 2010. Cross-evaluation of modelled and remotely sensed surface soil moisture with in situ data in southwestern France. Hydrol. Earth Syst. Sci. 14 (11), 2177–2191.
Albergel, C., De Rosnay, P., Gruhier, C., Muñoz-Sabater, J., Hasenauer, S., Isaksen, L., Kerr, Y., Wagner, W., 2012. Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations. Remote Sens. Environ. 118, 215–226.
Albergel, C., Rüdiger, C., Carrer, D., Calvet, J.-C., Fritz, N., Naeimi, V., Bartalis, Z., Hasenauer, S., 2009. An evaluation of ASCAT surface soil moisture products with in-situ observations in Southwestern France. Hydrol. Earth Syst. Sci. 13 (2), 115–124 https://hess.copernicus.org/articles/13/115/2009/.
Albergel, C., Rüdiger, C., Pellarin, T., Calvet, J.-C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., Piguet, B., Martin, E., 2008. From near-surface to root-zone soil moisture using an exponential filter: an assessment of the method based on in-situ observations and model simulations. Hydrol. Earth Syst. Sci. 12 (6), 1323–1337.
Babaeian, E., Sadeghi, M., Franz, T.E., Jones, S., Tuller, M., 2018. Mapping soil moisture with the optical trapezoid model (optram) based on long-term modis observations. Remote Sens. Environ. 211, 425–440.
Badewa, E., Unc, A., Cheema, M., Kavanagh, V., Galagedara, L., 2018. Soil moisture mapping using multi-frequency and multi-coil electromagnetic induction sensors on managed podzols. Agronomy 8 (10), 224.
Baroni, G., Ortuani, B., Facchi, A., Gandolfi, C., 2013. The role of vegetation and soil properties on the spatio-temporal variability of the surface soil moisture in a maize-cropped field. J. Hydrol. 489, 148–159.
Bell, J.E., Palecki, M.A., Baker, C.B., Collins, W.G., Lawrimore, J.H., Leeper, R.D., Hall, M.E., Kochendorfer, J., Meyers, T.P., Wilson, T., et al., 2013. Us climate reference network soil moisture and temperature observations. Journal of Hydrometeorology 14 (3), 977–988.
Bell, K.R., Blanchard, B., Schmugge, T., Witczak, M., 1980. Analysis of surface moisture variations within large-field sites. Water Resour. Res. 16 (4), 796–810.
Berg, A., Sheffield, J., Milly, P.C., 2017. Divergent surface and total soil moisture projections under global warming. Geophys. Res. Lett. 44 (1), 236–244.
Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 2017. Classification and regression trees. Routledge.
Brocca, L., Melone, F., Moramarco, T., Morbidelli, R., 2010. Spatial-temporal variability of soil moisture and its estimation across scales. Water Resour. Res. 46 (2).
Brocca, L., Melone, F., Moramarco, T., Wagner, W., Hasenauer, S., 2010. Ascat soil wetness index validation through in situ and modeled soil moisture data in central italy. Remote Sens. Environ. 114 (11), 2745–2755.
Cai, Y., Zheng, W., Zhang, X., Zhangzhong, L., Xue, X., 2019. Research on soil moisture prediction model based on deep learning. PloS one 14 (4), e0214508.
Cang, X., Luo, W., 2018. Spatial association detector (SPADE). International Journal of Geographical Information Science 32 (10), 2055–2075. https://doi.org/10.1080/13658816.2018.1476693.
Cang, X., Luo, W., 2018. Spatial association detector (spade). International Journal of Geographical Information Science 32 (10), 2055–2075.
Chaney, N.W., Roundy, J.K., Herrera-Estrada, J.E., Wood, E.F., 2015. High-resolution modeling of the spatial heterogeneity of soil moisture: Applications in network design. Water resources research 51 (1), 619–638.
Dari, J., Morbidelli, R., Saltalippi, C., Massari, C., Brocca, L., 2019. Spatial-temporal variability of soil moisture: Addressing the monitoring at the catchment scale. Journal of Hydrology 570 (October 2018), 436–444.
Chen, R., Yan, D., Wen, A., Shi, Z., Chen, J., Liu, Y., Chen, T., 2021. The regional difference in engineering-control and tillage factors of Chinese Soil Loss Equation. J. Mount. Sci. 18 (3), 658–670.
Das, N.N., Mohanty, B.P., 2008. Temporal dynamics of psr-based soil moisture across spatial scales in an agricultural landscape during smex02: A wavelet approach. Remote Sens. Environ. 112 (2), 522–534.
Didan, K., Munoz, A.B., Solano, R., Huete, A., 2015. MODIS Vegetation Index User 's Guide (Collection 6) 2015 (May), 31.
Dorigo, W., Himmelbauer, I., Aberer, D., Schremmer, L., Petrakovic, I., Zappa, L., Preimesberger, W., Xaver, A., Annor, F., Ardö, J., Baldocchi, D., Blöschl, G., Bogena, H., Brocca, L., Calvet, J.-C., Camarero, J., Capello, G., Choi, M., Cosh, M., Demarty, J., van de Giesen, N., Hajdu, I., Jensen, K., Kanniah, K.D., de Kat, I., Kirchengast, G., Rai, P.K., Kyrouac, J., Larson, K., Liu, S., Loew, A., Moghaddam, M., Martínez Fernández, J., Mattar Bader, C., Morbidelli, R., Musial, J., Osenga, E., Palecki, M., Pfeil, I., Powers, J., Ikonen, J., Robock, A., Rüdiger, C., Rummel, U., Strobel, M., Su, Z., Sullivan, R., Tagesson, T., Vreugdenhil, M., Walker, J., Wigneron, J.P., Woods, M., Yang, K., Zhang, X., Zreda, M., Dietrich, S., Gruber, A., van Oevelen, P., Wagner, W., Scipal, K., Drusch, M., Sabia, R., 2021. The International Soil Moisture Network: serving Earth system science for over a decade. Hydrology and Earth System Sciences Discussions (January), 1–83.
Dorigo, W.A., Gruber, A., De Jeu, R.A.M., Wagner, W., Stacke, T., Loew, A., Albergel, C., Brocca, L., Chung, D., Parinussa, R.M., Kidd, R., 2015. Evaluation of the ESA CCI soil

moisture product using ground-based observations. Remote Sens. Environ. 162, 380–395 https://www.sciencedirect.com/science/article/pii/S0034425714002727.
Dorigo, W.A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., Gruber, A., Drusch, M., Mecklenburg, S., van Oevelen, P., Robock, A., Jackson, T., 2011. The International Soil Moisture Network: a data hosting facility for global in situ soil moisture measurements. Hydrol. Earth Syst. Sci. 15 (5), 1675–1698 https://hess.copernicus.org/articles/15/1675/2011/.
Elkhrachy, I., 2018. Vertical accuracy assessment for SRTM and ASTER Digital Elevation Models: A case study of Najran city. Saudi Arabia. Ain Shams Engineering Journal 9 (4), 1807–1817 https://www.sciencedirect.com/science/article/pii/S2090447917300084.
Entekhabi, D., Njoku, E.G., O'Neill, P.E., Kellogg, K.H., Crow, W.T., Edelstein, W.N., Entin, J.K., Goodman, S.D., Jackson, T.J., Johnson, J., et al., 2010. The soil moisture active passive (smap) mission. Proc. IEEE 98 (5), 704–716.
Entin, J.K., Robock, A., Vinnikov, K.Y., Hollinger, S.E., Liu, S., Namkhai, A., 2000. Meteorologic i Land Surface. J. Geophys. Res. 105 (D9), 11865–11877.
Famiglietti, J., Wood, E.F., 1994. Multiscale modeling of spatially variable water and energy balance processes. Water Resour. Res. 30 (11), 3061–3078.
Famiglietti, J.S., Ryu, D., Berg, A.A., Rodell, M., Jackson, T.J., 2008. Field observations of soil moisture variability across scales. Water Resour. Res. 44 (1).
Fang, H., Baret, F., Plummer, S., Schaepman-Strub, G., 2019. An Overview of Global Leaf Area Index (LAI): Methods, Products, Validation, and Applications. Rev. Geophys. 57 (3), 739–799.
Green, J.K., Seneviratne, S.I., Berg, A.M., Findell, K.L., Hagemann, S., Lawrence, D.M., Gentine, P., 2019. Large influence of soil moisture on long-term terrestrial carbon uptake. Nature 565 (7740), 476–479.
Gruber, A., Dorigo, W.A., Zwieback, S., Xaver, A., Wagner, W., 2013. Characterizing Coarse-Scale Representativeness of in situ Soil Moisture Measurements from the International Soil Moisture Network. Vadose Zone Journal 12 (2). https://doi.org/10.2136/vzj2012.0170.
Han, J., Mao, K., Xu, T., Guo, J., Zuo, Z., Gao, C., 2018. A soil moisture estimation framework based on the cart algorithm and its application in china. Journal of hydrology 563, 65–75.
Hartigan, J.A., Wong, M.A., 1979. Algorithm as 136: A k-means clustering algorithm. Journal of the royal statistical society. series c (applied statistics) 28 (1), 100–108.
Hashimoto, H., Dungan, J.L., White, M.A., Yang, F., Michaelis, A.R., Running, S.W., Nemani, R.R., 2008. Satellite-based estimation of surface vapor pressure deficits using modis land surface temperature data. Remote Sens. Environ. 112 (1), 142–155.
Holzman, M.E., Rivas, R., Bayala, M., 2014. Subsurface soil moisture estimation by vi–lst method. IEEE Geosci. Remote Sens. Lett. 11 (11), 1951–1955.
Hou, A.Y., Kakar, R.K., Neeck, S., Azarbarzin, A.A., Kummerow, C.D., Kojima, M., Oki, R., Nakamura, K., Iguchi, T., 2014. The global precipitation measurement mission. Bull. Am. Meteorol. Soc. 95 (5), 701–722.
Hu, W., Shen, Q., Zhai, X., Du, S., Zhang, X., 2021. Impact of environmental factors on the spatiotemporal variability of soil organic matter: a case study in a typical small Mollisol watershed of Northeast China. J. Soils Sediments 21 (2), 736–747.
Jasiewicz, J., Stepinski, T.F., 2013. Geomorphons–a pattern recognition approach to classification and mapping of landforms. Geomorphology 182, 147–156.
Jawson, S.D., Niemann, J.D., 2007. Spatial patterns from eof analysis of soil moisture at a large scale and their dependence on soil, land-use, and topographic properties. Adv. Water Resour. 30 (3), 366–381.
Jensen, K.H., Illangasekare, T.H., 2011. Hobe: A hydrological observatory. Vadose Zone Journal 10 (1), 1–7.
Jiang, Y., Weng, Q., 2017. Estimation of hourly and daily evapotranspiration and soil moisture using downscaled LST over various urban surfaces. GIScience & Remote Sensing 54 (1), 95–117. https://doi.org/10.1080/15481603.2016.1258971.
Johnson, S.C., 1967. Hierarchical clustering schemes. Psychometrika 32 (3), 241–254.
Joyce, R.J., Xie, P., 2011. Kalman filter–based cmorph. Journal of Hydrometeorology 12 (6), 1547–1563.
Kang, S., Xu, Y., You, Q., Flügel, W.-A., Pepin, N., Yao, T., 2010. Review of climate and cryospheric change in the tibetan plateau. Environmental research letters 5 (1), 015101.
Konare, A., Zakey, A., Solmon, F., Giorgi, F., Rauscher, S., Ibrah, S., Bi, X., 2008. A regional climate modeling study of the effect of desert dust on the west african monsoon. Journal of Geophysical Research: Atmospheres 113 (D12).
Kumar, S.V., Dirmeyer, P.A., Peters-Lidard, C.D., Bindlish, R., Bolten, J., 2018. Information theoretic evaluation of satellite soil moisture retrievals. Remote sensing of environment 204, 392–400.
Kusangaya, S., Toucher, M.L.W., van Garderen, E.A., Jewitt, G.P.W., 2016. An evaluation of how downscaled climate data represents historical precipitation characteristics beyond the means and variances. Global Planet. Change 144, 129–141 https://www.sciencedirect.com/science/article/pii/S0921818116301229.
Lei, F., Crow, W.T., Shen, H., Su, C.-H., Holmes, T.R., Parinussa, R.M., Wang, G., 2018. Assessment of the impact of spatial heterogeneity on microwave satellite soil moisture periodic error. Remote sensing of environment 205, 85–99.
Li, T., Chen, Y., Han, L., Cheng, L., Lv, Y., Fu, B., Feng, X., Wu, X., 2021. Shortened duration and reduced area of frozen soil in the northern hemisphere. The Innovation.
Li, X., Al-Yaari, A., Schwank, M., Fan, L., Frappart, F., Swenson, J., Wigneron, J.-P., 2020. Compared performances of smos-ic soil moisture and vegetation optical depth retrievals based on tau-omega and two-stream microwave emission models. Remote Sens. Environ. 236, 111502.
Liang, S., Fang, H., 2021. Quantitative analysis of driving factors in soil erosion using geographic detectors in Qiantang River catchment, Southeast China. J. Soils Sediments 21 (1), 134–147.
Likas, A., Vlassis, N., Verbeek, J.J., 2003. The global k-means clustering algorithm. Pattern recognition 36 (2), 451–461.
Liu, D., Mishra, A.K., Yu, Z., Yang, C., Konapala, G., Vu, T., 2017. Performance of SMAP, AMSR-E and LAI for weekly agricultural drought forecasting over continental United States. J. Hydrol. 553, 88–104 https://www.sciencedirect.com/science/article/pii/S0022169417305140.
Liu, Y., Chen, Y., Wu, Z., Wang, B., Wang, S., 2021. Geographical detector-based stratified regression kriging strategy for mapping soil organic carbon with high spatial heterogeneity. Catena 196 (December 2019).
Luo, P., Song, Y., Wu, P., 2021. Spatial disparities in trade-offs: economic and environmental impacts of road infrastructure on continental level. GIScience & Remote Sensing 58 (5), 756–775.
Ma, H., Zeng, J., Chen, N., Zhang, X., Cosh, M.H., Wang, W., 2019. Satellite surface soil moisture from smap, smos, amsr2 and esa cci: A comprehensive assessment using global ground-based observations. Remote Sens. Environ. 231, 111215.
Ma, H., Zeng, J., Zhang, X., Fu, P., Zheng, D., Wigneron, J.-P., Chen, N., Niyogi, D., 2021. Evaluation of six satellite- and model-based surface soil temperature datasets using global ground-based observations. Remote Sens. Environ. 264, 112605 https://www.sciencedirect.com/science/article/pii/S0034425721003254.
Martínez-Fernández, J., Ceballos, A., 2005. Mean soil moisture estimation using temporal stability analysis. J. Hydrol. 312 (1–4), 28–38.
McColl, K.A., Alemohammad, S.H., Akbar, R., Konings, A.G., Yueh, S., Entekhabi, D., 2017. The global distribution and dynamics of surface soil moisture. Nat. Geosci. 10 (2), 100–104.
McNairn, H., Jackson, T.J., Wiseman, G., Bélair, S., Berg, A., Bullock, P., Colliander, A., Cosh, M.H., Kim, S.-B., Magagi, R., et al., 2014. The soil moisture active passive validation experiment 2012 (smapvex12): Prelaunch calibration and validation of the smap soil moisture algorithms. IEEE Trans. Geosci. Remote Sens. 53 (5), 2784–2801.
Molero, B., Leroux, D., Richaume, P., Kerr, Y., Merlin, O., Cosh, M., Bindlish, R., 2018. Multi-timescale analysis of the spatial representativeness of in situ soil moisture data within satellite footprints. Journal of Geophysical Research: Atmospheres 123 (1), 3–21.
Molero, B., Leroux, D.J., Richaume, P., Kerr, Y.H., Merlin, O., Cosh, M.H., Bindlish, R., 2018. Multi-Timescale Analysis of the Spatial Representativeness of In Situ Soil Moisture Data within Satellite Footprints. Journal of Geophysical Research: Atmospheres 123 (1), 3–21 https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017JD027478.
Montzka, C., Rötzer, K., Bogena, H.R., Sanchez, N., Vereecken, H., 2018. A new soil moisture downscaling approach for smap, smos, and ascat by predicting sub-grid variability. Remote sensing 10 (3), 427.
Mu, Q., Zhao, M., Running, S.W., 2013. Modis global terrestrial evapotranspiration (et) product (nasa mod16a2/a3). Algorithm Theoretical Basis Document, Collection 5.
Nielsen, D., Biggar, J., Erh, K., et al., 1973. Spatial variability of field-measured soil-water properties. Hilgardia 42 (7), 215–259.
Niu, Y.J., Zhou, J.W., Yang, S.W., Wang, G.Z., Liu, L., Hua, L.M., may 2017. [Quantitative apportionment of slope aspect and altitude to soil moisture and temperature and plant distribution on alpine meadow]. Ying yong sheng tai xue bao = The journal of applied ecology 28 (5), 1489–1497. URL https://doi.org/10.13287/j.1001-9332.201705.032.
Ochsner, T.E., Linde, E., Haffner, M., Dong, J., 2019. Mesoscale Soil Moisture Patterns Revealed Using a Sparse In Situ Network and Regression Kriging. Water Resour. Res. 55 (6), 4785–4800.
Peng, J., Loew, A., Merlin, O., Verhoest, N.E., 2017. A review of spatial downscaling of satellite remotely sensed soil moisture. Rev. Geophys. 55 (2), 341–366.
Peng, J., Niesel, J., Loew, A., 2015. Evaluation of soil moisture downscaling using a simple thermal-based proxy–the remedhus network (spain) example. Hydrol. Earth Syst. Sci. 19 (12), 4765–4782.
Perry, M.A., Niemann, J.D., 2007. Analysis and estimation of soil moisture at the catchment scale using eofs. J. Hydrol. 334 (3–4), 388–404.
Peters-Lidard, C., Zion, M., Wood, E.F., 1997. A soil-vegetation-atmosphere transfer scheme for modeling spatially variable water and energy balance processes. Journal of Geophysical Research: Atmospheres 102 (D4), 4303–4324.
Planchon, O., 2000. A study of the coastal climates in france using temperature and precipitation data (1961–1990). Meteorological Applications 7 (3), 217–228.
Purdy, A.J., Fisher, J.B., Goulden, M.L., Colliander, A., Halverson, G., Tu, K., Famiglietti, J.S., 2018. SMAP soil moisture improves global evapotranspiration. Remote Sens. Environ. 219 (September), 1–14. https://doi.org/10.1016/j.rse.2018.09.023.
Qu, Y., Zhu, Z., Montzka, C., Chai, L., Liu, S., Ge, Y., Liu, J., Lu, Z., He, X., Zheng, J., et al., 2021. Inter-comparison of several soil moisture downscaling methods over the qinghai-tibet plateau, china. J. Hydrol. 592, 125616.
Quinn, P., Beven, K., Culf, A., 1995. The introduction of macroscale hydrological complexity into land surface-atmosphere transfer models and the effect on planetary boundary layer development. J. Hydrol. 166 (3–4), 421–444.
Redding, T.E., Hope, G.D., Fortin, M.J., Schmidt, M.G., Bailey, W.G., 2003. Spacial patterns of soil temperature and moisture across subalpine forest-clearcut edges in the southern interior of British Columbia. Can. J. Soil Sci. 83 (1), 121–130.
Romshoo, S.A., 2004. Geostatistical analysis of soil moisture measurements and remotely sensed data at different spatial scales. Environ. Geol. 45 (3), 339–349.
Rosenbaum, U., Bogena, H.R., Herbst, M., Huisman, J.A., Peterson, T.J., Weuthen, A., Western, A.W., Vereecken, H., 2012. Seasonal and event dynamics of spatial soil moisture patterns at the small catchment scale. Water Resour. Res. 48 (10).
Sawada, Y., 2018. Quantifying Drought Propagation from Soil Moisture to Vegetation Dynamics Using a Newly Developed Ecohydrological Land Reanalysis. Remote Sensing 10 (8) https://www.mdpi.com/2072-4292/10/8/1197.

Schaefer, G.L., Cosh, M.H., Jackson, T.J., 2007. The usda natural resources conservation service soil climate analysis network (scan). Journal of Atmospheric and Oceanic Technology 24 (12), 2073–2077.
Shellito, P.J., Small, E.E., Livneh, B., 2018. Controls on surface soil drying rates observed by SMAP and simulated by the Noah land surface model. Hydrol. Earth Syst. Sci. 22 (3), 1649–1663.
Silva, B.M., Silva, S.H.G., de Oliveira, G.C., Peters, P.H.C.R., dos Santos, W.J.R., Curi, N., 2014. Soil moisture assessed by digital mapping techniques and its field validation. Ciência e Agrotecnologia 38 (2), 140–148 http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-70542014000200005&lng=en&tlng=en.
Song, Y., Shen, Z., Wu, P., Viscarra Rossel, R.A., 2021. Wavelet geographically weighted regression for spectroscopic modelling of soil properties. Sci. Rep. 11, 17503 https://doi.org/10.1038/s41598-021-96772-z.
Song, Y., Wang, J., Ge, Y., Xu, C., 2020. An optimal parameters-based geographical detector model enhances geographic characteristics of explanatory variables for spatial heterogeneity analysis: Cases with different types of spatial data. GIScience & Remote Sensing 57 (5), 593–610.
Song, Y., Wright, G., Wu, P., Thatcher, D., McHugh, T., Li, Q., Li, S.J., Wang, X., 2018. Segment-based spatial analysis for assessing road infrastructure performance using monitoring observations and remote sensing data. Remote Sensing 10 (11), 1696.
Song, Y., Wu, P., 2021. An interactive detector for spatial associations. International Journal of Geographical Information Science 1–26.
Song, Y., Yang, H., Peng, J., Song, Y., Sun, Q., Li, Y., 2015. Estimating pm2.5 concentrations in xi'an city using a generalized additive model with multi-source monitoring data. PLoS One 10 (11), e0142149.
Tao, L., Ryu, D., Western, A., Boyd, D., 2021. A New Drought Index for Soil Moisture Monitoring Based on MPDI-NDVI Trapezoid Space Using MODIS Data. Remote Sensing 13 (1) https://www.mdpi.com/2072-4292/13/1/122.
Vereecken, H., Huisman, J.A., Pachepsky, Y., Montzka, C., van der Kruk, J., Bogena, H., Weihermüller, L., Herbst, M., Martinez, G., Vanderborght, J., 2014. On the spatio-temporal dynamics of soil moisture at the field scale. J. Hydrol. 516, 76–96. https://doi.org/10.1016/j.jhydrol.2013.11.061.
Verhoest, N.E., Lievens, H., Wagner, W., Álvarez-Mozos, J., Moran, M.S., Mattia, F., 2008. On the soil roughness parameterization problem in soil moisture retrieval of bare surfaces from synthetic aperture radar. Sensors 8 (7), 4213–4248.
Wang, J.-F., Li, X.-H., Christakos, G., Liao, Y.-L., Zhang, T., Gu, X., Zheng, X.-Y., 2010. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the heshun region, china. International Journal of Geographical Information Science 24 (1), 107–127.
Wang, J.-F., Zhang, T.-L., Fu, B.-J., 2016. A measure of spatial stratified heterogeneity. Ecol. Ind. 67, 250–256.
Wang, T., Franz, T.E., Li, R., You, J., Shulski, M.D., Ray, C., 2017. Evaluating climate and soil effects on regional soil moisture spatial variability using eof s. Water Resour. Res. 53 (5), 4022–4035.
Wang, Y., Yang, J., Chen, Y., Wang, A., De Maeyer, P., 2018. The Spatiotemporal Response of Soil Moisture to Precipitation and Temperature Changes in an Arid Region. China. Remote Sensing 10 (3) https://www.mdpi.com/2072-4292/10/3/468.
Wei, X., Huang, C., Wei, N., Zhao, H., He, Y., Wu, X., 2019. The impact of freeze–thaw cycles and soil moisture content at freezing on runoff and soil loss. Land Degradation & Development 30 (5), 515–523.
Western, A.W., Blöschl, G., Grayson, R.B., 1998. Geostatistical characterisation of soil moisture patterns in the tarrawarra catchment. J. Hydrol. 205 (1–2), 20–37.
Western, A.W., Zhou, S.-L., Grayson, R.B., McMahon, T.A., Blöschl, G., Wilson, D.J., 2004. Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes. J. Hydrol. 286 (1–4), 113–134.
Wilson, D.J., Western, A.W., Grayson, R.B., 2005. A terrain and data-based method for generating the spatial distribution of soil moisture. Adv. Water Resour. 28 (1), 43–54.
Wu, X., Liu, M., 2012. In-situ soil moisture sensing: measurement scheduling and estimation using compressive sensing. In: 2012 ACM/IEEE 11th International Conference on Information Processing in Sensor Networks (IPSN). IEEE, pp. 1–11.
Xu, X., Dunbar, R., Derksen, C., Colliander, A., Kim, Y., Kimball, J., 2018. Smap l3 radiometer global and northern hemisphere daily 36 km ease-grid freeze/thaw state. NASA National Snow and Ice datacenter, Boulder, Colorado, USA.
Yang, K., Qin, J., Zhao, L., Chen, Y., Tang, W., Han, M., Chen, Z., Lv, N., Ding, B., Wu, H., et al., 2013. A multiscale soil moisture and freeze–thaw monitoring network on the third pole. Bull. Am. Meteorol. Soc. 94 (12), 1907–1916.
Zacharias, S., Bogena, H., Samaniego, L., Mauder, M., Fuß, R., Pütz, T., Frenzel, M., Schwank, M., Baessler, C., Butterbach-Bahl, K., et al., 2011. A network of terrestrial environmental observatories in germany. Vadose zone journal 10 (3), 955–973.
Zappa, L., Forkel, M., Xaver, A., Dorigo, W., 2019. Deriving Field Scale Soil Moisture from Satellite Observations and Ground Measurements in a Hilly Agricultural Region. Remote Sensing 11 (22) https://www.mdpi.com/2072-4292/11/22/2596.
Zeng, J., Chen, K.-S., Bi, H., Chen, Q., 2016. A preliminary evaluation of the smap radiometer soil moisture product over united states and europe using ground-based measurements. IEEE Trans. Geosci. Remote Sens. 54 (8), 4929–4940.
Zhao, L., Yang, K., Qin, J., Chen, Y., Tang, W., Montzka, C., Wu, H., Lin, C., Han, M., Vereecken, H., 2013. Spatiotemporal analysis of soil moisture observations within a tibetan mesoscale area and its implication to regional soil moisture measurements. J. Hydrol. 482, 92–104.
Zhuo, L., Dai, Q., Zhao, B., Han, D., 2020. Soil moisture sensor network design for hydrological applications. Hydrol. Earth Syst. Sci. 24 (5), 2577–2591 https://hess.copernicus.org/articles/24/2577/2020/.
Zhuo, L., Dai, Q., Zhao, B., Han, D., 2020. Soil moisture sensor network design for hydrological applications. Hydrol. Earth Syst. Sci. 24 (5), 2577–2591.