A 1-km daily soil moisture dataset ~~of~~over China ~~based~~ using ~~on~~ in-situ measurement ~~using~~ and machine learning

Qingliang Li[1,2], Gaosong Shi[2], Wei Shangguan [1,*], Vahid Nourani[3], Jianduo Li[~~3~~4], Lu Li[1], Feini Huang[1], Ye Zhang[1], Chunyan Wang[2] Dagang Wang[~~4~~5], Jianxiu Qiu[~~4~~5], Xingjie Lu[1], Yongjiu Dai[1]

[1]Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Guangdong Province Key Laboratory for Climate Change and Natural Disaster Studies, School of Atmospheric Sciences, Sun Yat-sen University, Guangzhou 510275, China;
[2]College of Computer Science and Technology, Changchun Normal University, Changchun 130032, China;
[3]Center of Excellence in Hydroinformatics and Faculty of Civil Engineering, University of Tabriz, 29 Bahman Ave., Tabriz , Iran
[~~3~~4]State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing 10081, China
[~~4~~5]School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China

Correspondence to: Wei Shangguan (Email: shgwei@mail.sysu.edu.cn)

Abstract. High quality gridded soil moisture products are essential for many Earth system science applications, while the recent reanalysis and remote sensing soil moisture data are often available at coarse resolution and remote sensing data are only for the surface soil.~~and they are usually available from remote sensing or model simulations with coarse resolution.~~ Here, we present a 1 km resolution long-term dataset of soil moisture derived through machine learning trained ~~with~~ by the *in-situ* measurements of 1,789 stations over China, named as SMCI1.0. Random Forest ~~is~~ used as a robust machine learning approach to predict soil moisture using ERA5-L~~l~~and time series, leaf area index, land cover type, topography and soil properties as ~~covariate~~predictors. SMCI1.0 provides 10-layer soil moisture with 10 cm intervals up to 100 cm deep at daily resolution over the period 200~~1~~0-2020. Using *in-situ* soil moisture as the benchmark, two independent experiments were~~are~~ conducted to ~~investigate~~ evaluate the estimation accuracy of the SMCI1.0: year-to-year ~~experiment~~ (*ubRMSE* ranges from 0.041-0.052 and *R* ranges from 0.883-0.919) and station-to-station experiments (*ubRMSE* ranges from 0.045-0.051 and *R* ranges from 0.866-0.893). SMCI1.0 generally has advantages over other gridded soil moisture products, including ERA5-Land, SMAP-L4 and SoMo.ml. However, the high errors of soil moisture often located in North China Monsoon Region. Overall, the highly accurate estimations of both the year-to-year and station-to-station experiments ensure the applicability of SMCI1.0 to ~~studies~~ study on the spatial-temporal patterns. As SMCI1.0 is based on *in-situ* data, it can be useful complements of existing model-based and satellite-based soil moisture datasets for various hydrological, meteorological, and ecological analyses and ~~modeling~~modelling. The DOI link for the dataset is http://dx.doi.org/10.11888/Terre.tpdc.272415 (Shangguan et al., 2022).

1 Introduction

Soil moisture (SM) plays a key role in land-atmosphere interactions through its strong impacts on water and carbon cycle (Entekhabi et al. 1996; Seneviratne et al. 2010; Wagner et al. 2007). The status of SM is closely related to ~~the variation in~~ climate and weather conditions (Dirmeyer et al. 2006). The high-quality SM data with ~~large~~ fine spatial-temporal scale can be valued as indispensable ~~factors~~ tools for observing the extreme weather events, e.g., ~~monitoring of~~ droughts (e.g., Chawla et al. 2020; Mishra et al. 2017; Tijdeman and Menzel 2021) ~~and~~, floods (e.g., Kim et al. 2019; Norbiato et al. 2008; Parinussa et al. 2016)~~. Hence, high-quality SM can be acted as a vital variable in wide range of applications such as flood and drought prediction~~ and carbon cycle modelling (Sungmin and Orth 2020). Further, SM is also identified as an important component of the Essential Climate Variables by the Global Observing System for Climate (GCOS 2016). However, high-quality SM data acquisition is a challenging task due to the ~~high variability of SM in space and time~~ complicated spatiotempral variations of the SM (Li and Lin 2018; Ojha et al. 2014; Vereecken et al. 2014). ~~The~~Such spatiotempral variations ~~in~~of SM are usually affected by the inherent heterogeneity of soils, land cover, and weather (Brocca et al. 2007; Crow et al. 2012; Vereecken et al. 2014).

At present, the ~~way~~methods for ~~of~~ SM data ~~acquisition~~ estimation can be divided into five categories: *in-situ* ~~SM stations~~observations, satellite observations, offline land surface model simulations, Earth system model simulations, and reanalysis products. For *in-situ* SM observations, SM data ~~is~~are usually measured by the probe measurement method (Orth and Seneviratne 2014), ~~they have~~in which as direct observations this method usually leads to lower errors than satellite observations, land surface model simulations, Earth system model simulations and reanalysis products (Pan et al. 2019). Although large number of stations have distributed all over the world, there are still many regions with no *in-situ* SM observations due to financial constraints (Karthikeyan and Kumar 2016) and ~~they~~ field stations are too sparse to capture adequate spatial coverage (Gruber et al. 2016). For satellite observations, SM data ~~is~~are mainly retrieved by microwave radiometer (frequencies are less than 12 GHz) on satellite (Entekhabi et al. 2010; Fujii et al. 2009; Kerr and Coauthors 2010) which can provide the global SM data with uniformly distribution. But for the microwave radiometer measured SM data from the near-surface, only the top layer SM (typically ~5 cm) can be retrieved and the data gaps exist in regions with dense vegetation, and snow-covered or frozen soils. The SM data of the ~~in~~ offline land surface model and Earth system model simulations span~~s~~ multiple soil layers and have seamless spatial distribution (Gu et al. 2019), but they both have the uncertain and different forcing factors due to the spatial sub-grid heterogeneity of soil properties and vegetation~~,~~ and thus leading to large differences from *in-situ* SM observations~~.~~ (Dirmeyer et al. 2006; Kumar et al. 2009). ~~. For r~~Reanalysis products~~, they~~ can also provide SM data with ~~well~~good temporal variations by assimilating observations into land surface models or Earth system models (Chen et al. 2021). ~~Meanwhile, t~~They can also provide SM data in deeper soil depth than satellite observations~~.~~, ~~However,~~but reanalysis products ~~still have the differences~~usually lead to higher disagreement with *in-situ* SM observations when the assimilated meteorological variables (e.g., precipitation) are biased (Balsamo et al. 2015).

In brief, the characteristic strong-points and shortcomings are both coexisted in each type of SM product. Hence, we are eager to develop the high-quality SM product which comprehensively have high-resolution seamless spatial distribution, long time periods, and low errors from the above SM products.

Recently, machine-learning (ML) models have been successfully applied ~~in SM prediction~~for predicting (Li et al. 2021; Mohamed et al. 2021; Xu et al. 2010)~~ and~~or downscaling~~e modeling~~ (Chakrabarti et al. 2014; Srivastava et al. 2013; Wei et al. 2019; Mao et al. 2022) the SM values. They capture the complex nonlinear relationship between SM and all available predictors related to SM variation (e.g., meteorological variables, land-cover and soil data)~~ and further achieve~~with better accuracy~~ accurate results~~. ML models provide ~~an alternative opportunity for estimating~~capacity to estimate high-quality SM data based on *in-situ* SM ~~stations~~ measurements (Sungmin and Orth 2020) and further to improve the generated SM product~~, that give full play to the roles of the *in-situ* SM observations~~ with low errors~~, and other SM products with~~ and seamless spatial distribution ~~and~~for long time periods. ~~Such as, Zeng et al. (Zeng et al. 2019) applied the r~~Random forest (RF) as such a ML method was applied by Zeng et al. (2019)~~ model~~ to generate 0.5 km~~ spatial and~~ daily ~~temporal resolution of~~ SM ~~observations over~~ data for the period from 2010 to 2014 ~~in~~ over Oklahoma based on *in-situ* SM ~~stations~~ records and satellite observations. The low root means square error (ranging from 0.038 to 0.050 m^3/m^3 for year-to-year test and 0.044 to 0.057 for station-to-station test) obtained from experiments, ~~which demonstrated the usability of their~~ demonstrating the accuracy of the gridded SM data. Sungmin et al. (~~Sungmin and Orth~~ 2020) used the Long Short-term Memory (LSTM) model as a deep learning approach to estimate daily SM data ~~in the~~over whole world ~~with about~~at 27.75 km spatial ~~and daily temporal~~ resolution ~~over~~ for the period from 2000 to 2019, stating the superiority of their SM data over ERA5 dataset~~. They represented that their SM data outperformed the SM datasets of ERA5~~. It was necessary to note that the above two studies both emphasized that the applied *in-situ* SM observations ~~did~~ could not cover the whole tested regions, leading to relatively high uncertainty outside the training conditions. In other words, the more *in-situ* SM stations ~~existed~~ in the test~~ed~~ region, the higher ~~-~~quality gridded SM data ~~can be generated~~ by ML models. Additionally, Carranza et al. (~~Carranza et al.~~ 2020) used RF model to estimate root zone SM within a small catchment from 2016 to 2018, and demonstrated that ML model had slightly higher accuracy than a process-based model combined with data assimilation for data-poor regions. Karthikeyan et al. (~~Karthikeyan and Mishra~~ 2021) applied Extreme Gradient Boosting (XGBoost) to estimate daily SM data ~~in~~ over the United States with about 1 km ~~spatial and daily temporal~~ resolution ~~over~~ for the period from 31 March 2015 to 29 February 2019 (only 1,431 days) and the results showed that ~~they~~ the esimations can well capture temporal variations of SM (~~*ubRMSE* less than 0.04 m^3/m^3~~).

China is one of the largest countries in the world, ~~which located~~expanded from central ~~and~~ to eastern Asia. The climate types are complex and diverse, which spans wet, semi humid, semi dry and dry climate types from southeast to northwest, the northward extent and intensity of summer monsoon often cause significant changes in precipitation and arid-humid climate (Cong et al. 2013). ~~As we know,~~Since SM and precipitation can interact with each other (Li et al. 2020), therefore in situ data based estimation of SM is a challenging task due to such heterogeneity and complex spatiotemporal variabilities. ~~which~~

~~also represents that the variability of China SM in space and time are complex and further takes serious challenges for estimating China SM data based on *in-situ* SM stations~~.

Previous studies have already ~~produced many~~provided several ~~SM~~ gridded SM products covering China or the world, but mainly based on remote sensing data and only for the surface layer (e.g., Chen et al., 2021, Meng et al., 2021, Song et al., 2022, Wang et al., 2021 and Zhang et al., 2021). However, there is still a big gap in technical literature about daily SM data with high quality (high-resolution seamless spatial distribution, long time periods, and low errors) at multiple layers based on *in-situ* measurements ~~do not exist~~ for China ~~yet~~. Although Sungmin et al. (~~Sungmin and Orth~~ 2020) generated the global SM data by ML model which includes the China region, only ~~less than~~data from about 20 *in-situ* SM stations have been applied for SM modelling for the whole China ~~in China were applied, which was hardly ensure the quality of China SM product~~. In addition, ~~this product's resolution~~the resolution of this product is 0.25 degree, which limits its use in regional applications ~~requiring~~ when high resolution SM are required.

To fill this research gap, in this study, we aimed ~~to generate~~at generating high quality gridded SM data ~~in~~ over China ~~with using~~ *in-situ* measurements ~~based on~~and RF model (Fig.1). The ~~covariate~~predictors were consisted of static data and time series variables, including ERA5-Land (the land component of the fifth generation of European Reanalysis, Muñoz Sabater, 2019; Muñoz Sabater, 2021~~Balsamo et al. 2015~~), USGS (United States Geological Survey) land cover type (Loveland et al. 2000), USGS DEM (Digital Elevation Model, Balenović et al. 2016), reprocessed MODIS LAI (Moderate-resolution Imaging Spectroradiometer Leaf Area Index, Yuan et al. 2011) and CSDL (China Soil Dataset for Land surface ~~modeling~~modelling, Shangguan et al. 2013). The *in-situ* SM observations from 1,648~~1,789~~ stations ~~after quality control procedures~~ were ~~acted~~ employed as ~~our~~ the SM modelling target ~~variables~~ after quality control procedures.~~, which were obtained from China Meteorological Administration (CMA).~~

~~Our~~ The new China gridded SM product (named SMCI1.0, Soil Moisture of China by *in-situ* data, version 1.0) provides SM data at ten layers, which include soil depth from 10cm to 100cm with an interval of 10 cm. Meanwhile, SMCI1.0 has ~1km (30 seconds) spatial resolution and daily temporal resolution over the period from 1 January 2010 to 31 December 2020. For the SMCI1.0 product, we mainly considered to answer the ~~four~~ following research questions ~~as follows~~:

(1) ~~How~~ What is the sensitivity of the in-situ SM data to ~~are *in-situ* SM and~~ all the ~~covariate~~predictors ~~related~~, including meteorological data (air temperature, precipitation, total evaporation, potential evaporation), soil data (SM and soil temperature at different soil layers, and static soil properties), leaf area index and land cover type.

(2) Can the RF model successful generate high quality gridded SM (high-resolution seamless spatial distribution, long time periods, and low errors) at multiple layers ~~in~~ over China based on *in-situ* SM observations?

(3) How does the RF model perform~~s~~ for spatiotemporal estimation of SM ~~for the space and time extrapolation experiment, in other words, can the RF model generate the SM data with low errors which take *in-situ* SM observations as the reference~~ under year-to-year and station-to-station ~~estimating~~scebarios?

(4) What are the conditions ~~can~~ in which SMCI1.0 SM data ~~have~~ may lead to lower errors ~~or~~ and higher errors against adjusted *in-situ* SM observations?

For the above issues, we make four contributions for generating and validating multi-layer gridded SM data over China. First, we record and make detailed analysis of the correlations between *in-situ* SM and all ~~covariate~~predictors. Then, we apply the RF to model the complex relationship between ~~covariate~~predictors and *in-situ* SM observations, and further validate ~~the~~ using year-to-year and station-to-station ~~estimating~~experiments. Finally, we intuitively display and analysis the quality of SMCI1.0 ~~with~~ at different conditions, which can help the researchers to ~~and it is expected to help researchers~~ improve the China gridded SM intentionally and strategically. ~~The schema of this work is listed below.~~ Section 2 describes the *in-situ* SM data, predicting data~~, data served as covariates~~, RF model and its application in SM estimating. Section 3 gives the validation results, experimental results, a sampled map on a day and relative importance of ~~covariate~~predictors. Sections 4 and 6 present the discussion conclusions, respectively.

2.Materials and Methods

2.1 *in-situ* SM observations

Target SM data for RF model was constructed from the CMA SM observations. The ~~observations~~ dataset contains hourly data from 1,789 stations over China ~~(18-N, 73-W) and have hourly temporal resolution over~~ for ~~the period from~~ 1 January 2010 to 31 December 2020. The spatial distribution of observations is shown in Fig. S1(a). ~~For our *in-situ* SM observations, two aspects deserve to be noted: one aspect is~~It should be noted that data from such a ~~the~~ large number of in situ stations ~~(i.e., 1,789), which~~ can help ML models to capture the complex nonlinear relationship between SM and ~~covariate~~predictors over various training conditions and thus to generate high quality ~~China~~ gridded SM data. ~~The other aspect is the bias and standard deviation correction of *in-situ* SM, which is vital for our study to allow the ML model to achieve the high-quality SM product. We applied the same correcting method with that of Sungmin et al. (2020), who adjusted the raw *in-situ* SM observations to match means and standard deviation of the ERA5-Land gridded SM data at the corresponding time periods, grid cells and layers.~~

The automated quality control of *in-situ* SM observations was performed before training the RF model. We first removed the null values over the long period (~~10~~ 10-days time step) and outlier/unreasonable SM values. ~~In checking~~To check the unreasonable SM values, four plausibility checks were performed, such as checking geophysical consistency using precipitation and soil temperature, spike detection, break detection and constant values detection. The details could be found in the Global Automated Quality Control method (Dorigo et al. 2013). Finally, the removed values were replaced by the linear interpolation method according to the remaining SM values at the same time period from five days ahead ~~and~~ to five days later. To facilitate generating 1km gridded SM data at multiple layers by the RF model, the CMA SM observations were processed to daily and the observations were averaged if there are more than one stations within a grid at 1km resolution. We simply averaged all the available observations in each day ~~at each *in-situ* SM measurement station for daily resolution~~ and all ~~the *in-situ* SM measurement~~ stations if there are more than one stations within each grid ~~for~~ with 1km resolution. In this way, we got 1, 648~~-1,789~~ spatial points (or grids) of observations. The description of *in-situ* SM could be found in Supplementary Material (Text S1 and Fig. S1). ~~The details for the target *in-situ* SM are represented in Fig. 2. Fig. 2(a) shows that stations are dense in the east part of China, but sparse in the west part. Fig. 2(b) represents that the sample size varies with soil depth, and large numbers of missing values exist at 70 and 90 cm soil depths. From Fig. 2 (c), we could see that the values of the *in-situ* SM at all soil depths were mainly concentrated in the range from 0.2 to 0.4 m^3/ m^3. Fig. 2(d) denotes that the data number in low standard deviation (0~0.05 m^3/ m^3) is smaller than that in high standard deviation (0.05~0.07 m^3/ m^3) from at 10 to 40 cm soil depths. But the opposite conclusion can be drawn from 50 to 100 cm soil depths (larger data number in low standard deviation is than that in high standard deviation). Meanwhile, Fig. 2(d) also hints that the standard deviation of SM at deeper soil depth (except that at 100 cm soil depth) is lower than that at upper soil depth. Decreasing standard deviation with increased soil depth denoted that *in-situ* SM is more stable in deep soil depth, which is consistent with the previous studies (Gao and Shao 2012; Wang et al. 2013). From Fig. 2 (e), we could see that the stations have 8 climate types, most observations belong to temperate climate with dry winter (Cw), temperate climate, fully humid (Cf) and snow climate with dry winter (Dw), and the data with tropical monsoon climate (Am) and snow climate, fully humid (Df) are sparse, which occupy only small parts of China.~~

After the above data processing~~generating daily SM based on CMA SM observations for each 1km grid where there is one or more *in-situ* stations~~, ~~we started to perform~~ the correction of deviation and variance ~~for~~of *in-situ* SM was performed, which can help the ML model to achieve the high-quality SM product. ~~i~~*In-situ* SM data ~~was~~ have been obtained by various sensor types~~, which had~~ with different calibration processes. Hence, to overcome the ~~artifacts~~ artefacts during the RF model training, we adjusted the observations to match means and standard deviation of the ERA5-Land SM at the corresponding time periods ~~and~~, grid cells and layers using the method proposed by ~~(~~Sungmin and Orth (2020). In this method, we first obtained a weight by dividing the standard deviations of the in-situ SM at each station by that of ERA5-Land SM at the corresponding grid, and then multiplied the original *in-situ* SM by this weight. After that, we computed the difference between the average value of the *in-situ* SM at each station and the ERA5-Land SM at the corresponding grid, and subtract the *in-situ* SM by the computed difference. This method made the target *in-situ* SM resemble the mean and standard deviation of ERA5-Land SM, and kept daily temporal variations which follow the original *in-situ* SM time series. As the soil depth of each soil layer of ERA5-Land SM was inconsistent with that of *in-situ* SM, we mapped the soil layer of ERA5-Land SM to the corresponding soil layers of *in-situ* SM. Hence, the *in-situ* SM data from 10 cm to 30 cm were adjusted based on the gridded SM at layer2 from ERA5-Land dataset (7-28 cm), and the *in-situ* SM data from 30 cm to 100 cm were adjusted based on the gridded SM at layer3 from ERA5-Land dataset (28-100 cm).

2.2 Datasets as ~~covariate~~predictors

Table 1 shows the used ~~datasets uses covariate~~predictors ~~used~~ for RF ~~modeling~~modelling. Most ~~covariate~~predictors were collected from the ERA5-Land reanalysis dataset, which is an enhanced version of ERA5 land component, forced by meteorological fields from ERA5.~~which was produced by the land component of European Centre for Medium-Range Weather Forecasts (ECMWF).~~ The reasons for selecting the ERA5-Land dataset as preference were as follows: (1) it is generated under a single simulation of a land surface model using ERA5 reanalysis as the forcing data, but with a series of improvements ~~making~~ which make it more accurate for all types of land applications (Muñoz-Sabater et al., 2021)~~(Albergel et al. 2018)~~; (2) ERA5-Land is currently updated with 2-3 months latency~~there are only several months latency for obtaining ERA5-Land datasets~~, which ~~allowed~~ allows us to update SMCI1.0 in time; (3) ERA5-Land~~the data~~ is long-term (since ~~1981~~1950) data and with seamless spatial distribution and multilayers, which ~~helps us~~makes it possible to generate high quality SMCI1.0 ~~easily~~. Compared with satellite observations, we can avoid the spatial-temporal gaps and limited time periods covered by using ERA5-Land reanalysis (Sungmin and Orth 2020). The static data of ~~covariate~~predictors were collected from USGS land cover type (Loveland et al. 2000) and DEM (Balenović et al. 2016), reprocessed MODIS LAI Version 6 for land surface and climate modelling (Yuan et al., 2011) and the China Soil Dataset for Land Surface Modeling (CSDL, Shangguan et al., 2013), including sand, silt and clay content, rock fragment, and bulk density. The reprocessed MODIS LAI Version 6 was improved by a two-step integrated method ~~that had the advantage of~~with continuity and consistency in space and time ~~series~~ domains (Yuan et al., 2011). It was worth noting that the temporal resolution of the reprocessed MODIS LAI Version 6 was 8 days, and the daily LAI between the 8 days was computed by linear interpolation of the nearest two LAI values at 8-day time step. CSDL was ~~developed for use in the land surface modeling. The spatial distribution of soil type, rock fragment, and bulk density was~~ derived by the polygon linkage method, ~~which were well represented and~~whose results are consistent with common knowledge of Chinese soil scientists (Shangguan et al., 2013). All predictors were processed to the same 1km by 1km grid system. ERA5-Land data with 9 km resolution were resampled into 1 km by the nearest neighbor method and MODIS LAI with 500 m resolution were aggregated into 1 km by averaging.

2.3 Random Forest ~~regression~~

Random Forest (RF) is an ensemble machine learning approach, which apply the decision trees and bagging methods for the classification and regression problem (Breiman 2001). The simple decision trees model ~~partitioned~~ partitions the variable space and further ~~grouped~~ groups dataset recursively based on similar instances. For the candidate variables from a set of ~~covariate~~predictors, a split ~~was~~ is determined by the values of ~~interesting~~ desired variable that is evolved into a tree structure with multiple parent and child nodes. Meanwhile, the response variance for decision regression trees ~~was~~ is applied as the criterion to~~for~~ maximize~~s~~ the purity of each node (the response variance ~~was~~ is applied to measure node purity) and further to find the optimal split. RF generates~~d~~ diverse decision trees to avoid overfitting through bagging method, which constructs~~ed~~ multiple training sub-dataset by resampling with replacement of the original dataset. For each training sub-dataset, a decision tree ~~was~~ is growing until the ~~selected~~ pre-assumed criterion ~~was~~ is reached (e.g., the value for the minimum node size). ~~After~~ When all ~~the~~ decision trees ~~were~~ are generated, the average ~~was taken from~~of all the estimations from each decision tree is computed.

The importance of the ~~covariate~~predictors obtained by the RF model ~~was~~ is also worth noting, which ~~computed~~ can be explored by a permutation ~~scheme~~method. In the permutation method, ~~the~~ different SM ~~was~~ are estimated by permuting all the ~~covariate~~predictors. Hence, the importance of ~~covariate~~predictors ~~could~~ can be ~~obtained~~ detected by comparing thei~~r~~

accuracy of SM estimation. Such as, if one ~~covariate~~predictor ~~was vital~~is dominant to estimate target SM, the estimated SM values accuracy ~~was~~ is expected to be decreased ~~for~~using ~~estimation by~~ the ~~remaining~~ other non-permuted ~~covariate~~predictors ~~without the covariate~~.

2.4 The application ~~for~~ of Random Forest model

In ~~our~~ this study, we first ~~selected~~ determined the optimal values of hyper-parameters in RF model based on the 10-fold cross-validation method. After ~~selecting the optimal~~calibration of the hyper-parameters, two independent experiments ~~are~~ were conducted to investigate the estimation accuracy of the developed SMCI1.0 ~~at~~ spatial-temporal ~~scale~~ data (year-to-year and ~~station~~ station-~~to~~ to-station experiments). In the year-to-year ~~estimating~~experiments, the data from 2010 to 2017 ~~years~~ in each station ~~was~~ were reserved for training set, and to evaluate the ~~estimation~~ accuracy of SMCI1.0 at temporal scale, we compared the generated SM by RF model ~~at each soil depth~~ with the ~~corresponding~~ *in-situ* SM data from 2018 to 2020 ~~years~~. In the station-to-station ~~estimating~~experiments, the randomly selected data from 2/3 of the stations ~~with randomly selection~~ from 2010 to 2020 were~~was~~ applied for training ~~the RF model~~, and the ~~remained 1/3~~data of the rest stations were used to evaluate the ~~estimation~~ accuracy of SMCI1.0 at spatial scale. Finally, the SMCI1.0 product was generated by RF model at 1km resolution~~, which was built~~ based on the *in-situ* SM and ~~the combined~~ ~~covariate~~predictors (shown in Table 1) from all stations and all years. In addition to the ~~1~~ 1-km resolution, we also produced a version of ~~9~~ 9-km resolution by aggregating the higher resolution ~~covariate~~predictors for the convenience of applications ~~which need only~~when coarser SM data are needed in broad scale studies. In addition to the period of 2010-2020 when in situ SM data are available, we also produced the gridded SM for the period of 2000-2009 when in situ SM data are unavailable, assuming that the relationship between SM and predictors remains the same in the last two decades. It is proper to deem that the data quality during 2000-2009 is poorer than that of 2010-2020. ~~SMCI1.0 can be accessed at 错误!超链接引用无效。~~

The number of randomly selected ~~candidate~~ variables from all the ~~covariate~~predictors (*max_features*) and the value ~~for~~ for the minimum node size (*min_samples_leaf*) ~~in RF model were~~ are the vital hyper-parameters for RF model which can affect the modelling performance. ~~The values of *max_features* and *min_samples_leaf* directly determined how the RF model grown.~~ Other hyper-parameters, such as number of trees (*n_estimators*), were no~~t tuned but simply~~ determined based on RF's own training. ~~The hyper-parameters *max_features* affected the split SM values and *min_samples_leaf* was acted as the criterion for stopping the decision tree growing.~~ Meanwhile, to prevent over-fitting problem, we applied the 10-fold cross-validation method to tune the values of *max_features* and *min_samples_leaf*~~, and they were selected from~~ in the range [1,25] with a single interval and [5,30] with 5 intervals via the gridded direct search method. ~~hyper-parameters method for preventing RF model over-fitting, which randomly divided the whole dataset into *k*-fold and a 10th of the sub-datasets was used as validation sample while the other sub-datasets were applied for training RF model. The root means square error (*RMSE*) was assessed for evaluating model accuracy by the 10-fold cross-validation method.~~ The accuracy of RF models with all hyper-parameters calibrated by the direct search method~~based on grid hyper-parameters method~~ at 10 cm soil depth were shown in Table ~~21~~S1. ~~We could see~~It can be seen that the root mean square error (*RMSE*)~~RMSE~~ obtained based on all the hyper-parameters ranged from 0.601 to 0.637 and the best accuracy (*RMSE*=0.601) ~~can~~ could be achieved when *max_features* and *min_samples_leaf* set to be 1 and 20, respectively, and they are used to the rest modelling of this study.~~. The optimal hyper-parameters (*max_features*=1 and *min_samples_leaf*=20) in RF model were used for further research.~~ The modelling performance and quality of SMCI1.0 product ~~was~~ were evaluated in terms of *ubRMSE*, *MAE* (Mean Absolut Error), *R* (correlation coefficient), R^2 (explained variation) and *Bias*, respectively. *ubRMSE* and *MAE* were applied to test the ability to estimate volatility and fluctuation amplitude, respectively. *R* denotes fluctuation pattern and R^2 represents the percentage of variance explained by the RF model. *Bias* was used to observe if the estimations were overestimated or underestimated. ~~The five~~These metrics were computed as follows:

$$ubRMSE = \sqrt{\frac{\sum_{i=1}^{N}[(x_i-\bar{X})-(y_i-\bar{Y})]^2}{N}}, \tag{1}$$

$$MAE = \frac{\sum_{i=1}^{N}|x_i-y_i|}{N}, \tag{2}$$

$$Bias = x_i - y_i, \tag{3}$$

$$R = \frac{\sum_{i=1}^{N}(x_i-\bar{X})(y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{Y})^2}}, \tag{4}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i-x_i)^2}{N\sum_{i=1}^{N}(y_i-\bar{Y})^2}, \tag{5}$$

where y_i and x_i denoted the *i*-th *in-situ* SM and gridded SM for all the stations and periods, respectively. $\bar{Y}$ and $\bar{X}$ represented the mean values of the *in-situ* SM and gridded SM, respectively.

3.Results

3.1 Validation of Random Forest based SM modelling ~~validation~~

To evaluate and validate the performance of RF model for generating SMCI1.0, we mainly discussed the ~~modeling~~modelling ability by year-to-year and station-to-station experiments, which could ensure that SMCI1.0 product has low errors in both temporal and spatial scales against *in-situ* SM records. Meanwhile, we also compared the results with the state-of-the-art global gridded datasets such as ERA5-Land, SMAP-L4 and SoMo.ml ~~datasets~~.

The scatter plot ~~between~~ of the mean values of SMCI1.0 and ~~that of~~ *in-situ* SM data for ~~at~~ each station, the frequency distributions of all SM values in SMCI1.0 and ~~that~~ in *in-situ* measurements, and the violin-plot for the distribution of daily SM from stations for each climate type ~~were~~ are ~~re~~presented for the year-to-year experiment in Fig. 2 (from 10 to 30 cm soil depths) and Fig. S2 (from 40 to 100 cm soil depths). As shown in Fig. 2 (a), we can conclude that there ~~was~~ is generally a good agreement between the mean of SMCI1.0 and that of *in-situ* SM at each station (the correlation ranges from 0.867 to 0.908), which demonstrate~~d~~ that the RF model can well capture spatial variations in *in-situ* SM. The RF model showed somewhat better results in deeper soil depths, such as the RF model at 30 cm soil depth had better performance than that at 10 and 20 cm soil depths ~~in~~ as shown by Fig. 2 (a), which was consistent with the previous studies (e.g., Sungmin and Orth 2020). ~~And the~~The ~~different~~ worst results ~~was~~ were achieved by the RF model at 70 cm and 90 cm soil depths ~~in~~ as shown by Fig. S2 (a)~~, where the performance was the worst in all the soil depths~~ (*ubRMSE*=0.053, *MAE*=0.038, *R*=0.867, R^2=0.731 at 70 cm soil depth; *ubRMSE*=0.052, *MAE*=0.036, *R*=0.883, R^2=0.759 at 90 cm soil depth). Meanwhile the best result was achieved by the RF model at 30 cm soil depth (*ubRMSE*=0.043, *MAE*=0.033, *R*=0.908, R^2=0.824~~ at 30 cm soil depth~~). ~~The reason may be that RF model is difficult to estimate accurate SM for only a few *in-situ* SM stations. From Fig. S1 (b), we can see that the total numbers of data at 70 cm and 90 cm soil depths is relatively small. In other words, more diversity of data was expected to help RF model 'learn' complete relationship between covariate~~predictor~~s and *in-situ* SM and further generated SMCI1.0 with low errors in China. Meanwhile, it also showed the superior quality for our SMCI1.0 product, because the larger numbers of *in-situ* SM data in China were applied for estimating seamless SM than that by the previous studies (Sungmin et al. 2020). In~~ As shown by Fig. 2 (b), although the SMCI1.0 ~~had~~ yielded ~~smaller~~ less variability in the values range from 0 to 0.18, 0.38 to 0.43, and 0.46 to 0.6 and ~~larger~~ higher variability in other ~~value~~ ranges, as a whole, ~~SM in~~ SMCI1.0 data generally agree~~d~~ well with *in-situ* SM values. The same conclusion can be drawn from 40 to 100 cm soil depths in Fig. S2 (b). The SMCI1.0 data were further evaluated for each climate type in Fig. 2 (c) and Fig. S2 (c). With regard to the violin-plot, RF model ~~can~~ could estimate consistent results with *in-situ* SM. However, the inconsistent SM was estimated in Tropical Monsoon Climate (Am) and Desert Climate (Bw). The reason could ~~also~~ be attributed to only few *in-situ* SM data in these climatic regions, ~~which~~ as ~~re~~presented in Fig S1 (e). Finally, we concluded that RF model can reproduce the temporal variation in *in-situ* SM data accurately at unseen period~~ accurately. Meanwhile, we also advocated that more diverse training data over various regions was needed for capturing the complex relationship between covariates and SM, and further improving the quality of high resolutions SM product.~~

It is clear f~~F~~rom Fig. 3 and Fig. S3~~, we could see~~ that although the results of the station-to-station experiment were inferior to ~~that~~ those of the year-to-year estimating, RF model ~~can~~ could also perform well in estimating seamless SM ~~in~~ over China at unseen locations. Additionally, similar to the year-to-year experiment, RF model performed ~~the best~~better at 30 cm soil depth than ~~that~~ those at other soil depths in the station-to-station experiment.

Finally, we ~~also~~ compared SMCI1.0 product with other gridded datasets (ERA5-Land, SoMo.ml and SMAP-L4) according to the median *ubRMSE*, *R*, *Bias* and *MAE*. ~~From~~ According to Fig. 4 and Fig. S4, SMCI1.0 product ~~had~~ provides the lowest median *ubRMSE* and *MAE* values ~~from~~for 10 cm to 100cm soil depths. ~~Regarding~~ considering the median *Bias* between gridded SM and *in-situ* SM observations, SMCI1.0 product ~~had~~ shows almost similar ~~quality~~ accuracy with ERA5-Land datasets for all ~~the soil~~ depths, but ~~had~~ higher ~~quality~~ accuracy than SoMo.ml and SMAP-L4 datasets. It was worth noting that the SMAP-L4 dataset ~~had~~ has the widest spread of errors and tended to underestimate *in-situ* measurements, which lead~~ed~~ to higher median *ubRMSE* and *MAE* values. Regarding the median *R* between gridded SM and *in-situ* SM observations, SMCI1.0 product ~~had~~ has slightly higher quality than SoMo.ml dataset for 10cm, 20cm, 80cm and 100cm soil depths and obvious advantages ~~than~~ over ERA5-Land and SMAP-L4 datasets for all ~~the soil~~ depths, while it had lower quality than SoMo.ml dataset for other soil depths. Considering all the above metrics, SMCI1.0 product ~~were~~ provides more robust data than ~~the~~ some other commonly used gridded datasets. ~~Interestingly, it was inconsistent for the results of *R*, *ubRMSE*, and *MAE* in Fig. 2 and Fig. 4, which had the same phenomenon with the previous studies (Sungmin and Orth 2020) (represented in their Fig. 4 and Fig. 5). For example, SMCI1.0 product had the *ubRMSE*, *MAE* and *R* being 0.046, 0.035 and 0.889 at 10 cm soil depth in Fig. 2. However, in Fig. 4, the box-plot represented the lowest *ubRMSE*, *MAE* and highest *R* of SMCI1.0 product were nearly 0.03, 0.02, and 0.7, respectively. The reason may be that the same metrics were calculated in different ways, the one in Fig. 2 was to count the results of all stations and temporal period, and the one in Fig. 4 was to count the results of only temporal period at one station.~~

Overall, the RF model ~~can~~ could be able to successfully generate the SM data with low errors taking *in-situ* SM observations as the reference at unseen periods and locations. According to the comparison analysis, the SMCI1.0 product outperforms ~~the existing~~ some other SM products ~~(~~including ERA5-Land, SoMo.ml and SMAP-L4~~) in the sense of statistic metrics~~.

3.2 The spatial and temporal evaluation of the SMCI1.0

Overall performance of the proposed modelling and accuracy of SMCI1.0 dataset were evaluated in section 3.1, but nothing presented there about variability and trend of this dataset at different temporal and spatial scales.~~As the section 3.1 evaluated the overall performance of estimated SM at the macro level, the variability and trends of the SMCI1.0 in temporal and spatial scale cannot be reflected.~~ Hence, to ~~take the evaluation of~~evaluate the temporal variation of the SMCI1.0 data ~~in temporal scale~~, we randomly selected stations from different climate ~~regional~~ regions for evaluating the ~~SM temporal~~ dynamics of the SM data in SMCI1.0, ERA5-Land, SMAP-L4, SoMo.ml and *in-situ* SM from 10 cm to 20 cm soil depths. ~~And~~ On the other hand, for the spatial scale, we represented the estimation performance for each *in-situ* SM station in terms of *ubRMSE*, *R*, and *bias*~~, respectively~~. Noticeably, ~~in order to evaluate each station as much as possible, we apply~~ year-to-year experiment was conducted to evaluate each station as much as possible ~~in this testing~~.

Fig. 5 compare~~d~~s the ~~SM~~ temporal dynamics of the SM data from SMCI1.0, ERA5-Land, SMAP-L4, SoMo.ml, and *in-situ* ~~SM~~ datasets at 10 cm soil depth along with local precipitation. We could see that although the SMCI1.0 product ~~had~~ shows large deviation compared ~~with~~ to the *in-situ* SM in snow climate, fully humid zone (Df-51431: E, N), it was almost consistent with *in-situ* SM in other regions. It ~~was~~ is necessary to note that the SM values in Desert Climate region (Bw-W1063: E, N) ~~had~~ show higher variability but with low precipitation from 231th to 325th days, the SMCI1.0 product could still adequately capture their relationship (represented in the light blue rectangle). Overall, and similar to in-suit data, SMCI1.0 data reasonably follow the consistency with climate condition as SM is increased and decreased in wet and dry conditions, respectively.~~the SMCI1.0 could follow the reasonable patterns which *in-situ* SM increased with wet condition and decreased with dry conditions. During the rainfall near 91th day across the Tropical Monsoon Climate zone (Am) and near 1st day across the Snow climate with dry winter zone (Dw), the *in-situ* SM did not increase with high precipitation, but the SMCI1.0 product could capture the increase in SM (denoted in the light blue rectangle). The reason may be that the applied covariates had bias with *in-situ* measurement and further affected estimation by RF model. Meanwhile, we also~~

~~found the RF model could overcome much bias in dry conditions, except for that from 196th to 305th days in the snow climate, fully humid zone (shown in the light red rectangle). In the case of 30 cm soil depth (represented in Fig. S4), we could see an agreement between several peak events, it could be attributed to the soil texture homogeneity in the 10 and 30 cm soil depths. Almost all climatic regions had lower dynamic ranges at 30 cm soil depth than that at 10 cm, this may be attributed to the persistent behavior of SM at 30 cm soil depth. For the evaluations of SM temporal dynamics from 10 to 30 cm, we can see that SMCI1.0 can broadly capture the temporal characteristic of *in-situ* SM and further demonstrated the high quality of SMCI1.0 product.~~

Fig. 6 represent~~sed~~s the *in-situ* testing performance according to the ~~fit statistics (~~*ubRMSE*, *R*, *Bias*, and *MAE*~~)~~ values. We could see that the SMCI1.0 product ~~had~~ led to relatively low *ubRMSE*, *Bias*, and *MAE* over most regions. ~~In combination~~Additionally, ~~with~~ Fig. 7~~, we also found~~ shows that the low errors of SMCI1.0 product ~~were~~ often appeared in the arid regions, which was consistent with the previous study (Zhang et al. 2019). However, the higher *ubRMSE*, *MAE* and lower *R* values could be seen in North China Monsoon Region. The North China Monsoon Region has typical temperate monsoon climate characteristics, where the annual temperature is high and the rainy season is concentrated. The SM variations in the North China Monsoon Region were complex, which may present great challenges for estimating SM by RF model. Except North China Monsoon Region, SMCI1.0 data mostly led to the R values larger than 0.5.~~Despite SMCI1.0 product had lower *R* in North China Monsoon Region than that in other climatic regions, the *R* values were mostly larger than 0.5 (within the acceptable limit). This highlighted the robustness of SMCI1.0 product.~~ According to the *Bias* in Fig. 6~~7~~, we could see that SMCI1.0 product tends to be underestimated in the northeast and southwest China, and be overestimated in the east China, which had the similar trend with ERA5-Land dataset ~~and~~, ~~we could also draw the similar conclusions for the~~which can also be confirmed by the box-plot of Bias in Fig. 5. SMCI1.0 product led to lower errors than SoMo.ml in estimating *in-situ* SM. Meanwhile, ~~it had the opposite estimations with SoMo.ml dataset in north China and Sichuan province (~~SMCI1.0 product are often underestimated in north China ~~and~~ but overestimated in Sichuan province (97°21'E-108°12'E, 26°03'N-34°19'N), ~~but~~ contrarily to the SoMo.ml dataset ~~was the opposite),~~ . ~~but SMCI1.0 product had lower errors in estimating *in-situ* SM.~~ According to the *R* values in Fig. 6~~7~~, SMCI1.0 product ~~had~~ led to ~~the~~ similar results with SoMo.ml dataset, and performed better than ERA5-Land and SMAP-L4 datasets, which could also be represented by the box-plot of *R* in Fig. 5. ~~In the case of 30 cm soil depth in Fig. S5, the SMCI1.0 product had higher accuracy than that at 10 cm soil depth, especially in terms of *ubRMSE* and *MAE* metrics. The reason may be the background aridity leaded to low variability of SM in the deeper layers (Karthikeyan and Mishra 2021). The RF model can capture the variation in SM easier.~~

3.3 Spatial patterns of SMCI1.0

To describe the general spatial patterns of SMCI1.0 over the China, as an example,~~we presented~~ the 1km SM maps are presented ~~at 1km spatial resolution~~ for 1st January 2016 by Fig. 7~~.~~, ~~From Fig. 7, we could see~~ which shows that the spatial contiguity of SM patterns for SMCI1.0 ~~was~~ could be captured well, and most high-resolution details of SM patterns in all the climatic region for SMCI1.0 had more detailed “expression” than that for other SM products. Meanwhile, the spatial pattern of SMCI1.0 ~~is~~ was more consistent with those of high-resolution ~~covariate~~predictors such as DEM and LAI in some regions, which ~~also denoted~~indicated that the SMCI1.0 could better reflect the detailed spatial distribution of SM. Southeast China is the tropical monsoon climate zone, where the rainy season was concentrated (~~re~~presented in Fig. 5). Hence, these regions are predominantly wet in the SM maps. Northwest China is the Desert Climate region, ~~which had~~ with few~~not any~~ rainfall and ~~further lead to the~~ dry conditions (also represented in Fig. 5). Qinghai province (89°35'E-103°04'E, 31°09'N-39°19'N) belongs to the tundra climate zone, where some soils are wet and others ~~soils~~ are dry. This is probably due to the complicated topography of Qinghai Province that some regions with woody plants can intercept rainfall, which may decrease the overall water input into the soil (Zwieback et al. 2019), and other regions with vegetation can decreases soil temperature and evaporation from the soil surface by shading, ~~which avoid~~preventing the loss of soil moisture (Kemppinen et al. 2021).

~~3.4 Relative importance of covariates~~

~~The relative importance of covariates at the ten soil depths was shown in Fig. 9 and Fig. S6. Bars represented the variability of relative importance across the covariates. As represented in Fig. 9, the ERA5-Land SM was the most important to estimate *in-situ* SM from 10 to 100 cm soil depths. In addition to ERA5-Land SM covariates, evapotranspiration, DEM, clay, reprocessed MODIS LAI (Version 6), porosity, LAI low vegetation, air temperature, LAI high vegetation and silt were followed. The importance of other covariates was less than 0.01, which were not detailed discussed in this study. As we know, had strong correlation with SM dynamic under water-limited conditions (Albertsona and Kiely 2001). So, evapotranspiration had greatly associated with SM in the regression model. Clay, porosity, rock fragment, silt and sand were properties in the soil. Bissonnais et al. (Bissonnais et al. 1995) tested SM for 31 soil types with different soil properties over Illinois region and denoted that the available SM varied by each soil group. They could help RF model identify variation in SM through different soil properties. LAI was a vital parameter in the land surface and controlled many complex processes in relation to vegetation, which determined evapotranspiration and further had impact on water balance (Chen et al. 2015). It is worth note that reprocessed MODIS LAI (Version 6) (Yuan et al. 2015) had larger impact on SM estimation than the LAI of reanalysis products. The reason may be that it had better quality than the LAI of reanalysis products. Air temperature and SM were closely related, such as the climate shifts from the hot to the cold, SM decreased for all land covers (Feng and Liu 2015). However, air temperature had significant effect in the RF model for upper soil layers (at 10 cm and 20 cm soil depths) while it began to weaken in the deeper soil (represented in Fig. S6), which was consistent with the previous studies (Hu and Zheng 2003). Interestingly, as widely known, the land cover type is highly related to the variation in SM. However, it had relative low importance (less than 0.01) for the RF model than the above covariates. Noticeably, its importance was computed at the 1 km spatial resolution, the different importance of land cover type may be found at higher spatial resolution. Such as land cover type had less important to SM at coarse spatial resolution (Gaur and Mohanty 2016; Joshi et al. 2010), but had strong correlation with *in-situ* SM (Baroni et al. 2013). Meanwhile, intuitively, precipitation was also closely related, SM-precipitation coupling had received increasing interest in recent years (Seneviratne et al. 2010). Although the importance of precipitation (less than 0.01) was not reflected in the RF model, this did not imply that precipitation had not impact on the~~

~~variation in SM. This could be attributed to the relatively small frequency for daily rainfall during several years periods, which led to a low ranking compared with other covariates based on the selection metric of RF importance ranking. It should be noted that the static variables and the reprocessed LAI provide information at 1km or 500m resolution, while ERA5-Land is at 9km resolution. So, the spatial details under 1km resolution came from the static variables and the reprocessed LAI rather than ERA5-Land. This aspect cannot reflect by the importance of RF as RF models were established to mainly reflect the temporal variation. This is because that we have much more samples of SM in the time dimension than those in the spatial dimension (1,789, the total number of stations). As a result, the importance of higher resolution variables (especially static variables) in estimating the spatial variation of SM was essentially underestimated by the importance of RF.~~

4. ~~4.~~Discussion

4.1 Relative importance of ~~covariate~~predictors

The relative importance of predictors at the ten soil depths is shown in Fig. 8 and Fig. S7. Bars present the variability of relative importance across the predictors. As presented in Fig. 8, the ERA5-Land SM is the most important to estimate *in-situ* SM from 10 to 100 cm soil depths. In addition to ERA5-Land SM, evapotranspiration, DEM, clay, reprocessed MODIS LAI (Version 6), porosity, LAI low vegetation, air temperature, LAI high vegetation and silt were followed. The importance of other predictors was less than 0.01, which were not discussed in this study. It was well known that evapotranspiration has strong correlation with SM dynamic under water-limited conditions (Albertsona and Kiely, 2001). So, evapotranspiration is greatly associated with SM in the regression model. Clay, porosity, rock fragment, silt and sand are soil properties that can affect SM values. Bissonnais et al. (1995) investigated SM for 31 soil types with different soil properties over Illinois and denoted that the available SM varied with regards soil groups. Soil properties could help RF model identify variation of SM more accurately. LAI is a vital parameter in the land surface and controls many complex processes in relation to vegetation, which determined evapotranspiration and further can impact on water balance (Chen et al., 2015). Air temperature and SM were closely related, so that from the hot to the cold, SM decreases for all kinds of land covers (Feng and Liu, 2015). However, air temperature shows significant effect on the RF based modelling performance for upper soil layers (at 10 cm and 20 cm soil depths) while it is less for the deeper soil (as presented in Fig. S7), as also stressed by Hu and Zheng (2003). It is commonly known that the land cover type is highly related to the variation of SM, but it got lower importance (less than 0.01) in the current RF modelling than the other predictors. Noticeably, this rate of importance was computed at the 1 km spatial resolution but other rates of importance for land cover type may be obtained at higher spatial resolution. Although land cover type shows less important to SM at coarse spatial resolution (Gaur and Mohanty, 2016; Joshi et al., 2010), it has strong correlation with *in-situ* SM data (Baroni et al., 2013). Meanwhile, intuitively, precipitation and SM were also closely related (Seneviratne et al., 2010). Although the importance of precipitation (less than 0.01) was not reflected in the RF modelling, this did not necessarily imply that precipitation could not impact on the variation of SM. This could be attributed to the relatively small frequency for daily rainfall during several years, which led to a low ranking compared with other predictors based on the RF importance ranking metrics. It should be noted that the static variables and the reprocessed LAI provide information is at 1km or 500m resolution, while ERA5-Land is at 9km resolution. So, the spatial details under 1km resolution came from the static variables and the reprocessed LAI rather than ERA5-Land. This aspect cannot be reflected well by the importance of RF as RF models were established to mainly reflect the temporal variation. This is because that we have much more samples of SM in the time dimension than those in the spatial dimension (1,648). As a result, the importance of higher resolution variables (especially static variables) in estimating the spatial variation of SM was essentially underestimated by the importance metric.

4.2 Sensitivity to precipitation, air temperature and radiation

We applied partial correlation to analysis the sensitivity between the meteorological variables (precipitation, air temperature and radiation) and SM data. As Fig. 9 shows, precipitation had stronger correlation with SM in SMCI1.0 and ERA5-Land data than that in SoMo.ml product across most regions in China, presenting significant positive partial correlations. Additionally, air temperature had significant positive partial correlation with SM in the north-western China, and negative partial correlations in north China and Liaoning province (118°53'E-125°46'E, 38°43'N-43°26'N) for SMCI1.0. The negative partial correlation between air temperature and SM is consistent with the physics of the process that higher evaporation is caused by higher air temperatures, leading to lower SM. In some of the plateau areas (73°19'E-104°47'E, 26°00'N-39°47'N), the shortwave radiation is the dominant factors for SM variability for SMCI1.0 product, physically sounds logic that the strong radiation in the plateau area has a great impact on the land surface process. Meanwhile, we also found that the shortwave radiation has the great influence on the SM variability in Tropical Monsoon Climate regions, which is also consistent with the previous study (Yao et al. 2011). The negative correlation between radiation and SM for SoMo.ml product in Temperature Climate region was stronger than that for SMCI1.0 product, which could explain more negative trends in SM in Temperature Climate region for SoMo.ml product. Compared with other SM products, the SMCI1.0 dataset shows similar spatial patterns for all the partial correlations. Overall, the SMCI1.0 product provides reasonable results in reflecting the relationship between SM and its related meteorological variables.

4.~~1~~23 Factors affecting the quality of SMCI1.0 dataset~~The quality of SMCI1.0 at spatial-temporal scale~~

~~In this study, the gridded soil moisture was estimated through RF method in China based on the ERA5-Land reanalysis, USGS land cover type and DEM, reprocessed LAI and soil properties from CSDL, which included soil depths from 10cm to 100cm and had 1km spatial and daily temporal resolution over the period from 1 January 2010 to 31 December 2020. The training efficiency was high (*RMSE*=0.601) due to the selection of important factors and vital hyper-parameters (*max_features*=1 and *min_samples_leaf*=20). In the year-to-year experiment, the *RMSE*, *MAE*, *R* and R^2 between gridded soil moisture and *in-situ* soil moisture ranged from 0.041-0.052, 0.03-0.036, 0.883-0.919 and 0.767-0.842, respectively. In the station-to-station experiment, the *RMSE*, *MAE*, *R* and R^2 between gridded soil moisture and *in-situ* soil moisture ranged from 0.045-0.051, 0.035-0.038, 0.866-0.893 and 0.749-0.798, respectively.~~

Fig. 2 and S2 show that SM results at 70 cm and 90 cm were significant worse than those at other depths. The reason may be that linked to the incapability of the RF model to estimate accurate SM when data from only a few *in-situ* SM stations are available. From Fig. S1 (b), we can see that the total numbers of data at 70 cm and 90 cm soil depths are quite small. In other words, more abundant of data could help RF model to 'learn' relationship between predictors and *in-situ* SM data reliably and further improve the quality of high-resolution SM estimation over China. Meanwhile, compared to the previous study of Sungmin et al. (2020), our SMCI1.0 showed the superior quality (Fig. 4-6), because the larger numbers of *in-situ* SM data of China wereapplied for the RF based modelling.

From Fig. 5, during the rainfall near 91th day across the Tropical Monsoon Climate zone (Am) and near 1st day across the Snow climate with dry winter zone (Dw), the *in-situ* SM values did not increase due to high precipitation, but the SMCI1.0 product could capture the increase in SM (denoted in the light blue rectangle). The reason may be that the applied predictors had bias with *in-situ* measurements and further affected the SM estimation by RF model. Meanwhile, we also found the RF model could overcome much bias in dry conditions, except for those from 196th to 305th days in the snow climate, fully humid zone (shown in the light red rectangle). In the case of 30 cm soil depth (Fig. S5), we could see an agreement between several peak events, it could be attributed to the soil texture homogeneity at the 10 and 30 cm soil depths. Almost all climatic regions had lower dynamic ranges at 30 cm soil depth than that at 10 cm, this may be attributed to the persistent behaviour of SM at 30 cm soil depth. In the case of 30 cm soil depth in Fig. S6, the SMCI1.0 product had higher accuracy than that at 10 cm soil depth (Fig. 6), especially in terms of *ubRMSE* and *MAE* metrics. The reason may be due to the background aridity which could lead to low variability of SM in the deeper layers (Karthikeyan and Mishra 2021) so that the RF model could capture the SM variation in SM straightforwardly.

Oppositely, it is inconsistent for the results of *R*, *ubRMSE*, and *MAE* in Fig. 2 and Fig. 4, which is similar to the previous study (Sungmin and Orth 2020) (represented in their Fig. 4 and Fig. 5). For example, SMCI1.0 product led to the *ubRMSE*, *MAE* and *R* values being 0.046, 0.035 and 0.889 at 10 cm soil depth in Fig. 2. However, in Fig. 4, the box-plot shows the lowest *ubRMSE*, *MAE* and highest *R* values of SMCI1.0 product as 0.03, 0.02, and 0.7, respectively. The reason may be due to the circumstances of computing the same metrics in different ways, so that the results of Fig. 2 are for all stations and temporal period, whereas Fig. 4 shows the results of temporal period at only one station.

The obtained results by RF method were also compared with those of some other ML models, including CatBoost (Dorogush et al. 2018), XgBoost (Chen et al. 2016), and Neural Network (Rosenblatt et al. 1958) models. We found that their performance is similar to RF models with a R^2 value around 0.79. Therefore, due to the comparable performance and wide application of RF to SM modelling (e.g., Carranza et al. 2021, Lin et al. 2022, Ly et al. 2021), and more importantly due to its cost effective run time, only the results of RF were considered to produce high-resolution SM data in this study.

4.~~3~~4 Requirement of further validations and improvements

SMCI1.0 product generally agrees well with *in-situ* SM data over ~~in~~ China ~~than~~ with regard to other considered datasets in ~~general, under the validations with~~ the year-to-year and station-to-station validation scenarios. However, we cannot ensure the same quality ~~of SMCI1.0 product in the whole~~over different parts of China. The reason is that *in-situ* SM stations are unevenly distributed over China~~ion, and the *in-situ* SM~~ with higher sparsity in the west~~ern China is sparse~~. We hope more *in-situ* SM stations are evenly deployed in China, which such data can ~~ensure~~ improve the quality of SM in most regions as far as possible. Triple collocation analysis (Karthikeyan and Mishra 2021) is also an alternative method for evaluating SMCI1.0 product. Meanwhile, there are many possible reasons for the failure of RF model, such as lack of ~~in~~sufficient data and the week ‘learning ~~ability~~’ of model-self. Hence, not only additional records from China are needed to be available, but also more robust estimated models ~~are~~ may be proposed and used for SM modelling ~~hoped to explored~~. ~~Such as~~For instance, the ~~single~~ deep learning models ~~are~~ can be built and optimized ~~in each~~for different homogeneous region (Karthikeyan and Mishra 2021), or the optical remote sensing ~~should~~ can be used for the human-induced regions (Chen et al. 2021), which ~~can~~ may lead to better estimation of ~~e~~ SM.

~~4.4 Higher-resolution SM estimating~~

~~As we know,~~It is well known that higher-resolution (<1km) SM estimation is typically considered as a complex and challenging task (Peng et al. 2020). The relative important ~~covariate~~predictors identified in Section 4.1 can ~~help estimating model~~ enhance modelling performance and generated data~~the~~ quality of higher-resolution SM product. The SMCI1.0 product may also ~~be acted~~be used as a vital ~~covariate~~predictor for improving the higher-resolution ~~(<1km)~~ SM products. ~~Next~~Moreover, downscaling to the higher-resolution SM product ~~generated~~ based on the lower-resolution ~~covariate~~predictors can also be ~~understand~~considered as super-resolution task in the computer science, and the advanced deep learning models ~~with high performance~~ can also be explored (Lei et al. 2020; Zhang et al. 2020; Zhu et al. 2021). ~~Of course, the target *in-situ* SM with dense distribution is also needed, thus can ensure the quality of high-resolution SM and further provide the reliable validation.~~

~~4.45 Sensitivity to precipitation and air temperature~~

~~We applied partial correlation to analysis the sensitivity between the meteorological variables (precipitation, air temperature and radiation) and SM. As Fig. 9 shown, precipitation had stronger correlation with SM in SMCI1.0 and ERA5-Land than that in SoMo.ml product across most regions in China, and it represented significant positive partial correlations. Additionally, air temperature had significant positive partial correlations with SM in the northwestern China, and negative partial correlations in north China and Liaoning province for SMCI1.0. The results with negative partial correlations between air temperature and SM were consistent with the physical knowledge that higher evaporation may be caused by higher air temperatures, and they also leaded to lower SM. In some of the plateau areas, the shortwave radiation was the dominant~~

~~factors of SM variability for SMCI1.0 product, it had the consistent physical knowledge which the strong radiation in the plateau area had a great impact on the land surface process. Meanwhile, we also found that the shortwave radiation had the great influence on the SM variability in Tropical Monsoon Climate regions, which was also consistent with the previous studies (Yao et al. 2011). The negative correlation between radiation and SM for SoMo.ml product in Temperature Climate region was stronger than that for SMCI1.0 product, which could explain more negative trends in SM in Temperature Climate region for SoMo.ml product. Compared with other SM products, the SMCI1.0 had similar spatial patterns for all the partial correlations. Overall, the SMCI1.0 product had reasonable quality in reflecting the relationship between SM and its related meteorological variables.~~

4.65 Comparison with previous products and implications for the soil moisture modelling

This section mainly described and discussed the comparison between SMCI1.0 and some other SM products, and the implications for the soil moisture modelling and attribution. From the results presented in Section 3, we can see that SMCI1.0 generally outperforms some other SM products (e.g., ERA5-Land, SoMo.ml and SMAP-L4) at most cases. The most important uniqueness of SMCI1.0 is taking the *in-situ* SM data as the training target with abundant sample size. Even though we used the ERA5-~~l~~Land to correct their means and standard deviation at each site, the temporal variation still came from the point observations. We have also examined the RF model training with the original SM observations and found that the performance of the model is much worse with a R^2 of 0.67 compared to the model with correction with a R^2 of 0.79. More importantly, the resulting SM maps demonstrated unreasonable noisy spatial distribution. These indicates that the *in-situ* SM in China have essential data inconsistency and the correction according to ERA5-Land is necessary which has physical consistency. Furthermore, SMCI1.0 has been provided with relatively high spatial and temporal resolutions (1-km, daily) for ten soil depths, which makes it possible for wider applications at finer scales and deep soils for the whole China, while reanalysis and remote sensing SM data are often at coarser resolution and remote sensing SM data are only for the surface soil.

As the limitation for the SMCI1.0, machine learning based model cannot always reflect the variation of SM well, especially for some extreme events or so called "tipping points" (Bury et al. 2021). From Fig,5, we can see that SMCI1.0 deviated from the *in-situ* SM data in some cases, though this also happened to the other three SM products. For example, from 35th day to 61th day across the Snow climate, fully humid (Df), SMCI1.0 and SoMo.ml overestimated, while SMAP_L4 underestimated. "Tipping points" denoted that slowly changing SM sparks a sudden shift to a new (Bury et al. 2021). This discontinuity creates a big challenge for estimating *in-situ* SM by ML models, because "tipping points" simplify the dynamics of complex system down to the limited number of possible "normal forms" (Bury et al. 2021). ML models cannot accurately capture such extreme events. Hence, for these extreme events, we hope ML models trained on a sufficiently diverse datasets of possible SM variation can well capture complex relationship between SM and predictors. As a suggestion for the future work, a possible solution for this limitation is to apply a Land surface model, such as Common Land Model (Dai et al. 2003), to simulate large numbers of SM data and select the local bifurcations in SM variation as supplementary samples to enhance the learning generality of the RF model.

5.Data and code availability

All resources of RF model, including training and testing code is publicly available at https://github.com/ljz1228/SMCI1.0_RF data with the resolution of 1 km and 9km can be accessed at http://dx.doi.org/10.11888/Terre.tpdc.272415 (Shangguan et al. 2022).

6.Conclusions

High resolution SM has several potential applications in flood and drought prediction and carbon cycle modelling. Currently available SM gridded products covering China or the world are ~~currently~~ often based on remote sensing data or based on numerical ~~modeling~~modelling. However, there is still a lack of SM data with high resolution at multiple layers based on *in-situ* measurements for China. In this study, the gridded SM data was estimated through RF method ~~in~~ over China based on the ERA5-Land reanalysis, USGS land cover type and DEM, reprocessed LAI and soil properties from CSDL, which included soil depths from 10cm to 100cm and had 1km spatial and daily temporal resolution over the period from 1 January 2010 to 31 December 2020. ~~Through this work, we generated a 1 km resolution long-term gridded SM data in China with *in-situ* measurements based on RF model, which has 10 layers up to 100 cm deep at daily resolution over the period 2010-2020.~~ Two independent experiments with *in-situ* soil moisture as the benchmark ~~are~~ were conducted to investigate the quality of SMCI1.0: year-to-year experiment (*ubRMSE* ranges from 0.041-0.052, *MAE* ranges from 0.03-0.036, *R* ranges from 0.883-0.919, and R^2 ranges from 0.767-0.842) and station-to-station experiment (*ubRMSE* ranges from 0.045-0.051, *MAE* ranges from 0.035-0.038, *R* ranges from 0.866-0.893, and R^2 ranges from 0.749-0.798). SMCI1.0 generally ~~has~~ showed advantages over other gridded ~~soil moisture~~SM products, including ERA5-Land, SMAP-L4 and SoMo.ml. Meanwhile, with regard to the ~~fit~~ agreement statistics (*ubRMSE*, *R*, *Bias*, and *MAE*), we could see that the SMCI1.0 product has relatively low *ubRMSE*, *Bias*, and *MAE* values over most regions. However, the high errors of ~~soil moisture~~SM obtained often locate~~d~~ in North China Monsoon Region. Moreover, SMCI1.0 has reasonable spatial pattern and demonstrate more spatial details compared with ~~existing~~ the compared SM products. As a result, the~~our~~ SMCI1.0 product based on *in-situ* data can be useful complements of existing model-based and satellite-based datasets for various hydrological, meteorological, and ecological analyses and ~~modeling~~modelling, especially for those applications requiring high resolution SM maps. Furter works may focus on improving the SM map by using advanced deep learning methods and adding more observations, especially for the west part of China. It is also possible to update and extent the time coverage of this data set before 2010 as long as in situ SM data becomes available.

7.Author contributions

WSG conceived the research and secured funding for the research. QLL and WSG performed the analyses. QLL wrote the first draft of the manuscript. GSS and QLL conducted the research. WSG and QLL reviewed and edited the manuscript~~paper~~ before submission. WSG, QLL and VN revised the manuscript. All other authors joined the discussion of the research.

8.Competing interests

The authors declare that they have no conflict of interest.

9.Acknowledgements

The authors are grateful to all the data contributors who made it possible to complete this research.

10.Financial support

The study was partially supported by the National Key Research and Development Program of China under grants 2017YFA0604303 and the National Natural Science Foundation of China, grant number 42105144, 41975122 and U1811464.

References

Albergel, C., Dutra, E., Munier, S., Calvet, J. C., Munoz-Sabater, J., de Rosnay, P., and Balsamo, G.: ERA-5 and ERA-Interim driven ISBA land surface model simulations: which one performs better?~~,~~ Hydrol. Earth Syst. Sci., 22, 3515-3532, https://doi.org/10.5194/hess-22-3515-2018, 2018.

Albertson, J. D. and Kiely, G.: On the structure of soil moisture time series in the context of land surface models, Journal of Hydrology, 243, 101-119, https://doi.org/10.1016/S0022-1694(00)00405-4, 2001.

Balenović, I., Marjanović, H., Vuletić, D., Paladinić, E., and Indir, K.: Quality assessment of high density digital surface model over different land cover classes, Periodicum Biologorum, 117, 459-470, https://doi.org/10.18054/pb.2015.117.4.3452, 2016.

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., Dee, D., Dutra, E., Muñoz-Sabater, J., Pappenberger, F., de Rosnay, P., Stockdale, T., and Vitart, F.: ERA-Interim/Land: a global land surface reanalysis data set, Hydrol. Earth Syst. Sci., 19, 389-407, https://doi.org/10.5194/hess-19-389-2015, 2015.

Baroni, G., Ortuani, B., Facchi, A., and Gandolfi, C.: The role of vegetation and soil properties on the spatio-temporal variability of the surface soil moisture in a maize-cropped field, Journal of Hydrology, 489, 148-159, https://doi.org/10.1016/j.jhydrol.2013.03.007, 2013.

Brocca, L., Morbidelli, R., Melone, F., and Moramarco, T.: Soil moisture spatial variability in experimental areas of central Italy, Journal of Hydrology, 333, 356-373, https://doi.org/10.1016/j.jhydrol.2006.09.004, 2007.

Bury Thomas, M., Sujith, R. I., Pavithran, I., Scheffer, M., Lenton Timothy, M., Anand, M., and Bauch Chris, T.: Deep learning for early warning signals of tipping points, Proceedings of the National Academy of Sciences, 118, e2106140118, https://doi.org/10.1073/pnas.2106140118, 2021.

Carranza, C., Nolet, C., Pezij, M., and van der Ploeg, M.: Root zone soil moisture estimation with Random Forest, Journal of Hydrology, 593, 125840, https://doi.org/10.1016/j.jhydrol.2020.125840, 2021.

Chakrabarti, S., Bongiovanni, T., Judge, J., Nagarajan, K., and Principe, J. C.: Downscaling Satellite-Based Soil Moisture in Heterogeneous Regions Using High-Resolution Remote Sensing Products and Information Theory: A Synthetic Study, IEEE Transactions on Geoscience and Remote Sensing, 53, 85-101, https://doi.org/10.1109/TGRS.2014.2318699, 2015.

Chawla, I., Karthikeyan, L., and Mishra, A. K.: A review of remote sensing applications for water security: Quantity, quality, and extremes, Journal of Hydrology, 585, 124826, https://doi.org/10.1016/j.jhydrol.2020.124826, 2020.

Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, ACM, https://doi.org/10.1145/2939672.2939785, 2016.

Chen, M., Willgoose, G. R., and Saco, P. M.: Investigating the impact of leaf area index temporal variability on soil moisture predictions using remote sensing vegetation data, Journal of Hydrology, 522, 274-284, https://doi.org/10.1016/j.jhydrol.2014.12.027, 2015.

Chen, Y., Feng, X., and Fu, B.: An improved global remote-sensing-based surface soil moisture (RSSSM) dataset covering 2003–2018, Earth Syst. Sci. Data, 13, 1-31, https://doi.org/10.5194/essd-13-1-2021, 2021.

Cong, N., Wang, T., Nan, H., Ma, Y., Wang, X., Myneni, R. B., and Piao, S.: Changes in satellite-derived spring vegetation green-up date and its linkage to climate in China from 1982 to 2010: a multimethod analysis, Global Change Biology, 19, 881-891, https://doi.org/10.1111/gcb.12077, 2013.

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

Dai, Y., Zeng, X., Dickinson, R. E., Baker, I., Bonan, G. B., Bosilovich, M. G., Denning, A. S., Dirmeyer, P. A., Houser, P. R., Niu, G., Oleson, K. W., Schlosser, C. A., and Yang, Z.-L.: The Common Land Model, Bulletin of the American Meteorological Society, 84, 1013-1024, https://doi.org/10.1175/BAMS-84-8-1013, 2003.

Dirmeyer, P. A., Gao, X., Zhao, M., Guo, Z., Oki, T., and Hanasaki, N.: GSWP-2: Multimodel Analysis and Implications for Our Perception of the Land Surface, Bulletin of the American Meteorological Society, 87, 1381-1398, https://doi.org/10.1175/BAMS-87-10-1381, 2006.

Dorigo, W. A., Xaver, A., Vreugdenhil, M., Gruber, A., Hegyiová, A., Sanchis-Dufau, A. D., Zamojski, D., Cordes, C., Wagner, W., and Drusch, M.: Global Automated Quality Control of In Situ Soil Moisture Data from the International Soil Moisture Network, Vadose Zone Journal, 12, vzj2012.0097, https://doi.org/10.2136/vzj2012.0097, 2013.

Dorogush, A. V., Ershov, V., and Gulin, A.: CatBoost: gradient boosting with categorical features support, arXiv, https://doi.org/10.48550/arXiv.1810.11363, 2018.

Entekhabi, D., Rodriguez-Iturbe, I., and Castelli, F.: Mutual interaction of soil moisture state and atmospheric processes, Journal of Hydrology, 184, 3-17, https://doi.org/10.1016/0022-1694(95)02965-6, 1996.

Entekhabi, D., Njoku, E. G., Neill, P. E. O., Kellogg, K. H., Crow, W. T., Edelstein, W. N., Entin, J. K., Goodman, S. D., Jackson, T. J., Johnson, J., Kimball, J., Piepmeier, J. R., Koster, R. D., Martin, N., McDonald, K. C., Moghaddam, M., Moran, S., Reichle, R., Shi, J. C., Spencer, M. W., Thurman, S. W., Tsang, L., and Zyl, J. V.: The Soil Moisture Active Passive (SMAP) Mission, Proceedings of the IEEE, 98, 704-716, https://doi.org/10.1109/JPROC.2010.2043918, 2010.

Feng, H. and Liu, Y.: Combined effects of precipitation and air temperature on soil moisture in different land covers in a humid basin, Journal of Hydrology, 531, 1129-1140, https://doi.org/10.1016/j.jhydrol.2015.11.016, 2015.

Fujii, H., Koike, T., and Imaoka, K.: Improvement of the AMSR-E Algorithm for Soil Moisture Estimation by Introducing a Fractional Vegetation Coverage Dataset Derived from MODIS Data, Journal of The Remote Sensing Society of Japan, 29, 282-292, https://doi.org/10.11440/rssj.29.282, 2009.

Gao, L. and Shao, M.: Temporal stability of shallow soil water content for three adjacent transects on a hillslope, Agricultural Water Management, 110, 41-54, https://doi.org/10.1016/j.agwat.2012.03.012, 2012.

Gaur, N. and Mohanty, B. P.: Land-surface controls on near-surface soil moisture dynamics: Traversing remote sensing footprints, Water Resources Research, 52, 6365-6385, https://doi.org/10.1002/2015WR018095, 2016.

Gcos, G. C. O. S.: The Global Observing System for Climate: Implementation Needs, https://doi.org/10.13140/RG.2.2.23178.26566, 2016.

Gruber, A., Su, C. H., Crow, W. T., Zwieback, S., Dorigo, W. A., and Wagner, W.: Estimating error cross-correlations in soil moisture data sets using extended collocation analysis, Journal of Geophysical Research: Atmospheres, 121, 1208-1219, https://doi.org/10.1002/2015JD024027, 2016.

Gu, X., Li, J., Chen, Y. D., Kong, D., and Liu, J.: Consistency and Discrepancy of Global Surface Soil Moisture Changes From Multiple Model-Based Data Sets Against Satellite Observations, Journal of Geophysical Research: Atmospheres, 124, 1474-1495, https://doi.org/10.1029/2018JD029304, 2019.

Guo, L. and Lin, H.: Chapter Two - Addressing Two Bottlenecks to Advance the Understanding of Preferential Flow in Soils, in: Advances in Agronomy, edited by: Sparks, D. L., Academic Press, 61-117, https://doi.org/10.1016/bs.agron.2017.10.002, 2018.

Hu, Q. and Feng, S.: A Daily Soil Temperature Dataset and Soil Temperature Climatology of the Contiguous United States, Journal of Applied Meteorology, 42, 1139-1156, https://doi.org/10.1175/1520-0450(2003)042<1139:ADSTDA>2.0.CO;2, 2003.

Joshi, C. and Mohanty, B. P.: Physical controls of near-surface soil moisture across varying spatial scales in an agricultural landscape during SMEX02, Water Resources Research, 46, https://doi.org/10.1029/2010WR009152, 2010.

Karthikeyan, L. and Kumar, D. N.: A novel approach to validate satellite soil moisture retrievals using precipitation data, Journal of Geophysical Research Atmospheres, 121, https://doi.org/10.1002/2016JD024829, 2016.

Karthikeyan, L. and Mishra, A. K.: Multi-layer high-resolution soil moisture estimation using machine learning over the United States, Remote Sensing of Environment, 266, 112706, https://doi.org/10.1016/j.rse.2021.112706, 2021.

Kemppinen, J., Niittynen, P., Virkkala, A.-M., Happonen, K., Riihimäki, H., Aalto, J., and Luoto, M.: Dwarf Shrubs Impact Tundra Soils: Drier, Colder, and Less Organic Carbon, Ecosystems, 24, 1378-1392, https://doi.org/10.1007/s10021-020-00589-2, 2021.

Kerr, Y. H., Waldteufel, P., Wigneron, J., Delwart, S., Cabot, F., Boutin, J., Escorihuela, M., Font, J., Reul, N., Gruhier, C., Juglea, S. E., Drinkwater, M. R., Hahne, A., Martín-Neira, M., and Mecklenburg, S.: The SMOS Mission: New Tool for Monitoring Key Elements ofthe Global Water Cycle, Proceedings of the IEEE, 98, 666-687, https://doi.org/10.1109/JPROC.2010.2043032, 2010.

Kim, S., Zhang, R., Pham, H., and Sharma, A.: A Review of Satellite-Derived Soil Moisture and Its Usage for Flood Estimation, Remote Sensing in Earth Systems Sciences, 2, 225-246, https://doi.org/10.1007/s41976-019-00025-7, 2019.

Kim, H., Wigneron, J.-P., Kumar, S., Dong, J., Wagner, W., Cosh, M. H., Bosch, D. D., Collins, C. H., Starks, P. J., Seyfried, M., and Lakshmi, V.: Global scale error assessments of soil moisture estimates from microwave-based active and passive satellites and land surface models over forest and mixed irrigated/dryland agriculture regions, Remote Sensing of Environment, 251, 112052, https://doi.org/10.1016/j.rse.2020.112052, 2020.

Kumar, S. V., Reichle, R. H., Koster, R. D., Crow, W. T., and Peters-Lidard, C. D.: Role of Subsurface Physics in the Assimilation of Surface Soil Moisture Observations, Journal of Hydrometeorology, 10, 1534-1547, https://doi.org/10.1175/2009JHM1134.1, 2009.

Le Bissonnais, Y., Renaux, B., and Delouche, H.: Interactions between soil properties and moisture content in crust formation, runoff and interrill erosion from tilled loess soils, CATENA, 25, 33-46, https://doi.org/10.1016/0341-8162(94)00040-L, 1995.

Lei, S., Shi, Z., and Zou, Z.: Coupled Adversarial Training for Remote Sensing Image Super-Resolution, IEEE Transactions on Geoscience and Remote Sensing, 58, 3633-3643, https://doi.org/10.1109/TGRS.2019.2959020, 2020.

Li, Q., Wang, Z., Shangguan, W., Li, L., Yao, Y., and Yu, F.: Improved daily SMAP satellite soil moisture prediction over China using deep learning model with transfer learning, Journal of Hydrology, 600, 126698, https://doi.org/10.1016/j.jhydrol.2021.126698, 2021.

Li, L., Shangguan, W., Deng, Y., Mao, J., Pan, J., Wei, N., Yuan, H., Zhang, S., Zhang, Y., and Dai, Y.: A Causal Inference Model Based on Random Forests to Identify the Effect of Soil Moisture on Precipitation, Journal of Hydrometeorology, 21, 1115-1131, https://doi.org/10.1175/JHM-D-19-0209.1, 2020.

Lin, L. and Liu, X.: Mixture-based weight learning improves the random forest method for hyperspectral estimation of soil total nitrogen, Computers and Electronics in Agriculture, 192, 106634, https://doi.org/10.1016/j.compag.2021.106634, 2022.

Ly, H. B., Nguyen, T. A., and Pham, B. T.: Estimation of Soil Cohesion Using Machine Learning Method: A Random Forest Approach, Advances in Civil Engineering, https://doi.org/10.1155/2021/8873993, 2021.

Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, Z., Yang, L., and Merchant, J. W.: Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data, International Journal of Remote Sensing, 21, 1303-1330, https://doi.org/10.1080/014311600210191, 2000.

Mao T, Shangguan W, Li Q, Li L, Zhang Y, Huang F, Li J, Liu W, and Zhang R. A Spatial Downscaling Method for Remote Sensing Soil Moisture Based on Random Forest Considering Soil Moisture Memory and Mass Conservation. Remote Sensing. 14, 3858, https://doi.org/10.3390/rs14163858, 2022.

Meng, X., Mao, K., Meng, F., Shi, J., Zeng, J., Shen, X., Cui, Y., Jiang, L., and Guo, Z.: A fine-resolution soil moisture dataset for China in 2002–2018, Earth Syst. Sci. Data, 13, 3239-3261, https://doi.org/10.5194/essd-13-3239-2021, 2021.

Mishra, A., Vu, T., Veettil, A. V., and Entekhabi, D.: Drought monitoring with soil moisture active passive (SMAP) measurements, Journal of Hydrology, 552, 620-632, https://doi.org/10.1016/j.jhydrol.2017.07.033, 2017.

Mohamed, E., Habib, E., Abdelhameed, A. M., and Bayoumi, M.: Assessment of a Spatiotemporal Deep Learning Approach for Soil Moisture Prediction and Filling the Gaps in Between Soil Moisture Observations, Frontiers in Artificial Intelligence, 4, https://doi.org/10.3389/frai.2021.636234, 2021.

Muñoz Sabater, J.: ERA5-Land hourly data from 1981 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS) [data set], https://doi.org/10.24381/cds.e2161bac, 2019.

Muñoz Sabater, J.: ERA5-Land hourly data from 1950 to 1980. Copernicus Climate Change Service (C3S) Climate Data Store (CDS) [data set], https://doi.org/10.24381/cds.e2161bac, 2021.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-N.: ERA5-Land: A state-of-the-art global reanalysis dataset for land applications, Earth Syst. Sci. Data,13, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021, 2021.

Norbiato, D., Borga, M., Degli Esposti, S., Gaume, E., and Anquetin, S.: Flash flood warning based on rainfall thresholds and soil moisture conditions: An assessment for gauged and ungauged basins, Journal of Hydrology, 362, 274-290, https://doi.org/10.1016/j.jhydrol.2008.08.023, 2008.

O, S. and Orth, R.: Global soil moisture data derived through machine learning trained with in-situ measurements, Scientific Data, 8, 170, https://doi.org/10.1038/s41597-021-00964-1, 2021.

O, S., Hou, X., and Orth, R.: Observational evidence of wildfire-promoting soil moisture anomalies, Scientific Reports, 10, 11008, https://doi.org/10.1038/s41598-020-67530-4, 2020.

Ojha, R., Morbidelli, R., Saltalippi, C., Flammini, A., and Govindaraju, R. S.: Scaling of surface soil moisture over heterogeneous fields subjected to a single rainfall event, Journal of Hydrology, 516, 21-36, https://doi.org/10.1016/j.jhydrol.2014.01.057, 2014.

Orth, R. and Seneviratne, S. I.: Using soil moisture forecasts for sub-seasonal summer temperature predictions in Europe, Climate Dynamics, 43, 3403-3418, https://doi.org/10.1007/s00382-014-2112-x, 2014.

Pan, J., Shangguan, W., Li, L., Yuan, H., Zhang, S., Lu, X., Wei, N., and Dai, Y.: Using data-driven methods to explore the predictability of surface soil moisture with FLUXNET site data, Hydrological Processes, 33, 2978-2996, https://doi.org/10.1002/hyp.13540, 2019.

Parinussa, R. M., Lakshmi, V., Johnson, F. M., and Sharma, A.: A new framework for monitoring flood inundation using readily available satellite data, Geophysical Research Letters, 43, 2599-2605, https://doi.org/10.1002/2016GL068192, 2016.

Peng, J., Albergel, C., Balenzano, A., Brocca, L., Cartus, O., Cosh, M. H., Crow, W. T., Dabrowska-Zielinska, K., Dadson, S., Davidson, M. W. J., de Rosnay, P., Dorigo, W., Gruber, A., Hagemann, S., Hirschi, M., Kerr, Y. H., Lovergine, F., Mahecha, M. D., Marzahn, P., Mattia, F., Musial, J. P., Preuschmann, S., Reichle, R. H., Satalino, G., Silgram, M., van Bodegom, P. M., Verhoest, N. E. C., Wagner, W., Walker, J. P., Wegmüller, U., and Loew, A.: A roadmap for high-resolution satellite soil moisture applications – confronting product characteristics with user requirements, Remote Sensing of Environment, 252, 112162, https://doi.org/10.1016/j.rse.2020.112162, 2021.

Rezaei Moghadam, H., Hosseinalizadeh, M., Sheikh, V., and Jafari, R.: SOIL MOISTURE ESTIMATION USING DIGITAL ELEVATION MODEL (DEM), JOURNAL OF RS AND GIS FOR NATURAL RESOURCES (JOURNAL OF APPLIED RS AND GIS TECHNIQUES IN NATURAL RESOURCE SCIENCE), 6, 61-71, https://www.sid.ir/en/journal/ViewPaper.aspx?id=517799, 2015.

Rosenblatt, F.: The perceptron: A probabilistic model for information storage and organization in the brain, Psychological Review, 65, 386-408, https://doi.org/10.1037/h0042519, 1958.

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

Shangguan, W., Li, Q., and Shi, G.: A 1-km daily soil moisture dataset over~~of~~ China based on in-situ measurement (200~~1~~0-2020), National Tibetan Plateau Data Center [dataset], https://doi.org/10.11888/Terre.tpdc.272415, 2022.

Shangguan, W., Dai, Y., Liu, B., Zhu, A., Duan, Q., Wu, L., Ji, D., Ye, A., Yuan, H., Zhang, Q., Chen, D., Chen, M., Chu, J., Dou, Y., Guo, J., Li, H., Li, J., Liang, L., Liang, X., Liu, H., Liu, S., Miao, C., and Zhang, Y.: A China data set of soil properties for land surface modeling, Journal of Advances in Modeling Earth Systems, 5, 212-224, https://doi.org/10.1002/jame.20026, 2013.

Srivastava, P. K., Han, D., Ramirez, M. R., and Islam, T.: Machine Learning Techniques for Downscaling SMOS Satellite Soil Moisture Using MODIS Land Surface Temperature for Hydrological Application, Water Resources Management, 27, 3127-3144, https://doi.org/10.1007/s11269-013-0337-9, 2013.

Tijdeman, E. and Menzel, L.: The development and persistence of soil moisture stress during drought across southwestern Germany, Hydrol. Earth Syst. Sci., 25, 2009-2025, https://doi.org/10.5194/hess-25-2009-2021, 2021.

Vereecken, H., Huisman, J. A., Pachepsky, Y., Montzka, C., van der Kruk, J., Bogena, H., Weihermüller, L., Herbst, M., Martinez, G., and Vanderborght, J.: On the spatio-temporal dynamics of soil moisture at the field scale, Journal of Hydrology, 516, 76-96, https://doi.org/10.1016/j.jhydrol.2013.11.061, 2014.

Wagner, W., Blöschl, G., Pampaloni, P., Calvet, J.-C., Bizzarri, B., Wigneron, J.-P., and Kerr, Y.: Operational readiness of microwave remote sensing of soil moisture for hydrologic applications, Hydrology Research, 38, 1-20, https://doi.org/10.2166/nh.2007.029, 2007.

Wang, X.., Pan, Y.., Zhang, Y.., Dou, D., Hu, R., and Zhang, H.: Temporal stability analysis of surface and subsurface soil moisture for a transect in artificial revegetation desert area, China, Journal of Hydrology, 507, 100-109, https://doi.org/10.1016/j.jhydrol.2013.10.021, 2013.

Wang, Y., Mao, J., Jin, M., Hoffman, F. M., Shi, X., Wullschleger, S. D., and Dai, Y.: Development of observation-based global multilayer soil moisture products for 1970 to 2016, Earth Syst. Sci. Data, 13, 4385-4405, https://doi.org/10.5194/essd-13-4385-2021, 2021.

Wei, Z., Meng, Y., Zhang, W., Peng, J., and Meng, L.: Downscaling SMAP soil moisture estimation with gradient boosting decision tree regression over the Tibetan Plateau, Remote Sensing of Environment, 225, 30-44, https://doi.org/10.1016/j.rse.2019.02.022, 2019.

Xu, J. W., Zhao, J. F., Zhang, W. C., and Xu, X. X.: A Novel Soil Moisture Predicting Method Based on Artificial Neural Network and Xinanjiang Model, Advanced Materials Research, 121-122, 1028-1032, https://doi.org/10.4028/www.scientific.net/AMR.121-122.1028, 2010.

Yao Yun-Jun, Qin Qi-Ming, Zhao Shao-Hua, and Wei-Lin, Y.: Retrieval of soil moisture based on MODIS shortwave infrared spectral feature, J.Infrared Millim.Waves, 30, 9-14, http://journal.sitp.ac.cn/hwyhmb/hwyhmbcn/article/abstract/100118, 2011.

Yuan, H., Dai, Y., Xiao, Z., Ji, D., and Shangguan, W.: Reprocessing the MODIS Leaf Area Index products for land surface and climate modelling, Remote Sensing of Environment, 115, 1171-1187, https://doi.org/10.1016/j.rse.2011.01.001, 2011.

Yuan, Q., Xu, H., Li, T., Shen, H., and Zhang, L.: Estimating surface soil moisture from satellite observations using a generalized regression neural network trained on sparse ground-based measurements in the continental U.S, Journal of Hydrology, 580, 124351, https://doi.org/10.1016/j.jhydrol.2019.124351, 2020.

Zeng, L., Hu, S., Xiang, D., Zhang, X., Li, D., Li, L., and Zhang, T.: Multilayer Soil Moisture Mapping at a Regional Scale from Multisource Data via a Machine Learning Method, Remote Sensing, 11, https://doi.org/10.3390/rs11030284, 2019.

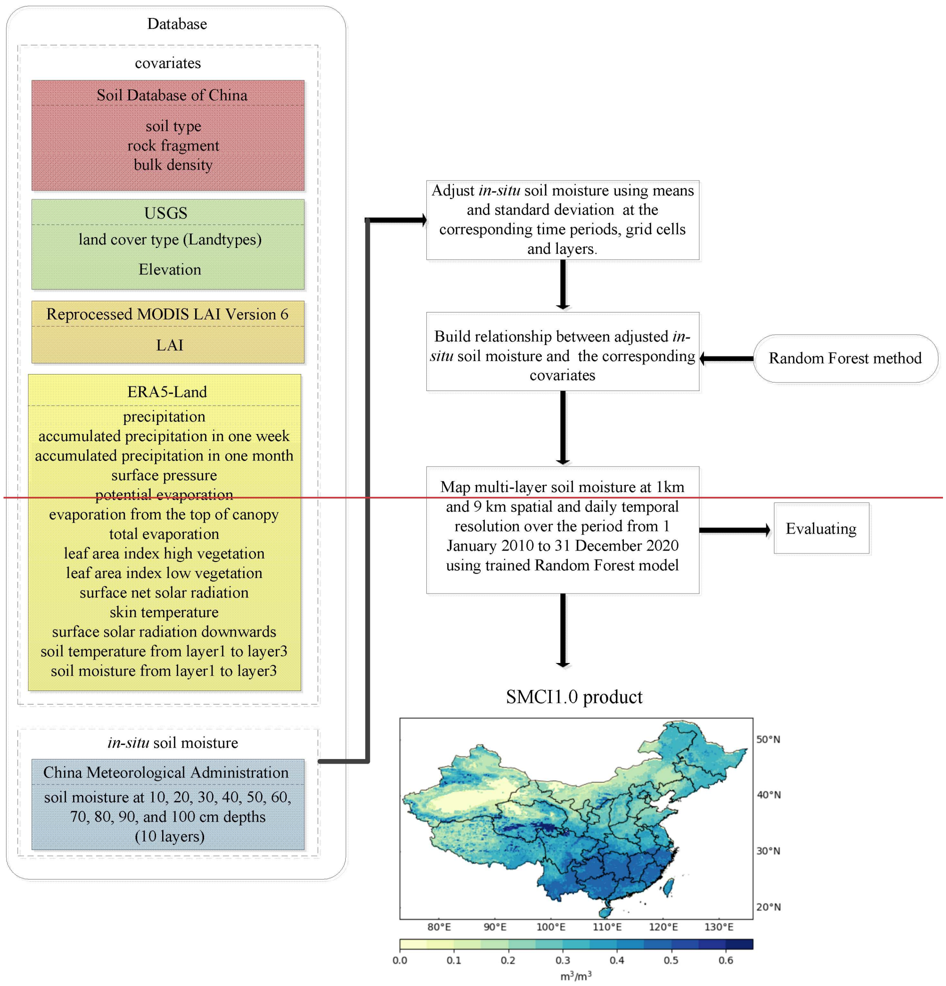
Zhang, H., Wang, P., and Jiang, Z.: Nonpairwise-Trained Cycle Convolutional Neural Network for Single Remote Sensing Image Super-Resolution, IEEE Transactions on Geoscience and Remote Sensing, 59, 4250-4261, https://doi.org/10.1109/TGRS.2020.3009224, 2021.

Zhang, R., Kim, S., and Sharma, A.: A comprehensive validation of the SMAP Enhanced Level-3 Soil Moisture product using ground measurements over varied climates and landscapes, Remote Sensing of Environment, 223, 82-94, https://doi.org/10.1016/j.rse.2019.01.015, 2019.

Zhang, Q., Yuan, Q., Li, J., Wang, Y., Sun, F., and Zhang, L.: Generating seamless global daily AMSR2 soil moisture (SGD-SM) long-term products for the years 2013–2019, Earth Syst. Sci. Data, 13, 1385-1401, https://doi.org/10.5194/essd-13-1385-2021, 2021.

Zhu, X., Guo, K., Ren, S., Hu, B., Hu, M., and Fang, H.: Lightweight Image Super-Resolution With Expectation-Maximization Attention Mechanism, IEEE Transactions on Circuits and Systems for Video Technology, 32, 1273-1284, https://doi.org/10.1109/TCSVT.2021.3078436, 2022.

Zwieback, S., Chang, Q., Marsh, P., and Berg, A.: Shrub tundra ecohydrology: rainfall interception is a major component of the water balance, Environmental Research Letters, 14, 055005, https://doi.org/10.1088/1748-9326/ab1049, 2019.


Database
covariates
Soil Database of China
soil type
rock fragment
bulk density
USGS
land cover type (Landtypes)
Elevation
Reprocessed MODIS LAI Version 6
LAI
ERA5-Land
precipitation
accumulated precipitation in one week
accumulated precipitation in one month
surface pressure
potential evaporation
evaporation from the top of canopy
total evaporation
leaf area index high vegetation
leaf area index low vegetation
surface net solar radiation
skin temperature
surface solar radiation downwards
soil temperature from layer1 to layer3
soil moisture from layer1 to layer3
in-situ soil moisture
China Meteorological Administration
soil moisture at 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100 cm depths (10 layers)
Adjust in-situ soil moisture using means and standard deviation at the corresponding time periods, grid cells and layers.
Build relationship between adjusted in-situ soil moisture and the corresponding covariates
Random Forest method
Map multi-layer soil moisture at 1km and 9 km spatial and daily temporal resolution over the period from 1 January 2010 to 31 December 2020 using trained Random Forest model
Evaluating
SMCI1.0 product
50°N
40°N
30°N
20°N
80°E
90°E
100°E
110°E
120°E
130°E
0.0
0.1
0.2
0.3
0.4
0.5
0.6
m3/m3

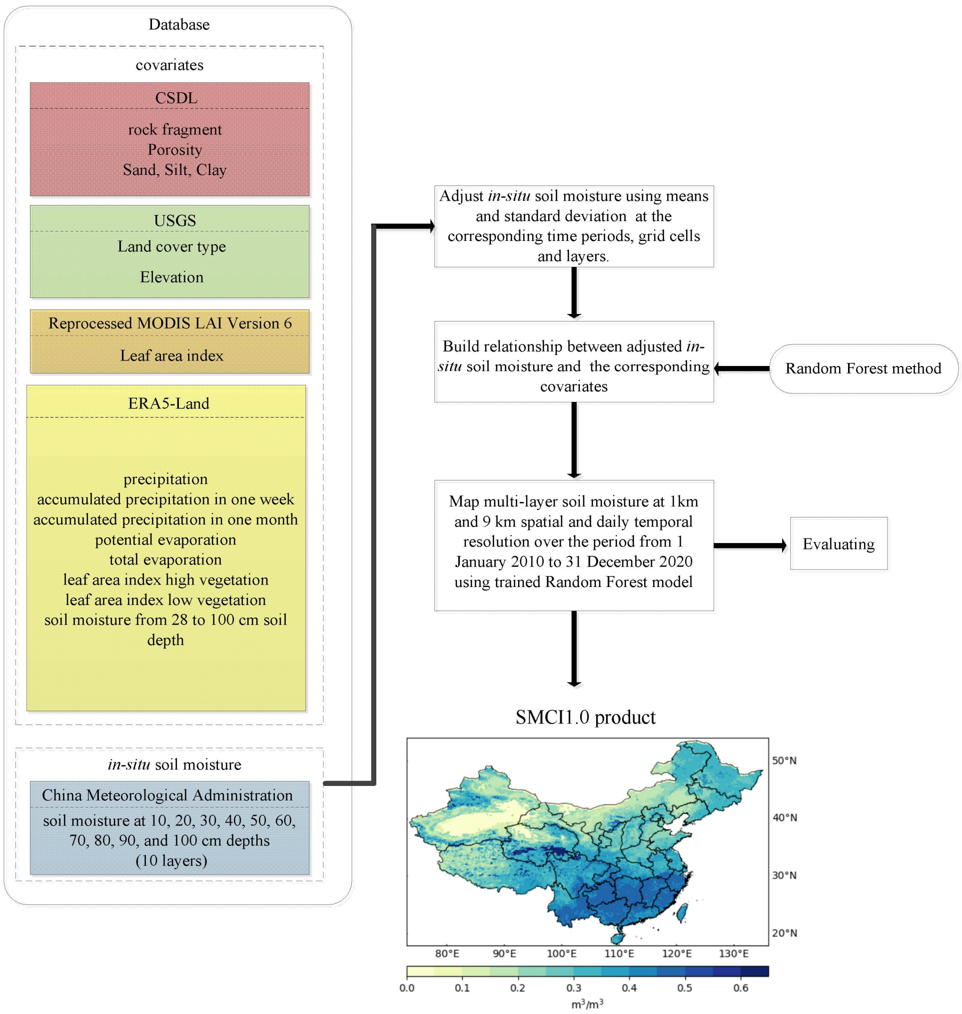
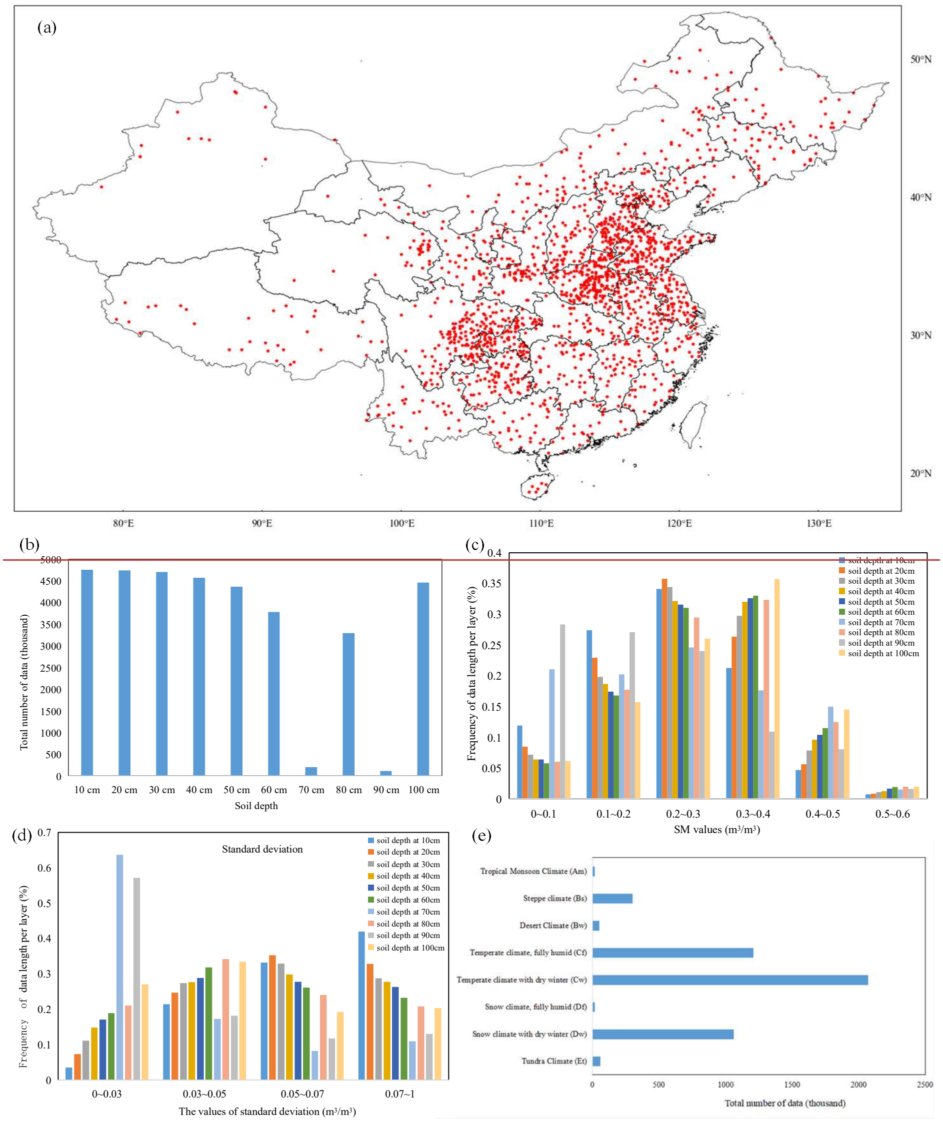


Figure 1: Generation process for the SMCI1.0 product with 1km spatial resolution and daily temporal resolution over the period from 1 January ~~2010~~ 2000 to 31 December 2020 over China.

域代码已更改


(a)
50°N
40°N
30°N
20°N
80°E
90°E
100°E
110°E
120°E
130°E
(b)
Total number of data (thousand)
10 cm
20 cm
30 cm
40 cm
50 cm
60 cm
70 cm
80 cm
90 cm
100 cm
Soil depth
(c)
Frequency of data length per layer (%)
0~0.1
0.1~0.2
0.2~0.3
0.3~0.4
0.4~0.5
0.5~0.6
SM values (m³/m³)
soil depth at 10cm
soil depth at 20cm
soil depth at 30cm
soil depth at 40cm
soil depth at 50cm
soil depth at 60cm
soil depth at 70cm
soil depth at 80cm
soil depth at 90cm
soil depth at 100cm
(d)
Standard deviation
Frequency of data length per layer (%)
0~0.03
0.03~0.05
0.05~0.07
0.07~1
The values of standard deviation (m³/m³)
(e)
Tropical Monsoon Climate (Am)
Steppe climate (Bs)
Desert Climate (Bw)
Temperate climate, fully humid (Cf)
Temperate climate with dry winter (Cw)
Snow climate, fully humid (Df)
Snow climate with dry winter (Dw)
Tundra Climate (Et)
Total number of data (thousand)

~~**Figure 2: (a) The locations of all stations in China; (b) Total data number per soil depth; (c) Frequency of data length per layer for SM values; (d) Frequency of data length per layer for standard deviation; (e) Total data number per climate zone.**~~

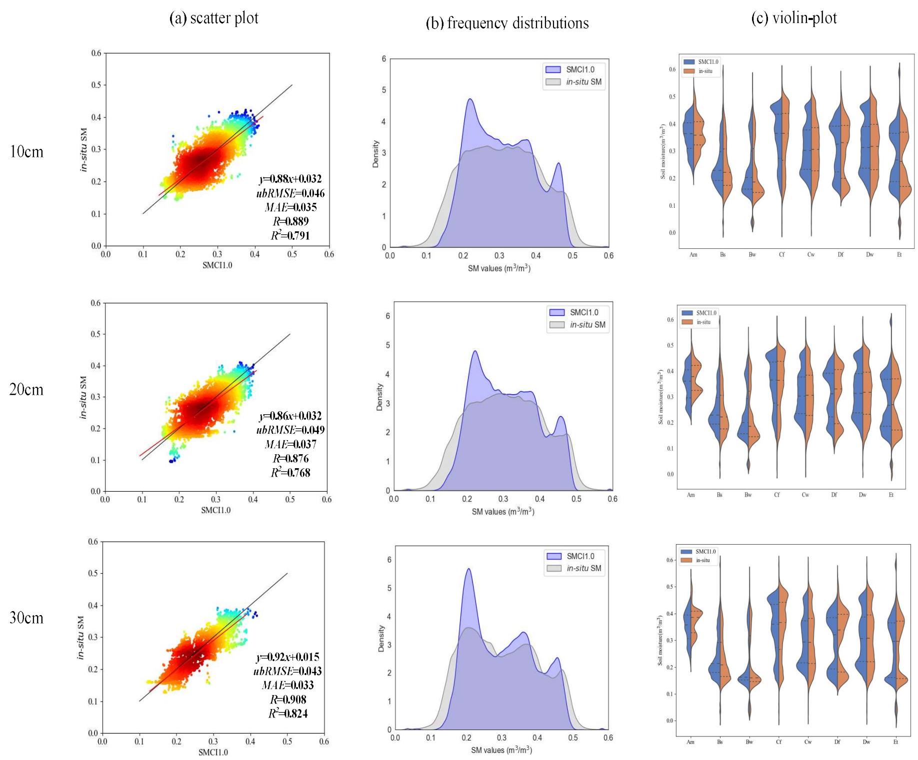


Figure 2: Comparisons between SMCI1.0 and *in-situ* SM from 10 to 30 cm soil depth in year-to-year experiment: comparison of (a) the scatter plot between the mean of SMCI1.0 and that of *in-situ* SM at each station, (b) the frequency distributions of all SM values in SMCI1.0 and that in *in-situ* measurements, (c) the violin-plot for the distribution of daily SM from stations for each climate type.

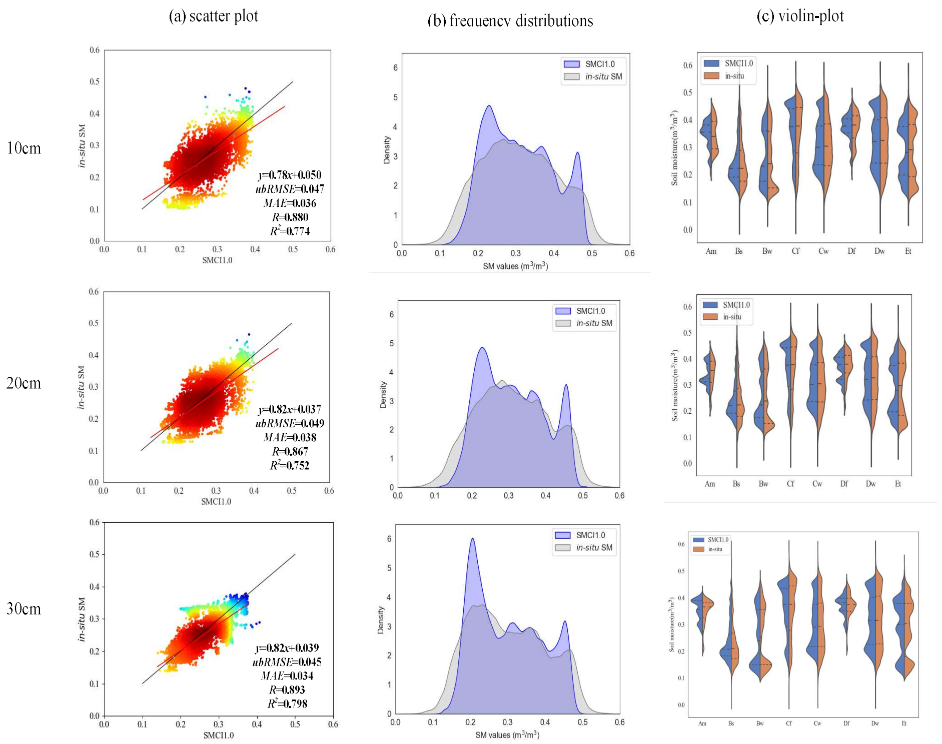
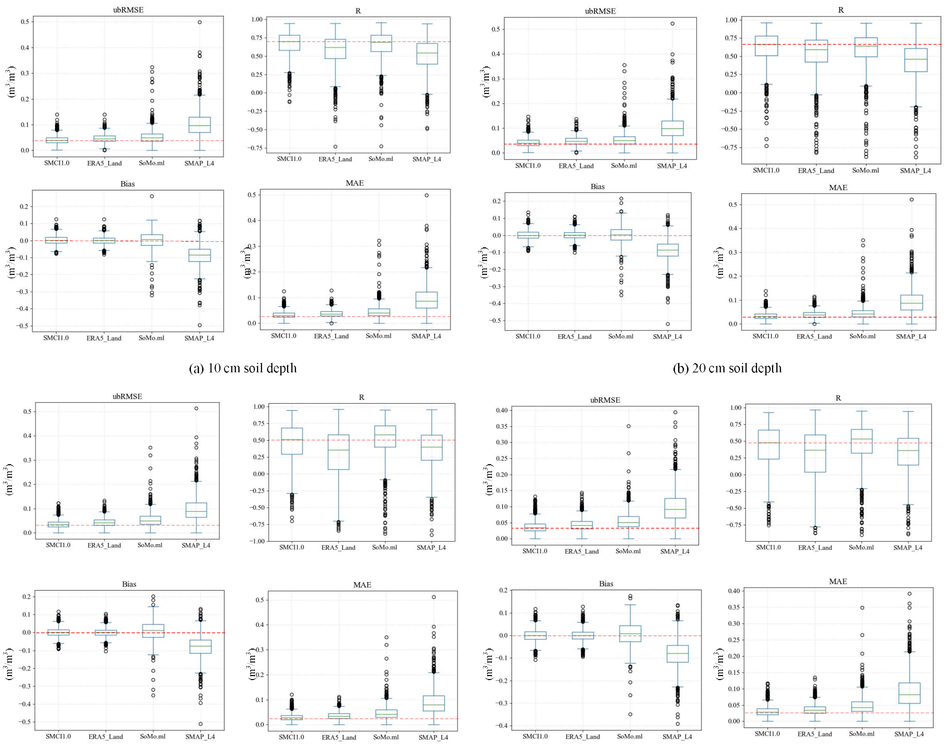


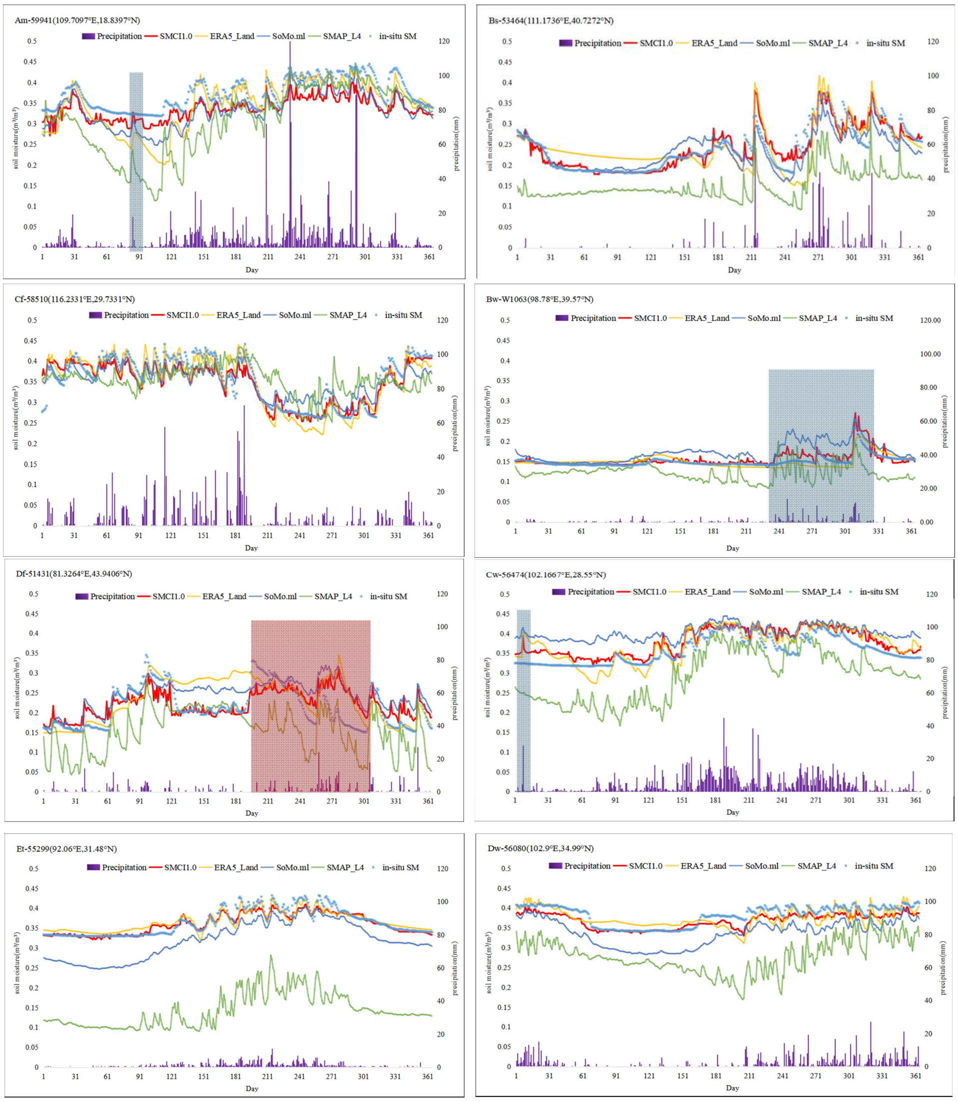
Figure 3: Same as Fig. 32 but for station-to-station estimating.



(a) 10 cm soil depth (b) 20 cm soil depth

(c) 30 cm soil depth (d) 40 cm soil depth

Figure 4: Comparison between gridded datasets (SMCI1.0, ERA5-Land, SoMo.ml and SMAP_L4) at soil depths of (a) 10 cm, (b) 20 cm, (c) 30 cm, and (d) 40 cm. The red lines indicate the zero value for Bias and the best performance among datasets for *ubRMSE*, *R* and *MAE*.



Am-59941(109.7097°E,18.8397°N)
Bs-53464(111.1736°E,40.7272°N)
Cf-58510(116.2331°E,29.7331°N)
Bw-W1063(98.78°E,39.57°N)
Df-51431(81.3264°E,43.9406°N)
Cw-56474(102.1667°E,28.55°N)
Et-55299(92.06°E,31.48°N)
Dw-56080(102.9°E,34.99°N)
Precipitation SMCI1.0 ERA5_Land SoMo.ml SMAP_L4 in-situ SM
soil moisture(m³/m³)
precipitation(mm)
Day

Figure 5: Time series of *in-situ* and estimated SM by RF model at 10 cm soil depth along with daily precipitation in different climatic zones.

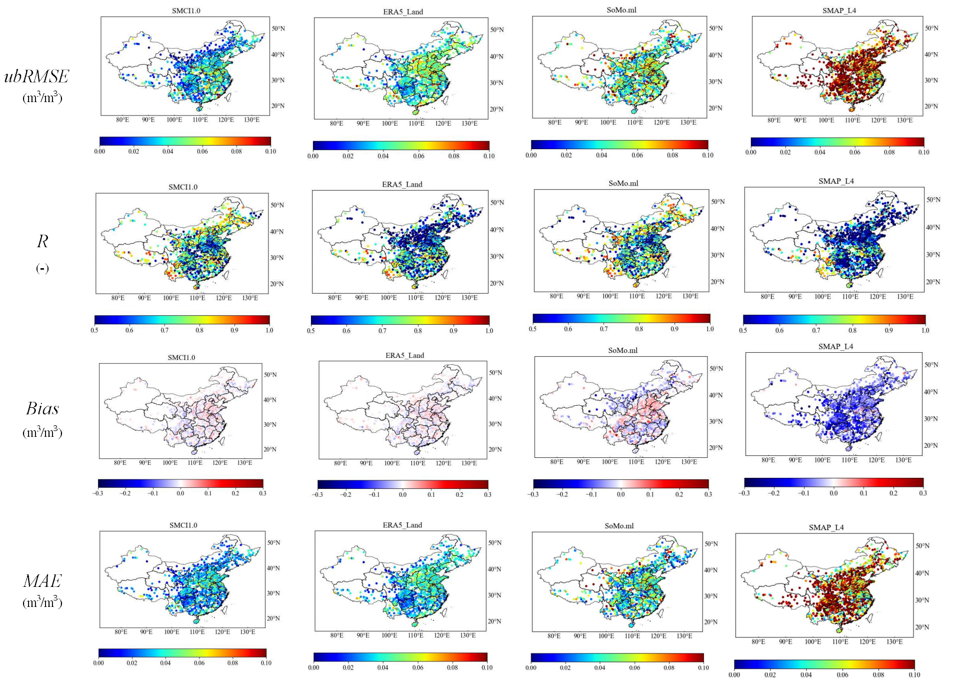
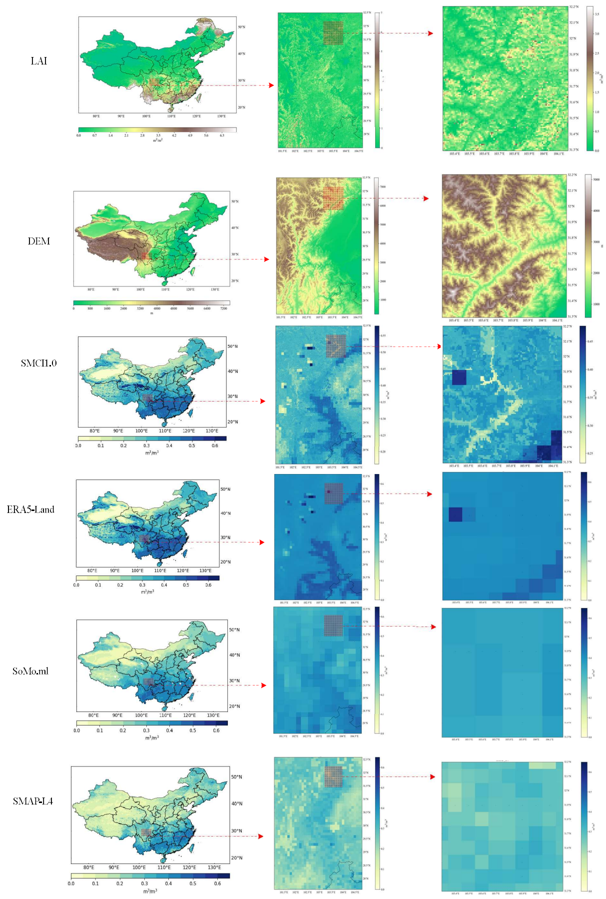


Figure 6: Goodness of fit statistics (*ubRMSE*, *R*, *Bias*, and *MAE*) at 10 cm soil depth for the RF model during the tested period.



880 **Figure 7: Soil moisture maps from different products on 1st January 2016. The resolution is 1km for SMC1.0, 9km for ERA5-Land and SMAP-L4 and 0.25 degree for SoMo.ml.**

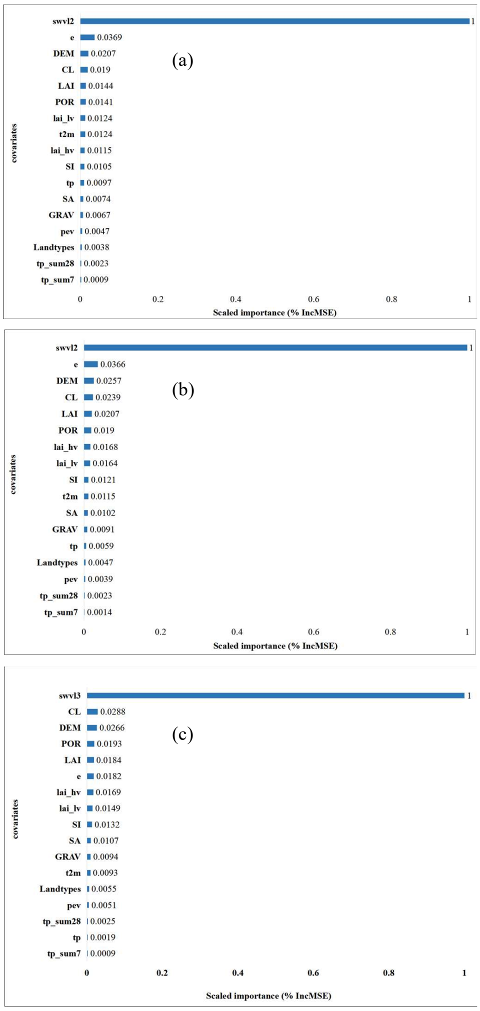

(a)
swvl2
e
DEM
CL
LAI
POR
lai_lv
t2m
lai_hv
SI
tp
SA
GRAV
pev
Landtypes
tp_sum28
tp_sum7
1
0.0369
0.0207
0.019
0.0144
0.0141
0.0124
0.0124
0.0115
0.0105
0.0097
0.0074
0.0067
0.0047
0.0038
0.0023
0.0009
covariates
0
0.2
0.4
0.6
0.8
1
Scaled importance (% IncMSE)
(b)
swvl2
e
DEM
CL
LAI
POR
lai_hv
lai_lv
SI
t2m
SA
GRAV
tp
Landtypes
pev
tp_sum28
tp_sum7
1
0.0366
0.0257
0.0239
0.0207
0.019
0.0168
0.0164
0.0121
0.0115
0.0102
0.0091
0.0059
0.0047
0.0039
0.0023
0.0014
covariates
0
0.2
0.4
0.6
0.8
1
Scaled importance (% IncMSE)
(c)
swvl3
CL
DEM
POR
LAI
e
lai_hv
lai_lv
SI
SA
GRAV
t2m
Landtypes
pev
tp_sum28
tp
tp_sum7
1
0.0288
0.0266
0.0193
0.0184
0.0182
0.0169
0.0149
0.0132
0.0107
0.0094
0.0093
0.0055
0.0051
0.0025
0.0019
0.0009
covariates
0
0.2
0.4
0.6
0.8
1
Scaled importance (% IncMSE)

Figure 8: Relative importance of ~~covariate~~predictors for the random forest (RF) model at soil depths of (a) 10 cm, (b) 20 cm, (c) 30 cm.

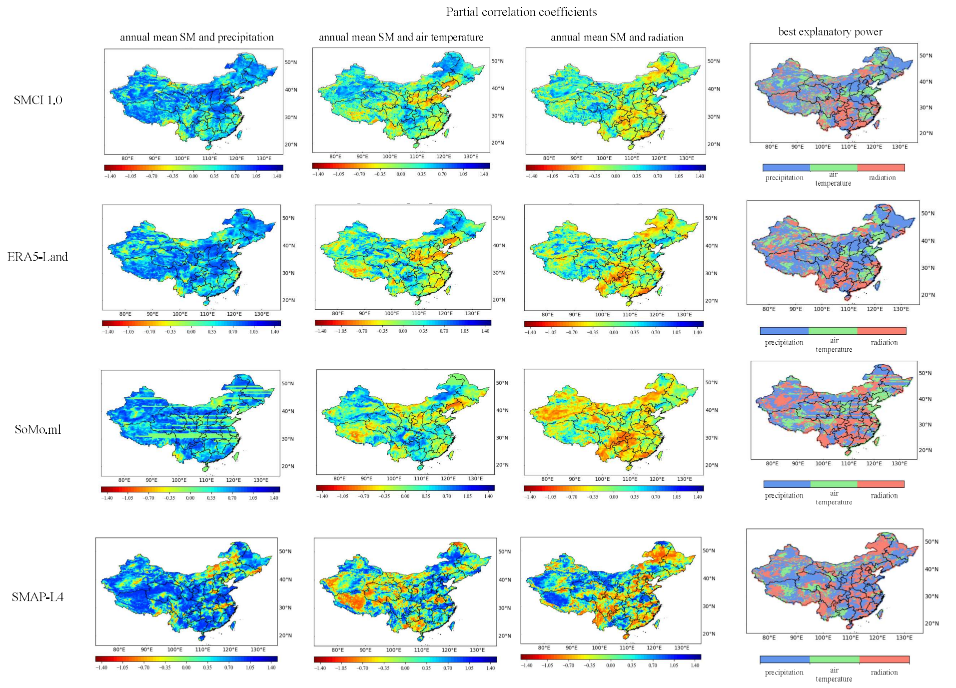


Figure 9: Partial correlation coefficients between annual mean SM and precipitation (the first column), air temperature (the second column), and radiation (the third column) for the different gridded SM products. The fourth column represents best explanatory power (highest absolute partial correlation) for the interannual variability in SM for the different gridded SM products.

Table 1. Details of the ~~covariate~~predictors for training the Random Forest model.

Source	Type	Variable (code)	Description	Time span	Spatial Resolution	Temporal Resolution
ERA5-Land (Land component of the fifth generation of European Reanalysis)	Time series	precipitation (tp) accumulated precipitation in one week (tp_sum7) accumulated precipitation in one month (tp_sum28) air temperature (t2m) potential evaporation (pev) total evaporation (e) leaf area index high vegetation (lai_hv) leaf area index low vegetation (lai_lv) soil moisture from ~~28~~ 7 to 100 cm soil depth (swvl2 to swvl3)	meteorological forcings and land surface variables	2010~2020	~9 km	hourly
CSDL (China Soil Dataset for Land surface modeling)	Static	rock fragment (GRAV) Porosity (POR) Sand, Silt, Clay (SA, SI, CL)	Soil ~~covariate~~predictors	---	~1 km	---
USGS (Unite States Geology Survey)	Static	Land cover type (Landtypes) Elevation (DEM)	Predominant land cover type and elevation	---	~1 km	---
Reprocessed MODIS LAI	Time series	Leaf area index (LAI)	Reprocessed LAI using a two-step integrated	2010~2020	~500 m	8-day

Version 6	method