Journal of Hydrology 566 (2018) 150–163

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

An advanced error correction methodology for merging in-situ observed and model-based soil moisture

Zhiyong Wu[a,b,*], Jianhong Zhou[a,b], Hai He[a,b], Qingxia Lin[a,b], Xiaotao Wu[a,b], Zhengguang Xu[a,b]

[a] Institute of Water Problems, Hohai University, Nanjing 210098, China
[b] College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China

A R T I C L E I N F O

This manuscript was handled by A. Bardossy, Editor-in-Chief, with the assistance of Saman Razavi, Associate Editor

Keywords:
VIC model
SM merging
Error correction
CDF matching
Kriging technique
OI technique

A B S T R A C T

In order to obtain an improved soil moisture (SM) dataset at large scale, an advanced SM merging methodology based on error correction methods was constructed to merge the model-based and in-situ observed SM data. The SM datasets in a 0–40 cm soil layer were derived from 10 km × 10 km Variable Infiltration Capacity (VIC) model and 797 in-situ stations, respectively. The merging methodology was conducted grid by grid, and mainly included two parts: bias correction and random error correction. Firstly, the bias correction was performed for the VIC simulations by applying the Cumulative Distribution Function (CDF) matching approach combined with the kriging technique. Secondly, the random error of the VIC simulations was corrected using an Optimal Interpolation (OI) technique based on a spatio-temporal correlation function which was proposed and constructed in this study. Through validations against in-situ observations, the merged SM was evaluated, and the merging errors in each step were analyzed and discussed. The results showed that the merged SM product was improved compared to the original SM data, both temporally and spatially. The SM merging methodology is effective and reliable in combining the accurate but sparse in-situ observations and the continuous VIC simulations. In addition, the spatial mismatch impact on the representativeness of in-situ stations was limited, and the merging errors were mainly produced in the CDF estimation process. The random error information in the spatial dimension exhibited a bigger impact on the random error correction comparing to that in the temporal dimension. This study provided strong encouragement for the efficient use of in-situ SM observations and provided valuable methods for combining multi-sources SM datasets.

1. Introduction

Soil moisture (SM) is recognized as a key variable in hydro-meteorological applications, since it controls the hydrological cycle and land-atmosphere interactions such as evapotranspiration, infiltration, and runoff (Koster et al., 2004; Seneviratne et al., 2010). SM can be measured accurately through in-situ stations, however, the measured data is inadequate in both spatial coverage and temporal frequency. Compared with the in-situ observations, the continuous SM data sources over large areas provided by satellite remote sensing are widely used in related researches (e.g., Lievens et al., 2015; Dorigo et al., 2015; Colliander et al., 2017). In particular, the microwave remote sensing SM products have great potential in SM monitoring compared to the optical remote sensing SM product which is significantly affected by weather and vegetation cover (Parrens et al., 2012). Examples of relevant microwave-based SM products are the Advanced SCATterometer (ASCAT) (Wagner et al., 2007), the Advanced Microwave Scanning Radiometer-Earth observing system (AMSR-E) (McCabe et al., 2005), the Soil Moisture and Ocean Salinity (SMOS) (Kerr, 2007), and the Soil Moisture Active Passive (SMAP) (Cai et al., 2017) product. Unfortunately, the satellite sensors can only provide SM data for the surface soil layer (0.2–5 cm) (Escorihuela et al., 2010), while the SM values of deeper soil layers are considered to exert influence in hydro-meteorological applications (such as rainfall-runoff prediction, drought monitoring, and water resource management) (Seneviratne et al., 2010; Choi et al., 2013). Moreover, the accuracy of satellite-based SM product is closely related to surface roughness, soil type, and vegetation coverage which may increase the observational uncertainties (Gruhier et al., 2010).

Another different strategy to obtain the SM data is by the application of hydrological models, based on an understanding of physical processes, and long-term meteorological observations. The SM simulations can efficiently reflect the spatio-temporal variations of SM which are observed by the in-situ stations and satellites (Du et al., 2016).

* Corresponding author at: Institute of Water Problems, Hohai University, Nanjing 210098, China.
E-mail address: wzyhhu@gmail.com (Z. Wu).

https://doi.org/10.1016/j.jhydrol.2018.09.018
Received 23 December 2017; Received in revised form 6 September 2018; Accepted 8 September 2018
Available online 10 September 2018

However, the SM simulations suffer from significant errors that are greatly influenced by the particular model structure, uncertainties of model parameters, and the meteorological forcings (Dumedah and Coulibaly, 2013; Massari et al., 2015). In general, SM data derived from above methods are not perfect and each method possesses characteristic uncertainties.

One common solution to obtain an improved SM dataset is to assimilate observations into hydrological models. Over the last couple of decades, data assimilation methods such as extended or ensemble Kalman filters (Draper et al., 2012; de Rosnay et al., 2013) were widely used for SM data assimilation (Liu et al., 2012b; Lahoz and De Lannoy, 2014). Many of these applications assimilate in-situ observed (e.g., Han et al., 2012; Gruber et al., 2018) or satellite-based (e.g., Sahoo et al., 2013; Lievens et al., 2015) SM data into hydrological models to obtain a continuous and more accurate SM dataset. Nevertheless, prior knowledge of modeling and observational uncertainties are required in data assimilation (Yilmaz et al., 2012; Yang et al., 2016), and ad-hoc error statistics are frequently used for describing the errors in assimilated observations, model parameters, model structure, and model forcings. As a result, the relative weights between the model and the observations are theoretically subjective, which may compromise the representativeness of each data source, and even further reduce the simulation accuracy (Crow and Van Loon, 2006; Maggioni et al., 2013). In addition, the assimilated SM data is still a modeled product, and it is unclear whether the assimilations retain the spatio-temporal characteristics detected by the observations (Liu et al., 2011).

An improved SM product can be also expected if different SM data sources are merged. For example, Liu et al. (2011, 2012c) firstly integrated passive (AMSR-E) and active (ASCAT) microwave satellite SM estimates for a combined SM product based on a modeled SM product. After that, the SM product was extended and improved within the Climate Change Initiative (CCI) program of European Space Agency (ESA) by combining more satellite-based SM products (Dorigo et al., 2015; Enenkel et al., 2016). The latest CCI product (ESA CCI SM v04.2, http://www.esa-soilmoisture-cci.org/dataregistration) covers a period from January 1978 to December 2016, which has a great potential for climate trend assessments. However, the CCI product is still subject to the observational depth and surface conditions, and discontinuities existed in its time series owing to the different observation systems. A more objective merging method was constructed by Yilmaz et al. (2012), who merged the model-, remote sensing-based SM products using a Triple Collocation (TC) method in a least squares framework. Nevertheless, the uncertainties estimated from the TC method for each data source are constant in time and the corrective information is not temporally propagated forward (Crow and Berg, 2010; Dorigo et al., 2010). Moreover, the merged or assimilated SM data obtained through the methods above are still biased, owing to the fact that the basic dataset such as model- or satellite-based SM are systematically different from the actual SM observations.

Generally, the errors of SM data can be divided into bias and random errors (Lahoz and De Lannoy, 2014). While biased SM data may not be a significant issue for some applications, such as some drought monitoring (e.g. Wu et al., 2011; Choi et al., 2013) and the climate change research (e.g. Dorigo and de Jeu, 2016), an accurate documentation of absolute magnitudes is critical for most applications, especially in agricultural production estimation and SM-based model calibration (Lahoz and De Lannoy, 2014). In order to obtain an improved SM product both in absolute magnitude and dynamic change, in this study, the relatively accurate but sparse in-situ SM data is selected as the observational source instead of biased satellite-based SM data, and the model-based SM data over a large area is selected as the continuous SM data source. We focused on developing an error correction method for merging in-situ observed and model-based SM data. Similar methods were typically used in the merging of meteorological variables, in which the Optimal Interpolation (OI) technique (Derber and Rosati, 1989) is supposed to be the most popular and effective method (Wang and Xie, 2007; Xie and Xiong, 2011; Pan et al., 2012). The OI technique is well-established in meteorological data assimilation to work out the optimal values only within a certain range, which is theoretically suitable for the SM merging. However, comparatively little investigation has been done on SM merging based on the OI technique, since the error characteristics of SM are quite different from that of meteorological data. Even for the OI-based SM analysis system of European Center for Medium-range Weather Forecasts (ECMWF), which is based on the assimilation of temperature and relative humidity instead of SM observations (Scipal et al., 2008).

The application of SM merging via the OI technique mainly faces two difficulties: the bias correction and the random error statistics (Xie and Xiong, 2011). Bias correction is a significant pre-requisite for the OI technique which requires the presence of unbiased data sources. In meteorological applications (e.g. Xie and Xiong, 2011; Pan et al., 2012), the bias of satellite-based or other data sources were often corrected by matching their Probability Density Function (PDF) with that of the observations using the PDF matching approach (Xie and Xiong, 2011). However, the PDF of meteorological observations are collected from surrounding stations within a large range, which is unsuitable for SM observations, since their in-situ station coverage is limited and SM values are highly variable. Moreover, SM bias correction over large scale in previous studies (e.g. Schneider et al., 2014; Kolassa et al., 2017) based on spatially continuous products such as model- and satellite-based products. It is difficult to correct the bias of simulations using the in-situ observations derived from limited amount of stations with limited observational times. For the random error statistics, SM is not only spatially correlated with meteorological variables, soil properties, and vegetation, but also temporally affected by the antecedent SM information (Maggioni et al., 2013; Penna et al., 2013). This is quite different from the meteorological variables. It is a challenge for considering the two dimensional SM errors in the OI technique.

The observational SM time series is almost normally distributed and can be parameterized using two moments: the mean value and the standard deviation (Brocca et al., 2010a). The two moments are quite stable over a large scale, and can be estimated over space using the kriging technique (Webster and Oliver, 2001). On this basis, the model-based SM data can be corrected in each grid using the Cumulative Distribution Function (CDF) matching approach (Reichle and Koster, 2004; Schneider et al., 2014), which is similar to the PDF matching but more commonly used in land surface applications. Moreover, the random error of model-based SM values are auto-correlated, both spatially and temporally, thus, a spatio-temporal correlation function was proposed for the OI technique.

Accordingly, in this study we adapted to these challenges by constructing a two-step SM merging methodology based on the CDF matching approach and the OI technique. The two SM merging sources were derived from 797 in-situ stations, and a daily Variable Infiltration Capacity (VIC) model with a resolution of 10 km × 10 km, respectively, during the study period from January 2008 to December 2016. Finally, an accurate SM product over a large scale was produced, and evaluated through various validation experiments. We have mainly achieved two innovations in the SM merging methodology: (1) The VIC simulations are corrected over large scale using the in-situ observations derived from a limited amount of stations with limited observational times; (2) Both the spatial and the temporal error characteristics are quantified in the OI technique.

2. Data and methodology

A two-step SM merging methodology was constructed in this study. To conduct the merging methodology, the Huang-Huai-Hai River Basin (3HRB) (Fig. 1) was selected as the study area, since it is a key area of agricultural production with the densest population in China (Lu et al., 2012).