

Fusion of In-Situ Soil Moisture and Land Surface Model Estimates Using Localized Ensemble Optimum Interpolation over China

Lipeng JIANG^{1,2}, Chunxiang SHI^{1,2*}, Shuai SUN², and Xiao LIANG²

¹ School of Geographic Sciences, Nanjing University of Information Science & Technology, Nanjing 210044

² National Meteorological Information Center, China Meteorological Administration, Beijing 100081

(Received February 23, 2020; in final form August 5, 2020)

ABSTRACT

Land data assimilation (DA) is an effective method to provide high-quality spatially and temporally continuous soil moisture datasets that are crucial in weather, climate, hydrological, and agricultural research. However, most existing land DA applications have used remote sensing observations, and are based on one-dimensional (1D) analysis, which cannot be directly employed to reasonably assimilate the recently expanded in-situ soil moisture observations in China. In this paper, a two-dimensional (2D) localized ensemble-based optimum interpolation (EnOI) scheme for assimilating in-situ soil moisture observations from over 2200 stations into land surface models (LSMs) is introduced. This scheme uses historical LSM simulations as ensemble samples to provide soil moisture background error covariance, allowing the in-situ observation information to be propagated to surrounding pixels. It is also computationally efficient because no additional ensemble simulations are needed. A set of ensemble sampling and localization length scale sensitivity experiments are performed. The EnOI performs best for in-situ soil moisture fusion over China with an ensemble sampling of hourly soil moisture from the previous 7 days and a localization length scale of 100 km. Following the evaluation, simulations for in-situ soil moisture fusion are also performed from May 2016 to September 2016. The EnOI analysis is notably better than that without in-situ observation fusion, as the wet bias of $0.02 \text{ m}^3 \text{ m}^{-3}$ is removed, the root-mean-square error (RMSE) is reduced by about 37%, and the correlation coefficient is increased by about 25%. Independent evaluation shows that the EnOI analysis performs considerably better than that without fusion in terms of bias, and marginally better in terms of RMSE and correlation.

Key words: soil moisture, fusion, land surface model (LSM), ensemble optimum interpolation, localization

Citation: Jiang, L. P., C. X. Shi, S. Sun, et al., 2020: Fusion of in-situ soil moisture and land surface model estimates using localized ensemble optimum interpolation over China. *J. Meteor. Res.*, **34**(6), 1335–1346, doi: 10.1007/s13351-020-0033-7.

1. Introduction

High-quality spatially and temporally continuous soil moisture datasets are urgently needed as they have important roles in weather, climate, hydrology, agriculture, and many other fields (Yeh et al., 1984; Engman, 1991; Scipal et al., 2008). There are two main ways to obtain soil moisture information: (1) in-situ measurements or satellite remote sensing observations; and (2) land surface model (LSM) simulations (Moradkhani, 2008). Remote sensing provides the ability to continuously monitor

soil moisture over large regions. Active and passive microwave measurements are the two main approaches used in soil moisture remote sensing. For example, those from the Soil Moisture Active and Passive (SMAP) mission (Entekhabi et al., 2009), the Soil Moisture and Ocean Salinity (SMOS) mission (Kerr et al., 2001), and the scatterometer and advanced scatterometer onboard the European Remote Sensing satellites (ERS-1 and ERS-2) and Meteorological Operational (MetOp) satellites, respectively (Bartalis et al., 2007; Naeimi et al., 2009). Measurements are also available from the Ad-

Supported by the China Meteorological Administration Special Public Welfare Research Fund (GYHY201506002), National Key Research and Development Program of China (2018YFC1506601), National Natural Science Foundation of China (91437220), and National Innovation Project for Meteorological Science and Technology (CMAGGTD003-5).

*Corresponding author: shicx@cma.gov.cn.

vanced Microwave Scanning Radiometer for Earth Observing System (AMSR-E; Kawanishi et al., 2003), the Advanced Microwave Scanning Radiometer 2 (AMSR2; Kim et al., 2015), the Special Sensor Microwave/Imager (Paloscia et al., 2001), and the Chinese Fengyun-3 (FY-3) satellites (Sun et al., 2014). LSMs are another important source of temporally and spatially continuous soil moisture information. Many global and regional land surface analysis or reanalysis datasets have been produced based on LSMs, such as the land surface dataset of the ECMWF's Interim Reanalysis (ERA-Interim/Land; Balsamo et al., 2015), the land surface dataset of the Modern-Era Retrospective Analysis for Research and Applications (MERRA-Land; Reichle et al., 2011), and the weakly coupled the NCEP's Global Land Data Assimilation System for the Climate Forecast System Reanalysis (CFSR/GLDAS; Meng et al., 2012).

Data assimilation (DA) is a powerful tool to combine observations with model data using a mathematical framework. To date, many operational regional and global land DA systems (LDAS) have been developed, such as the North American LDAS (Cosgrove et al., 2003; Xia et al., 2014); the Canadian LDAS, which assimilates L-band passive brightness temperature to reduce surface and root-zone soil moisture errors (Carrera et al., 2015); the China Meteorological Administration (CMA) LDAS (Shi et al., 2011); NASA's GLDAS (Rodell et al., 2004); the NCEP's GLDAS (Meng et al., 2012); and the ECMWF's GLDAS, which includes a simplified extended Kalman filter-based soil moisture DA (De Rosnay et al., 2013). Numerous studies and applications have focused on the assimilation of remotely sensed observations, including the assimilation of SMOS brightness temperature or soil moisture retrievals (Lievens et al., 2015; De Lannoy and Reichle, 2016), the assimilation of AMSR-E or AMSR-2 observations or retrievals (Yang et al., 2007; Jia et al., 2009; Tian et al., 2009), and SMAP (Draper et al., 2012; Kolassa et al., 2017). To assimilate in-situ observations directly, Gruber et al. (2018) introduced a two-dimensional (2D) Kalman filter using spatial error information provided by triple collocation techniques to assimilate spatially sparse in-situ soil moisture observations with a simplified linear model that only considered precipitation accumulation and time-independent soil moisture loss coefficients (Gruber et al., 2015, 2018). The Ensemble Kalman Filter (EnKF) is one of the most successfully applied DA methods (Evensen, 2003) among these applications.

The availability of soil moisture data from the CMA has grown considerably since the observation network was transformed in July 2013 from 10-day manual meas-

urements to hourly automatic monitoring (Wang and He, 2015). A network of more than 2200 in-situ soil moisture monitoring stations over China currently provides soil moisture observations operationally, and these in-situ observations have the potential to improve soil moisture analysis because of their more direct measurement of soil moisture than via remote sensing. However, most existing land DA applications are one-dimensional (1D) analyses, which only update the horizontally collocated LSM grid. These 1D analyses are well suited to the assimilation of remote sensing data, but they cannot be used directly for the assimilation of in-situ observations. Tools that can be used to merge in-situ measurements with LSMs are an urgent requirement to test and evaluate the potential value of these rapidly developing in-situ observations over China.

The main aim of this study was to propose a computationally efficient method to interpolate in-situ measurements to an LSM, allowing observations at each site to update the surrounding LSM grids. The ensemble-based optimum interpolation (EnOI) scheme first introduced by Evensen (2003) was used as the DA method because of its low computational cost and comparable performance to the EnKF (Blyverket et al., 2019). We determined the ensemble samples and localization length scale through a set of sensitivity experiments. In-situ soil moisture experiments with the selected ensemble samples and localization length scale from May to September 2016 were performed, followed by detailed evaluation.

The remainder of this paper is organized as follows. Section 2 describes the in-situ soil moisture observations and atmospheric forcing of LSM utilized in the analysis. Section 3 provides background on the formulation of EnOI and introduces the EnOI-based scheme for blending in-situ soil moisture observations and LSM estimates. Results are presented in Section 4 and summarized in Section 5.

2. Data

2.1 *In-situ observations of soil moisture*

The CMA's soil moisture monitoring network was transformed in July 2013 from 10-day manual measurements to hourly automatic monitoring. Currently, observations from more than 2200 stations are collected and archived by the National Meteorological Information Center (NMIC) of the CMA in real time. The observation profile of each station includes 10 vertical layers: 0–10, 10–20, 20–30, 30–40, 40–50, 50–60, 60–70, 70–80, 80–90, and 90–100 cm. The observations of the first layer (0–10 cm) were used in this study. The spatial

distribution of the observation stations is shown in Fig. 1. The observations are dense in Southeast and Central China and relatively sparse in Northwest China. Among all the stations, 230 stations (red points in Fig. 1), which are evenly distributed in space and continuously monitored in time, were selected for independent evaluation in Section 4.3.4.

2.2 Atmospheric forcing

The CMA Land DA System (CLDAS) was put into operation by the NMIC in 2013 and has since been providing real-time hourly near-surface atmospheric forcing data (e.g., air temperature, specific humidity, surface pressure, wind speed, precipitation, and radiation) and land surface products (e.g., ground temperature, soil temperature, and soil moisture) with a resolution of 0.0625° (Shi et al., 2019). The CLDAS near-surface atmospheric forcing data are generated by merging multi-source data, including ground-based observations and satellite-retrieved products, as well as numerical weather prediction model outputs. Then, the CLDAS land surface products are simulated by multiple LSMs [e.g., Community Land Model version 3.5 (CLM3.5), Common Land Model (CoLM), and Noah LSM with multiple parameterization (Noah-MP)] driven by the CLDAS near-surface atmospheric forcing data. CLDAS products have been widely used and validated in many research institutes, universities, and industries. In this study, CL-

DAS forcing data were used to drive the LSM in the experiments described below.

3. Methods

3.1 Land surface model

This study applied the community Noah-MP options (Niu et al., 2011) for soil moisture simulation. Noah-MP was designed to facilitate climate predictions with physical-based ensembles, and developed with substantial upgrades from the Noah LSM to better represent several parameters including surface-layer radiation balances, snow depth, soil moisture and heat fluxes, leaf area–rainfall interaction, vegetation and canopy temperature distinction, soil column and drainage of soil, and runoff. Multiple parameterization options are available in Noah-MP for key land–atmosphere interaction processes, such as snow, dynamic vegetation and surface water infiltration, and runoff. To better predict the climate, Noah-MP is capable of coupling the NCEP’s Global Forecasting System and Climate Forecasting System. Noah-MP contains four soil layers with thicknesses of 10, 30, 60, and 100 cm. In this paper, the default parameterization option of Noah-MP (Table 1) was used to simulate soil moisture.

To obtain a reasonable initial condition, every land model requires a spin-up period to reach the specific equilibrium state. We used the CLDAS atmospheric for-

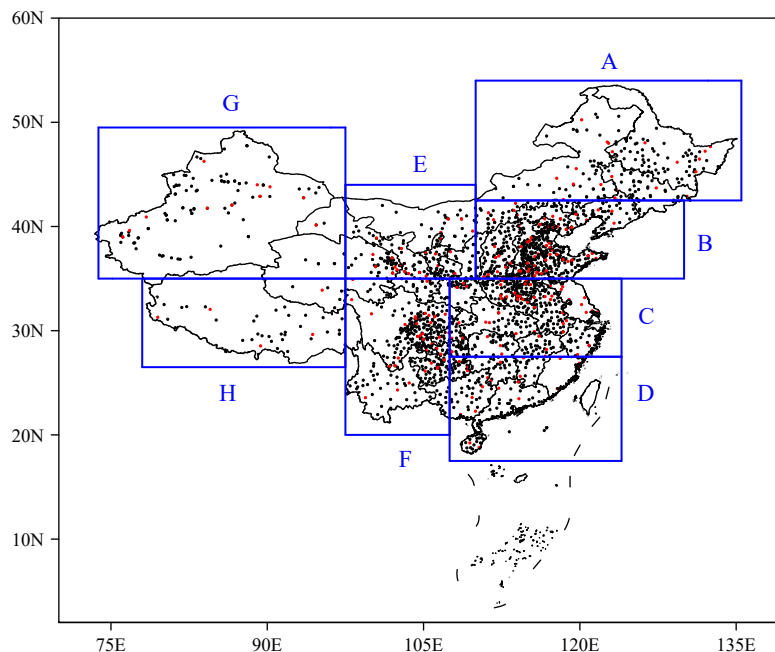


Fig. 1. Spatial distribution of automatic soil moisture observation stations for routine operation in China. A: Northeast China, B: North China, C: Jianghuai subregion, D: Southeast China, E: Inner Mongolia, F: Southwest China, G: Xinjiang subregion, and H: Tibetan Plateau. The red points are the 230 sites used for independent evaluation in Section 4.3.4.

Table 1. The Noah-MP parameterization options used in this study

Parameterization option	Physical configuration
Vegetation model option: 4	Use table leaf area index (LAI); use maximum vegetation fraction
Canopy stomatal resistance option: 1	Ball-Berry
Soil moisture factor for stomatal resistance option: 1	Noah (soil moisture)
Runoff and groundwater option: 1	TOPMODEL with groundwater (Niu et al., 2007)
Surface layer drag coefficient option: 2	Original Noah (Chen97)
Supercooled liquid water option: 1	No iteration (Niu and Yang, 2006)
Frozen soil permeability option: 1	Linear effects, more permeable (Niu and Yang, 2006)
Radiation transfer option: 1	Modified two-stream
Snow surface albedo option: 2	CLASS
Rainfall and snowfall option: 1	Jordan (1991)
Lower boundary of soil temperature option: 1	Zero heat flux from bottom
Snow and soil temperature time scheme: 1	Semi-implicit

cing described in Section 2.2 to drive a 20-yr (1998–2018) spin-up run with Noah-MP, and the values of the last time were taken as the initial conditions on 1 January 1998. The produced soil moisture simulations were used as the background states for the following in-situ soil moisture fusion experiments.

3.2 Localized EnOI system

According to the EnOI scheme proposed by Evensen (2003), the analysis (X^a) can be given as below:

$$X^a = X^b + K(Y - HX^b), \quad (1)$$

where $X^b \in \mathbb{R}^{N_m}$ is the model forecast state, $X^a \in \mathbb{R}^{N_m}$ is the analysis, N_m is the dimension of the model state vector, $Y \in \mathbb{R}^{N_y}$ is the observation vector, N_y is the number of observations, K is the gain matrix, and H is the observation operator. The gain matrix K is calculated by

$$K = \alpha(\rho \circ B)H^T [\alpha H(\rho \circ B)H^T + R]^{-1}, \quad (2)$$

where $B \in \mathbb{R}^{N_m \times N_m}$ is the ensemble-estimated background error covariance matrix; R is the observation error covariance matrix; the localized ensemble-estimated background error covariance matrix $\rho \circ B$ is the Schür product of matrices ρ and B , which is a matrix whose (i, j) entries are given by $\rho_{i,j} \cdot B_{i,j}$; and $\alpha \in (0, 1]$ is the parameter used to tune the different weights on the ensemble versus observations. The ensemble-estimated background error covariance is estimated from the equation

$$B = \frac{A' A'^T}{N-1}, \quad (3)$$

where $A' = [A'^1, A'^2, \dots, A'^N]$, N is the number of ensemble samples, and the k th element of A' is calculated by

$$A'^k = \left(X^k - \frac{1}{N-1} \sum_{i=1}^N X^i \right). \quad (4)$$

In the EnOI scheme, a relatively stationary ensemble of model state samples can be taken from a long-term en-

semble of model perturbations (anomalies) generated from a long-term model run (Evensen, 2003). Without the need for an ensemble forecast, the EnOI scheme can typically save N times the computational cost than the EnKF. In fact, many previous studies have employed similar historical ensemble methods to simplify the ensemble generation procedure in the assimilation. For example, Pan et al. (2009) used downscaled forcing ensemble forecasts from the NOAA/NCEP Climate Forecast System (CFS) as the input forcing ensembles in their hydrological assimilation system. Pan and Wood (2009) proposed a pattern-based sampling approach in which random samples were drawn from a historical rainfall database according to the pattern of the satellite rainfall, and Pan and Wood (2010) directly used the rainfall data from the Tropical Rainfall Measuring Mission (TRMM) satellite products as the rainfall ensembles to force their assimilation experiments. The selection of ensemble samples in this study is described in Section 4.2.

Another critical question in ensemble-based DA is the localization technique, which is a widely used solution to reduce sampling error, especially when the ensemble size is small (Hamill et al., 2001; Oke et al., 2007). We used the following fifth-order piecewise rational function (Gaspari and Cohn, 1999) to construct the localization matrix ρ :

$$\rho(i, j) = C_0(d_{i,j}/d), \quad (5)$$

where C_0 is defined as

$$C_0(I) = \begin{cases} -\frac{1}{4}I^5 + \frac{1}{2}I^4 + \frac{5}{8}I^3 - \frac{5}{3}I^2 + 1, & 0 \leq I \leq 1, \\ \frac{1}{12}I^5 - \frac{1}{2}I^4 + \frac{5}{8}I^3 + \frac{5}{3}I^2 - 5I + 4 - \frac{2}{3}I^{-1}, & 1 < I \leq 2, \\ 0, & 2 < I, \end{cases} \quad (6)$$

where $I = d_{i,j}/d$, in which d is the localization length scale and $d_{i,j}$ is the horizontal spatial distance between the i th and j th grid points. The localization length scale d

indicates the significance range of a measurement.

3.3 Evaluation methods

The evaluation criteria used in this study were the bias (Bias), root-mean-square error (RMSE), and correlation coefficient (Corr), which are calculated as follows:

$$\text{Bias} = \frac{1}{N-1} \sum_{i=1}^N (M_i - O_i), \tag{7}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (M_i - O_i)^2}{N-1}}, \tag{8}$$

$$\text{Corr} = \frac{\sum_{i=1}^N (M_i - \bar{M})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^N (M_i - \bar{M})^2} \sqrt{\sum_{i=1}^N (O_i - \bar{O})^2}}, \tag{9}$$

$$\bar{O} = \frac{1}{N-1} \sum_{i=1}^N O_i, \tag{10}$$

$$\bar{M} = \frac{1}{N-1} \sum_{i=1}^N M_i, \tag{11}$$

where M is the simulated (merged) soil moisture to be evaluated, O represents the in-situ soil moisture observations used for the evaluation, N is the number of observations, O_i is the i th observation, M_i is the simulated (merged) soil moisture collocated with the i th observation, \bar{O} is the average value of all observations used for the evaluation, and \bar{M} is the average value of simulated (merged) soil moisture at all the collocated locations.

4. Results

4.1 Optimal localization length scale for in-situ soil moisture fusion over China

The localization technique is useful for the reduction

of ensemble sampling error. One of the most important parameters for localization is the length scale. We began by analyzing the characteristics of soil moisture spatial correlation using in-situ observations over China. Then, a series of experiments with different localization length scales were performed for determining the optimal localization length scale.

4.1.1 Spatial correlation analysis based on the in-situ observation network in China

Using in-situ hourly observations from 1185 sites in July 2016, of which the observations are time-continuous and valid, we examined the characteristics of soil moisture spatial correlation over China. First, the nearest neighboring site for each site was identified, and then the spatial distance and correlation coefficient between each site and its nearest neighboring site were calculated.

As shown in Fig. 2, the number of stations within distances of 0–10 and 0–30 km from their nearest neighboring site is 215 and 683, respectively, and the distances between most stations (95.6%) and their nearest neighbors are within 100 km. The number of stations with correlations of 0.6–0.8 and 0.8–1 is 343 and 420, respectively, and 78.8% of the observation stations have correlations above 0.5 with their nearest neighboring site. Figure 3 shows the spatial distribution of the correlation coefficients between each site and their nearest neighboring site. Most of the correlations are higher than 0.4 in regions with dense observations, such as North, South, Southwest, and Central China. In contrast, in regions where the sites are spatially sparse, such as Inner Mongolia, Gansu, Tibet, and Xinjiang, the correlations are relatively weaker.

4.1.2 Localization length scale

Based on the above knowledge about the spatial distances and correlations of the in-situ soil moisture moni-

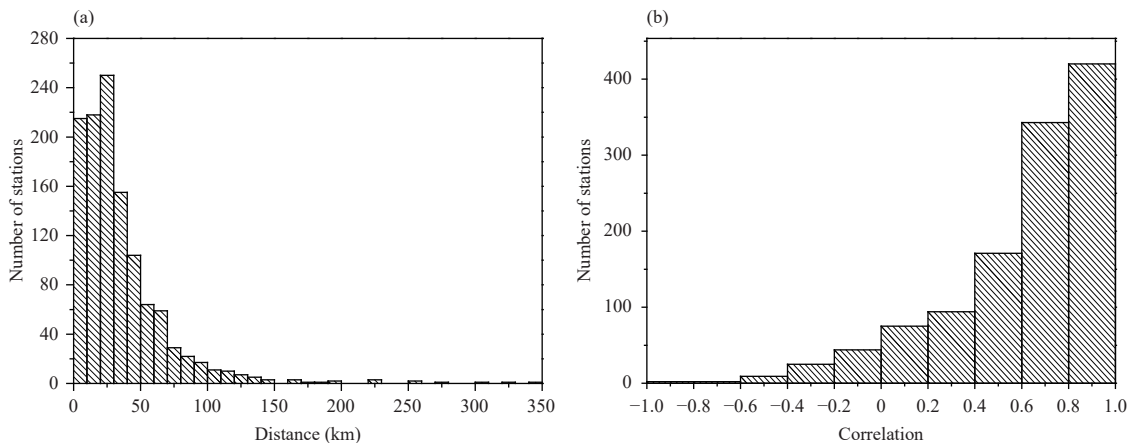


Fig. 2. Histograms of the (a) distances and (b) correlation coefficients between each site and its nearest neighboring site.

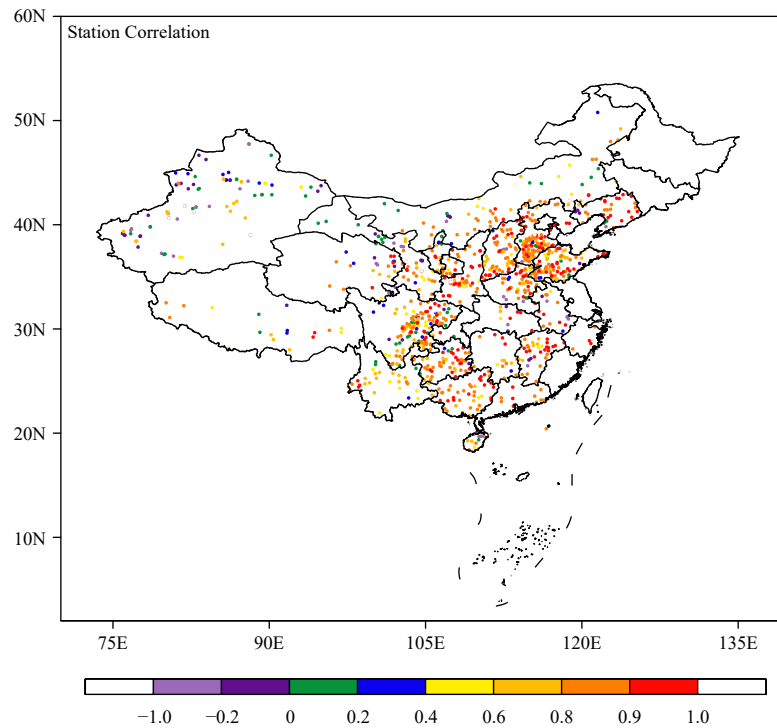


Fig. 3. Spatial distribution of correlation coefficients between each site and its nearest neighboring site.

oring network over China, a set of experiments with different localization length scales, including 200, 150, 100, 80, 50, 30, and 10 km, were performed, while the other configurations were kept the same (observation error $R = 0.01 \text{ m}^3 \text{ m}^{-3}$; ensemble samples from the previous 7-day hourly LSM simulations; $\alpha = 1$). In these experiments, the Noah-MP-simulated soil moisture was merged with the in-situ soil moisture observations during 1–31 July 2016, among which the abovementioned 230 sites were left for validation and the others were used in the fusion experiments. The fusion results were interpolated to the 230 sites and validated against the observations.

Figure 4 shows boxplots of the bias, RMSE, and correlations of the Noah-MP simulations (open loop) and fusion results under different localization length scales, validated against the 230 in-situ observations. The maximum values (the upper bound), minimum values (the lower bound), mean values (the horizontal line), median values (the smallest rectangles), and the outliers (the asterisks) are given in the boxplots. For the bias, the simulated soil moisture shows systematically positive biases between 0.017 and $0.035 \text{ m}^3 \text{ m}^{-3}$, whereas the merged soil moisture performs better with lower bias values. Furthermore, the biases of the merged soil moisture decrease with the increase in localization length scale and reach their lowest value with the length scale of 100 km. In terms of RMSE, the RMSEs of the simulated soil

moisture are between 0.063 and $0.0775 \text{ m}^3 \text{ m}^{-3}$, and the RMSEs of the merged soil moisture reduce as the localization length scale increases, again reaching their lowest values with the scale of 100 km. Regarding the correlation coefficients, the correlations of the simulated soil moisture and merged moisture under different localization length scales are stable with values of around 0.7. Relatively speaking, the merged soil moisture with localization length scales of 80, 100, 150, and 200 km have higher correlations. In conclusion, the localization length scale of 100 km performs best for in-situ soil moisture fusion over China.

4.2 Impact of ensemble sampling on soil moisture fusion

Reasonable construction of ensemble members is crucial for accurate estimation of background error covariance. A series of soil moisture fusion experiments covering the period 1–31 July 2016 were implemented using different historical ensemble samples, including the hourly samples from the previous 7 days ($7 \times 24 = 168$ members, marked as “168_en”), the previous 5 days ($5 \times 24 = 120$ members, marked as “120_en”), the previous 3 days ($3 \times 24 = 72$ members, marked as “72_en”), the previous 1 day ($1 \times 24 = 24$ members, marked as “24_en”), the same Julian day during 1998–2015 ($18 \times 24 = 432$ members, marked as “432_en”), and from the same hour and Julian day during 1998–2015 ($18 \times 1 = 18$ members,

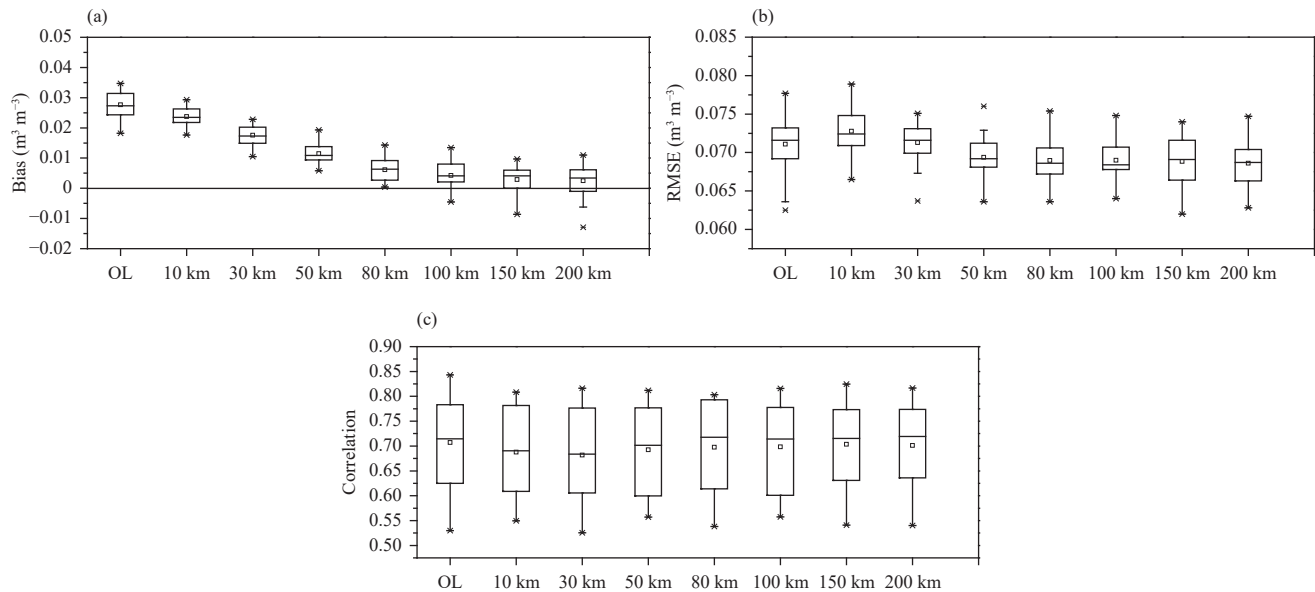


Fig. 4. Boxplots of (a) bias, (b) RMSE, and (c) correlations of soil moisture simulations and fusion results under different localization length scales.

marked as “18_en”). Aside from the ensemble sampling, the other configurations remained the same ($R = 0.01 \text{ m}^3 \text{m}^{-3}$, $d = 100 \text{ km}$, and $\alpha = 1$). The observations from 230 sites (red points in Fig. 1) were retained for independently validating the fusion results, and observations from the other sites were used for merging with the Noah-MP simulated soil moisture in the fusion experiments.

Boxplots of the bias, RMSE, and correlations of the merged soil moisture using different ensemble samples are presented in Fig. 5. In terms of the bias, the merged soil moisture using hourly samples from the previous 1 day (24_en in Fig. 5) has the largest bias, while that using hourly samples from the previous 7 days (168_en in Fig. 5) has the lowest bias. In terms of the RMSE, the experiment using hourly samples from the same hour and Julian day during 1998–2015 (18_en in Fig. 5) performs the worst, while the experiment collecting hourly samples from the previous 7 days (168_en in Fig. 5) performs the best. The correlation coefficients of the merged soil moisture are near 0.7 and vary little under different combinations of ensemble members. The experiments using hourly samples from the previous 5 days (120_en in Fig. 5) and from the previous 7 days (168_en in Fig. 5) have higher correlations. In summary, the experiment with ensemble samples from the previous 7-day hourly simulations performs best.

4.3 Soil moisture fusion experiments

According to the results of the previous sensitivity experiments, three further experiments were designed to evaluate the value of in-situ soil moisture observation fu-

sion. The first was a Noah-MP open loop experiment (marked OL hereafter), which did not use in-situ observations. The second experiment (marked ANL-2200) merged in-situ observations from 2200 sites (red and black dots in Fig. 1) with the output of OL, while the third (marked ANL-1970) merged in-situ observations from only 1970 sites (black dots in Fig. 1), also with the output of OL. In the ANL-2200 and ANL-1970 experiments, we took the previous 7-day hourly Noah-MP simulations as ensemble samples for calculating the background error covariance (number of ensemble members: 168). The horizontal localization length scale was set as 100 km. The experiments were performed from 1 May to 30 September 2016. Observations from 230 stations (red dots in Fig. 1) were used for evaluation.

4.3.1 Spatial distributions of OL versus ANL-2200

Monthly means of OL and ANL-2200 were calculated for each month from May to September in 2016. Figure 6 shows that both OL and ANL-2200 reflect the rational soil moisture distribution over China (humid in Southeast China and dry in Northwest China and Inner Mongolia) in July 2016. ANL-2200 provides more detailed information and is drier than OL in Inner Mongolia, Sichuan Basin, and South China. ANL-2200 and OL are quite similar over the Tibetan Plateau and West China, where in-situ observations are sparse.

4.3.2 Comparison of daily statistics over China

Daily average soil moisture values based on OL and ANL-2200 were calculated and then evaluated against the daily mean of the hourly observations used in EnOI.

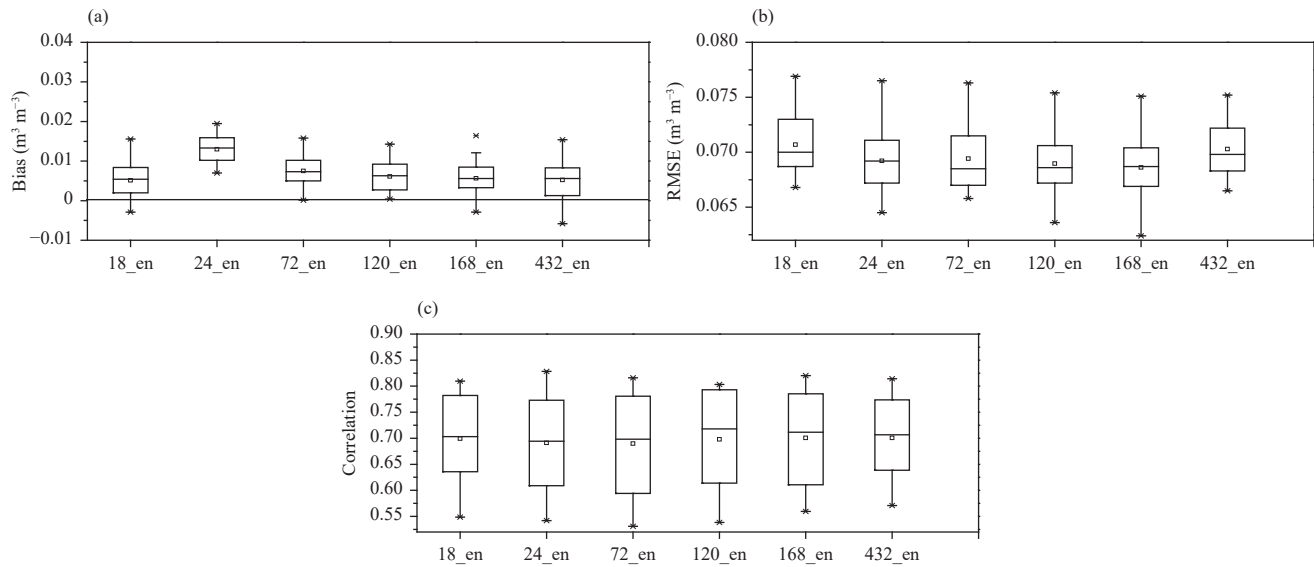


Fig. 5. Boxplots of soil moisture fusion errors in terms of (a) bias, (b) RMSE, and (c) correlations, using different ensemble samples.

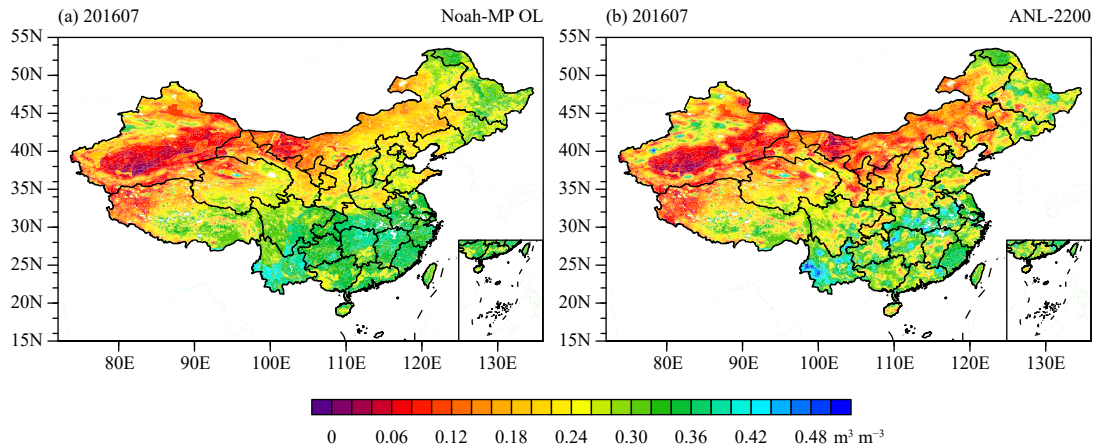


Fig. 6. Monthly means of the (a) Noah-MP open-loop and (b) ANL-2200 experiments for July 2016.

Figure 7 shows the bias of ANL-2200 is closer to zero ($0.0002 \text{ m}^3 \text{ m}^{-3}$), while OL has a notable wet bias from 0.01 to $0.035 \text{ m}^3 \text{ m}^{-3}$. The RMSE of ANL-2200 (0.035 – $0.055 \text{ m}^3 \text{ m}^{-3}$) is smaller than that of OL (0.065 – $0.075 \text{ m}^3 \text{ m}^{-3}$). ANL-2200 also has a higher correlation (around 0.9) than OL (0.5–0.8). Overall, the EnOI analysis is notably better than that without in-situ observation fusion, as the wet bias of $0.02 \text{ m}^3 \text{ m}^{-3}$ is removed, the RMSE is reduced by about 37% (0.071 – $0.045 \text{ m}^3 \text{ m}^{-3}$), and the correlation is increased by about 25% (0.71 – 0.89).

4.3.3 Comparison of statistics over different subregions

We also calculated the bias, RMSE, and correlation against observation used in EnOI from May to September 2016 over eight subregions (Fig. 1), including Northeast China, North China, the JiangHuai subregion, Southeast China, Inner Mongolia, Southwest China, the Xinji-

ang subregion, and the Tibetan Plateau, following the subregion definition of Ma et al. (2005). Figure 8 shows that the bias and RMSE of ANL-2200 (red) are significantly reduced compared with OL (blue) over all subregions. The biases of ANL-2200 over all subregions are within $\pm 0.01 \text{ m}^3 \text{ m}^{-3}$, while they can reach $0.07 \text{ m}^3 \text{ m}^{-3}$ for the Tibetan Plateau and $0.04 \text{ m}^3 \text{ m}^{-3}$ for Inner Mongolia. The RMSEs of ANL-2200 over all subregions are between 0.023 and $0.05 \text{ m}^3 \text{ m}^{-3}$, while most subregions have values exceeding $0.07 \text{ m}^3 \text{ m}^{-3}$ for the OL simulations. The correlations are notably increased for all subregions after EnOI fusion.

4.3.4 Independent evaluation

As described in Section 2.1, 230 stations that are evenly distributed in space were selected for independent evaluation. Figure 9 shows the averages of soil mois-

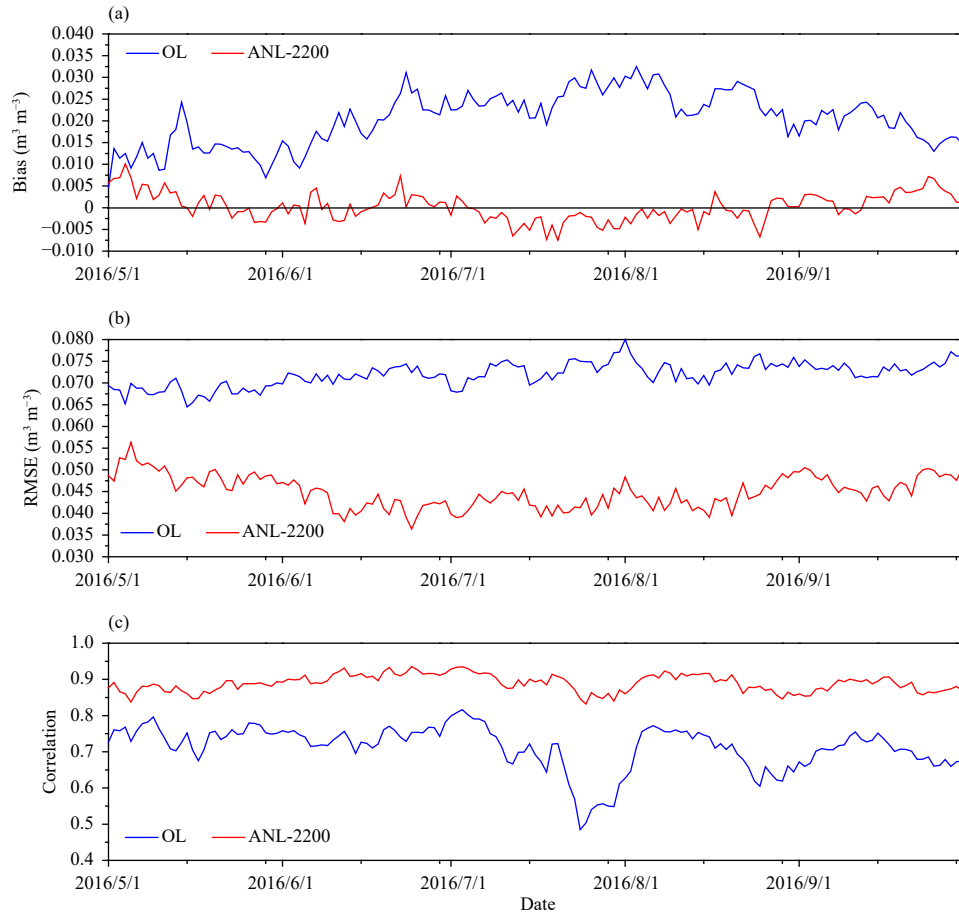


Fig. 7. Time series of daily statistics of (a) bias, (b) RMSE, and (c) correlation over China for the OL (blue) and ANL-2200 (red) experiments from 1 May to 30 September 2016.

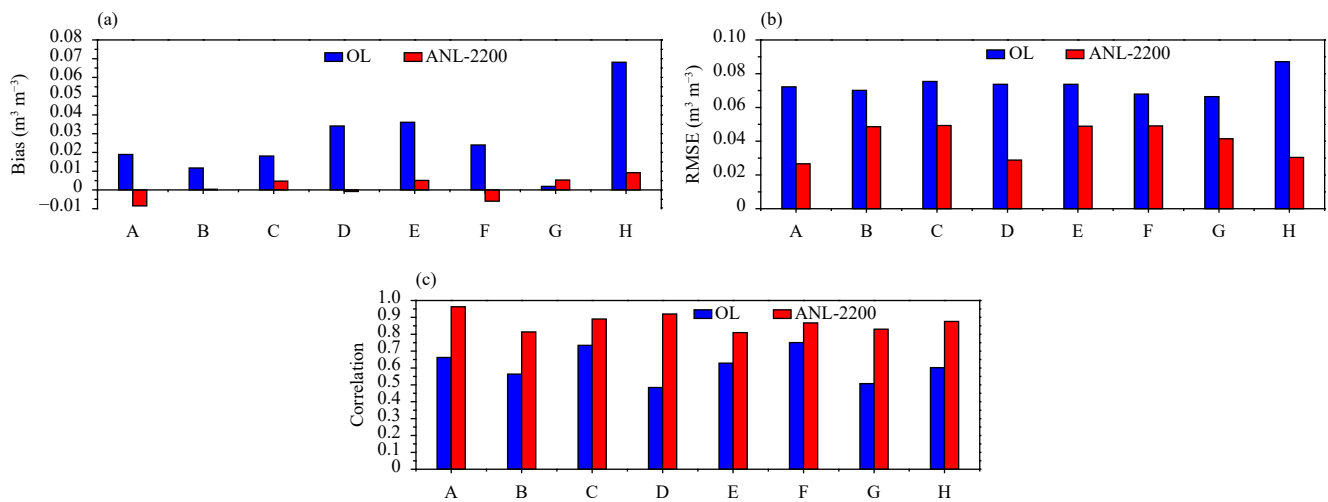


Fig. 8. (a) Bias, (b) RMSE, and (c) correlation of the OL and ANL-2200 experiments over eight subregions. A: Northeast China (21 stations), B: North China (63 stations), C: Jianghuai subregion (53 stations), D: Southeast China (20 stations), E: Inner Mongolia (23 stations), F: Southwest China (34 stations), G: Xinjiang subregion (11 stations), and H: Tibetan Plateau (5 stations).

ture from observations and the OL and ANL-1970 experiments at these 230 stations. The temporal trends of OL

and ANL-1970 are quite consistent with those from the observations, while OL has a clear wet bias and ANL-

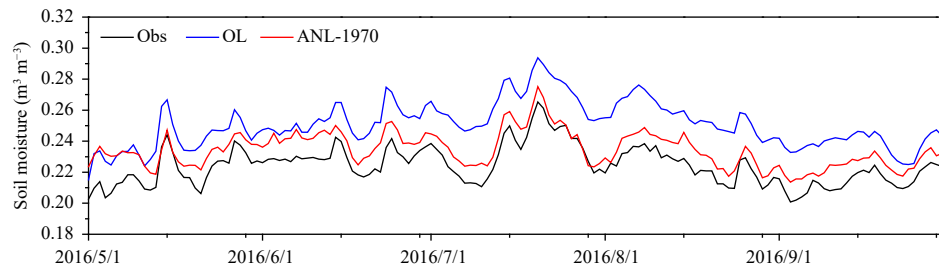


Fig. 9. Soil moisture time series based on observations (black) and the OL (blue) and ANL-1970 (red) experiments at the collated 230 sites not used in EnOI.

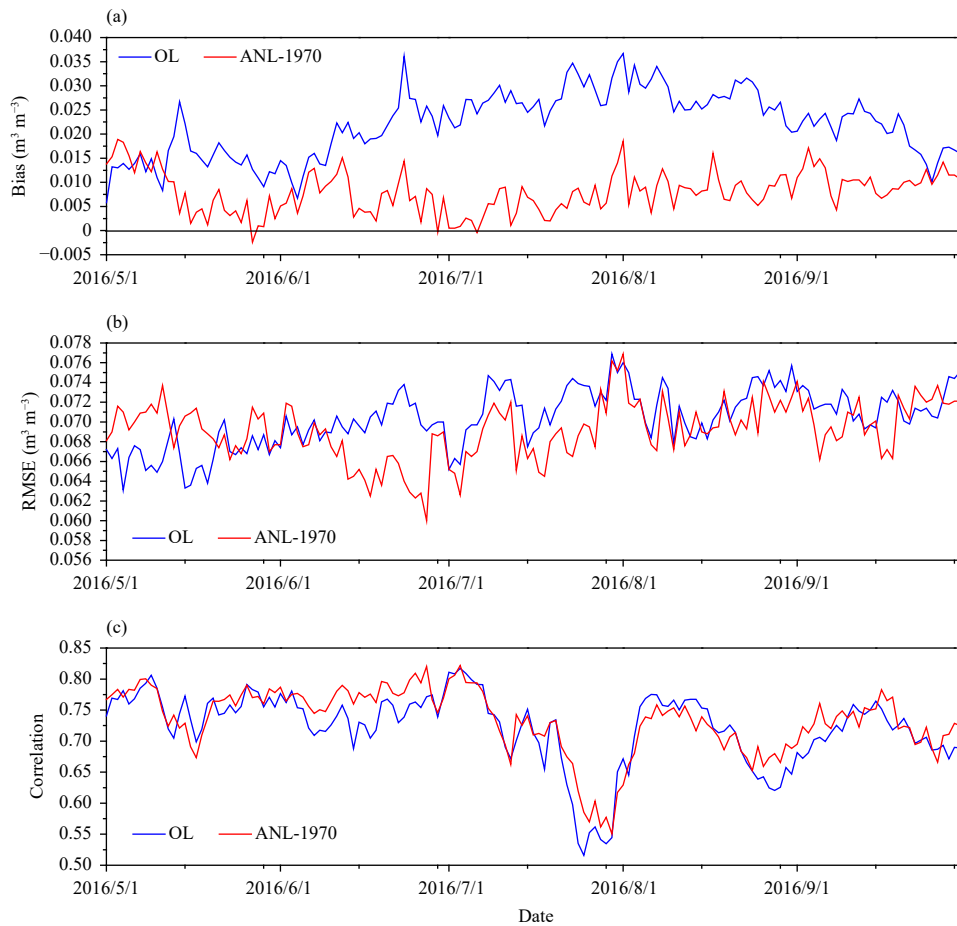


Fig. 10. Time series of daily (a) bias, (b) RMSE, and (c) correlation at 230 independent evaluation stations for the OL (blue) and ANL-1970 (red) experiments from 1 May to 30 September 2016.

1970 is much closer to the observations.

Figure 10 shows the time series of daily statistics at the 230 stations not used in EnOI for the OL and ANL-1970 experiments from 1 May to 30 September 2016. As with the evaluation results reported in Section 4.3.2, the bias of ANL-1970 is reduced by about 60% (from 0.02 to $0.008 \text{ m}^3 \text{ m}^{-3}$). The RMSE of ANL-1970 ($0.069 \text{ m}^3 \text{ m}^{-3}$) is also smaller than that of OL ($0.071 \text{ m}^3 \text{ m}^{-3}$), and the correlation of ANL-1970 (0.73) is marginally higher than for OL (0.71). Overall, the independent evaluation res-

ults show that ANL-1970 is considerably better than OL in terms of bias, and performs marginally better with respect to RMSE and correlation.

5. Summary

Given the recent rapid development of the soil moisture in-situ measurement network in China, there is no doubt that these observations can and should be used increasingly more widely, not only for calibration and val-

idation purposes but also for direct fusion into soil moisture products. In this context, the present study introduces an EnOI-based 2D soil moisture analysis scheme that allows in-situ observations at each site to update the surrounding LSM grids. The scheme uses a relatively stationary ensemble of model state samples to calculate the background error covariance matrix, which makes it very inexpensive computationally. A set of ensemble sampling and localization length scale sensitivity experiments were performed, and the results show that the EnOI scheme with ensemble sampling from the previous 7 days of hourly soil moisture states and a localization length scale of 100 km performs best for in-situ soil moisture fusion over China.

In-situ soil moisture fusion experiments were then performed from May to September 2016. The spatial distributions of monthly mean soil moisture over China with and without EnOI fusion are similar and reasonable. The EnOI analysis shows more detailed information in subregions such as Inner Mongolia, the Sichuan basin, and South China, where observations are denser. Evaluation against observations used in EnOI shows that the EnOI analysis is notably better than that without in-situ observation fusion for all the employed statistical metrics (bias, RMSE, and correlation). Independent evaluation shows that the EnOI analysis performs considerably better for bias, and marginally better for RMSE and correlation.

REFERENCES

- Balsamo, G., C. Albergel, A. Beljaars, et al., 2015: ERA-Interim/Land: A global land surface reanalysis data set. *Hydrol. Earth Syst. Sci.*, **19**, 389–407, doi: 10.5194/hess-19-389-2015.
- Bartalis, Z., W. Wagner, V. Naeimi, et al., 2007: Initial soil moisture retrievals from the METOP-A advanced scatterometer (ASCAT). *Geophys. Res. Lett.*, **34**, L20401, doi: 10.1029/2007GL031088.
- Blyverket, J., P. D. Hamer, L. Bertino, et al., 2019: An evaluation of the EnKF vs. EnOI and the assimilation of SMAP, SMOS and ESA CCI soil moisture data over the contiguous US. *Remote Sens.*, **11**, 478, doi: 10.3390/rs11050478.
- Carrera, M. L., S. Bélair, and B. Bilodeau, 2015: The Canadian land data assimilation system (CaLDAS): Description and synthetic evaluation study. *J. Hydrometeorol.*, **16**, 1293–1314, doi: 10.1175/JHM-D-14-0089.1.
- Cosgrove, B. A., D. Lohmann, K. E. Mitchell, et al., 2003: Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *J. Geophys. Res. Atmos.*, **108**, 8842, doi: 10.1029/2002JD003118.
- De Rosnay, P., M. Drusch, D. Vasiljevic, et al., 2013: A simplified Extended Kalman Filter for the global operational soil moisture analysis at ECMWF. *Quart. J. Roy. Meteor. Soc.*, **139**, 1199–1213, doi: 10.1002/qj.2023.
- De Lannoy, G. J. M., and R. H. Reichle, 2016: Assimilation of SMOS brightness temperatures or soil moisture retrievals into a land surface model. *Hydrol. Earth Syst. Sci.*, **20**, 4895–4911, doi: 10.5194/hess-20-4895-2016.
- Draper, C. S., R. H. Reichle, G. J. M. De Lannoy, et al., 2012: Assimilation of passive and active microwave soil moisture retrievals. *Geophys. Res. Lett.*, **39**, L04401, doi: 10.1029/2011GL050655.
- Engman, E. T., 1991: Applications of microwave remote sensing of soil moisture for water resources and agriculture. *Remote Sens. Environ.*, **35**, 213–226, doi: 10.1016/0034-4257(91)90013-V.
- Entekhabi, D., E. Njoku, and P. O'Neill, 2009: The Soil Moisture Active and Passive mission (SMAP): Science and applications. Proc. 2009 IEEE Radar Conference, IEEE, Pasadena, CA, USA, 1–3, doi: 10.1109/RADAR.2009.4977030.
- Evensen, G., 2003: The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367, doi: 10.1007/s10236-003-0036-9.
- Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757, doi: 10.1002/qj.49712555417.
- Gruber, A., W. Crow, W. Dorigo, et al., 2015: The potential of 2D Kalman filtering for soil moisture data assimilation. *Remote Sens. Environ.*, **171**, 137–148, doi: 10.1016/j.rse.2015.10.019.
- Gruber, A., W. T. Crow, and W. A. Dorigo, 2018: Assimilation of spatially sparse in situ soil moisture networks into a continuous model domain. *Water Resour. Res.*, **54**, 1353–1367, doi: 10.1002/2017WR021277.
- Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 2776–2790, doi: 10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2.
- Jia, B. H., Z. H. Xie, X. J. Tian, et al., 2009: A soil moisture assimilation scheme based on the ensemble Kalman filter using microwave brightness temperature. *Sci. China Ser. D: Earth Sci.*, **52**, 1835, doi: 10.1007/s11430-009-0122-z.
- Jordan R., 1991: A one-dimensional temperature model for a snow cover: Technical documentation for SNTherm. 89 (No. CRREL-SR-91-16). Cold Regions Research and Engineering Lab Hanover NH, 61 pp.
- Kawanishi, T., T. Sezai, Y. Ito, et al., 2003: The Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E), NASDA's contribution to the EOS for global energy and water cycle studies. *IEEE Trans. Geosci. Remote Sens.*, **41**, 184–194, doi: 10.1109/TGRS.2002.808331.
- Kerr, Y. H., P. Waldteufel, J. P. Wigneron, et al., 2001: Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission. *IEEE Trans. Geosci. Remote Sens.*, **39**, 1729–1735, doi: 10.1109/36.942551.
- Kim, S., Y. Y. Liu, F. M. Johnson, et al., 2015: A global comparison of alternate AMSR2 soil moisture products: Why do they differ? *Remote Sens. Environ.*, **161**, 43–62, doi: 10.1016/j.rse.2015.02.002.
- Kolassa, J., R. H. Reichle, Q. Liu, et al., 2017: Data assimilation to extract soil moisture information from SMAP observations. *Remote Sens.*, **9**, 1179, doi: 10.3390/rs9111179.
- Lievens, H., S. K. Tomer, A. Al Bitar, et al., 2015: SMOS soil

- moisture assimilation for improved hydrologic simulation in the Murray Darling Basin, Australia. *Remote Sens. Environ.*, **168**, 146–162, doi: 10.1016/j.rse.2015.06.025.
- Ma, Z. G., G. Huang, W. Q. Gan, et al., 2005: Multi-scale temporal characteristics of the dryness/wetness over northern China during the last century. *Chinese J. Atmos. Sci.*, **29**, 671–681. (in Chinese)
- Meng, J., R. Q. Yang, H. L. Wei, et al., 2012: The land surface analysis in the NCEP Climate Forecast System Reanalysis. *J. Hydrometeor.*, **13**, 1621–1630, doi: 10.1175/JHM-D-11-090.1.
- Moradkhani, H., 2008: Hydrologic remote sensing and land surface data assimilation. *Sensors*, **8**, 2986–3004, doi: 10.3390/s8052986.
- Naeimi, V., K. Scipal, Z. Bartalis, et al., 2009: An improved soil moisture retrieval algorithm for ERS and METOP scatterometer observations. *IEEE Trans. Geosci. Remote Sens.*, **47**, 1999–2013, doi: 10.1109/TGRS.2008.2011617.
- Niu, G. Y., and Z. L. Yang, 2006: Effects of frozen soil on snowmelt runoff and soil water storage at a continental scale. *J. Hydrometeor.*, **7**, 937–952, doi: doi.org/10.1175/JHM538.1.
- Niu, G. Y., Z. L. Yang, R. E. Dickinson, et al., 2007: Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data. *J. Geophys. Res. Atmos.*, **112** doi: doi:10.1029/2006JD007522.
- Niu, G. Y., Z. L. Yang, K. E. Mitchell, et al., 2011: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *J. Geophys. Res. Atmos.*, **116**, D12109, doi: 10.1029/2010JD015139.
- Oke, P. R., P. Sakov, and S. P. Corney, 2007: Impacts of localisation in the EnKF and EnOI: Experiments with a small model. *Ocean Dyn.*, **57**, 32–45, doi: 10.1007/s10236-006-0088-8.
- Paloscia, S., G. Macelloni, E. Santi, et al., 2001: A multifrequency algorithm for the retrieval of soil moisture on a large scale using microwave data from SMMR and SSM/I satellites. *IEEE Trans. Geosci. Remote Sens.*, **39**, 1655–1661, doi: 10.1109/36.942543.
- Pan, M., and E. F. Wood, 2009: A multiscale ensemble filtering system for hydrologic data assimilation. Part II: Application to land surface modeling with satellite rainfall forcing. *J. Hydrometeor.*, **10**, 1493–1506, doi: 10.1175/2009JHM1155.1.
- Pan, M., and E. F. Wood, 2010: Impact of accuracy, spatial availability, and revisit time of satellite-derived surface soil moisture in a multiscale ensemble data assimilation system. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.*, **3**, 49–56, doi: 10.1109/JSTARS.2010.2040585.
- Pan, M., E. F. Wood, D. B. McLaughlin, et al., 2009: A multiscale ensemble filtering system for hydrologic data assimilation. Part I: Implementation and synthetic experiment. *J. Hydrometeor.*, **10**, 794–806, doi: 10.1175/2009JHM1088.1.
- Reichle, R. H., R. D. Koster, G. J. M. De Lannoy, et al., 2011: Assessment and enhancement of MERRA land surface hydrology estimates. *J. Climate*, **24**, 6322–6338, doi: 10.1175/JCLI-D-10-05033.1.
- Rodell, M., P. R. Houser, U. Jambor, et al., 2004: The global land data assimilation system. *Bull. Amer. Meteor. Soc.*, **85**, 381–394, doi: 10.1175/BAMS-85-3-381.
- Scipal, K., M. Drusch, and W. Wagner, 2008: Assimilation of a ERS scatterometer derived soil moisture index in the ECMWF numerical weather prediction system. *Adv. Water Res.*, **31**, 1101–1112, doi: 10.1016/j.advwatres.2008.04.013.
- Shi, C. X., Z. H. Xie, H. Qian, et al., 2011: China land soil moisture EnKF data assimilation based on satellite remote sensing data. *Sci. China Earth Sci.*, **54**, 1430–1440, doi: 10.1007/s11430-010-4160-3.
- Shi, C. X., Y. Pan, J. X. Gu, et al., 2019: A review of multi-source meteorological data fusion products. *Acta Meteor. Sinica*, **77**, 774–783. (in Chinese)
- Sun, R. J., Y. P. Zhang, S. L. Wu, et al., 2014: The FY-3B/MWRI soil moisture product and its application in drought monitoring. Proc. 2014 IEEE Geoscience and Remote Sensing Symposium, IEEE, Quebec, Canada, doi: 10.1109/IGARSS.2014.6947184.
- Tian, X. J., Z. H. Xie, A. G. Dai, et al., 2009: A dual-pass variational data assimilation framework for estimating soil moisture profiles from AMSR-E microwave brightness temperature. *J. Geophys. Res. Atmos.*, **114**, D16102, doi: 10.1029/2008JD011600.
- Wang, L. Y., and Y. B. He, 2015: Research on outlier threshold of automatic soil moisture observation data. *Meteor. Mon.*, **41**, 1017–1022, doi: 10.7519/j.issn.1000-0526.2015.08.011. (in Chinese)
- Xia, Y. L., J. Sheffield, M. B. Ek, et al., 2014: Evaluation of multi-model simulated soil moisture in NLDAS-2. *J. Hydrol.*, **512**, 107–125, doi: 10.1016/j.jhydrol.2014.02.027.
- Yang, K., T. Watanabe, T. Koike, et al., 2007: Auto-calibration system developed to assimilate AMSR-E data into a land surface model for estimating soil moisture and the surface energy budget. *J. Meteor. Soc. Japan*, **85A**, 229–242, doi: 10.2151/jmsj.85A.229.
- Yeh, T. C., R. T. Wetherald, and S. Manabe, 1984: The effect of soil moisture on the short-term climate and hydrology change—A numerical experiment. *Mon. Wea. Rev.*, **112**, 474–490, doi: 10.1175/1520-0493(1984)112<0474:TEOSMO>2.0.CO;2.