

Article

Novel Soil Moisture Estimates Combining the Ensemble Kalman Filter Data Assimilation and the Method of Breeding Growing Modes

Yize Li ^{1,*}, Hong Shu ^{1,2}, B. G. Mousa ^{1,3}  and Zhenhang Jiao ¹

¹ State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China; shu_hong@whu.edu.cn (H.S.); dr.bahaa1@azhar.edu.eg (B.G.M.); 2016286190128@whu.edu.cn (Z.J.)

² Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China

³ Faculty of Engineering, Al-Azhar University, Cairo 11884, Egypt

* Correspondence: 2017106190034@whu.edu.cn

Received: 15 January 2020; Accepted: 9 March 2020; Published: 10 March 2020



Abstract: Soil moisture plays an important role in climate prediction and drought monitoring. Data assimilation, as a method of integrating multi-geographic spatial data, plays an increasingly important role in estimating soil moisture. Model prediction error, an important part of the background field information, occupies a position that could not be ignored in data assimilation. The model prediction error in data assimilation consists of three parts: forcing data error, initial field error, and model error. However, the influence of model error in current data assimilation methods has not been completely considered in many studies. Therefore, we proposed a theoretical framework of the ensemble Kalman filter (EnKF) data assimilation based on the breeding of growing modes (BGM) method. This framework used the BGM method to perturb the initial field error term w of EnKF, and the EnKF data assimilation to assimilate the data to obtain the soil moisture analysis value. The feasibility and superiority of the proposed framework were verified, taking into consideration breeding length and ensemble size through experiments. We conducted experiments and evaluated the accuracy of the BGM and the Monte Carlo (MC) methods. The experiment showed that the BGM method could improve the estimation accuracy of the assimilated soil moisture and solve the problem of model error which is not fully expressed in data assimilation. This study can be widely used in data assimilation and has a significant role in weather forecast and drought monitoring.

Keywords: soil moisture; data assimilation; breeding of growing modes; ensemble Kalman filter

1. Introduction

Soil moisture is a key variable that controls major physical processes in numerical weather prediction, climate simulation, crop growth models, and flood forecasting by controlling the surface energy distribution of sensible heat and latent heat fluxes in terrestrial–atmosphere coupling and energy and water cycling [1–5]. Accurate observation of high precision soil moisture data is of great significance for crop growth, drought monitoring, and monitoring global climate change [6–9].

At present, soil moisture data are obtained through three main methods: (1) Simulating soil moisture using land surface process models and hydrological models [10–12]. (2) Measuring soil moisture through ground meteorological observatories [13]. (3) Using remote sensing data to retrieve soil moisture [14–18]. Currently, optical and microwave technologies are widely used for monitoring soil moisture, which compensates for the space-time discontinuity in the site to some extent and can obtain large-scale soil moisture observation data by conducting large-scale and even global observations. However, the retrieval method is highly susceptible to the uncertainties associated with

the instruments and algorithms. Moreover, the coverage of surface vegetation also greatly affects the accuracy of remote sensing retrieval. In summary, making the soil moisture data of each system complement each other and using the high-precision large-scale soil moisture data with a continuous time scale is a worthy research topic.

Land surface data assimilation technology, as a method of integrating the multi-geographic spatial data, combines new observation data in the dynamic operation of numerical models to obtain datasets with temporal, spatial, and physical heterogeneity [19,20]. The land surface data assimilation algorithm is mainly divided into two categories: continuous data assimilation algorithm and sequential data assimilation algorithm [21–26]. The ensemble Kalman filter (EnKF) algorithm proposed by Evensen in 1994 is an efficient sequential data assimilation algorithm [27,28]. It is based on the Monte Carlo (MC) method that obtains the background error covariance matrix in practical applications using the idea of ensemble and reduces the number of calculations, having a good effect on data assimilation when dealing with nonlinear systems. Reichle and Koster [29] assimilated the global scale soil moisture data from the scanning multichannel microwave radiometer (SMMR) into the National Aeronautics and Space Administration (NASA) catchment land surface model (CLSM) using the EnKF, and verified them with ground-based measurements. The experimental results showed that the values obtained by data assimilation were better than those of satellite data and the model simulated data and that the data assimilation improved the estimation accuracy of soil moisture in the root zone. It proved that data assimilation could obtain better soil moisture estimation by combining the model simulation data with independent information from remote sensing observation data. Gruber et al. [30] used the triple collocation (TC) method to obtain error covariance statistics for both the modelling and observation subcomponents of a data assimilation system and the linear antecedent precipitation index (API) model, and the soil moisture data from the ground observation sites were assimilated using a Kalman filter, to analyze the advantages of the assimilation method for existing remote sensing observation data. Renzullo et al. [31] developed a terrestrial soil moisture data assimilation system framework for the Australian Water Resources Assessment (AWRA) modelling system. The advanced microwave scanning radiometer—earth observing system (AMSR-E) and advanced scatterometer (ASCAT) satellite soil moisture inversion data were assimilated into the AWRA-L system and compared with the soil moisture site measurements and also with a new network of cosmic-ray moisture probes (CosmOz) spread across the country. It has been demonstrated that the assimilation of AMSR-E and ASCAT soil moisture products into the AWRA-L model could improve the soil moisture estimates, especially in the root-zones layer. Kolassa et al. [32] studied the data assimilation of soil moisture active passive (SMAP) observations, compared different methods for extracting soil moisture information, and used different methods of adjusting observations to determine the effective method to assimilate the SMAP soil moisture observations into the CLSM. The previous studies have shown that data assimilation through the inversion of soil moisture products by microwave remote sensing can improve the estimation accuracy of soil moisture.

In land surface data assimilation, it is usually assumed that the model is perfect (for example, ignoring model errors) [33]. However, this assumption is problematic even in the application of relatively simpler models, and with more complexities in the model, the influence of model errors on the data assimilation results becomes more significant. Generally, model prediction error comes from three factors: (1) initial conditions, (2) forcing data, and (3) model equations [34]. In the EnKF data assimilation, different research areas, different land surface models, and different forcing data will bring a large number of uncertainties to the background field error and data assimilation results. Moreover, the real distribution of the model errors is also complex and diverse. Reichle et al. [35] used non-Gaussian errors to generate a large-scale ensemble and experimentally proved that the ‘good’ ensemble had a great influence on the background field error variance estimation of EnKF data assimilation. Turner et al. [34] systematically generalized the error sources of three models and analyzed the error conditions of the initial conditions and forcing data among the three error factors, perturbing the initial field and forcing data to assimilate the sea surface temperature. Kai et al. [36]

and Reichle et al. [37] used the standard distributed random number as a random error, to perturb the forcing data to generate ensembles for the EnKF model error estimation. Kumar et al. [38] and Kumar et al. [39] applied the standard distributed random number perturbation to forcing data, model parameters, and soil moisture initial fields to generate ensembles. As an important part of the EnKF data assimilation algorithm, the ensemble is the key to express the model error. The estimation of model error from the dispersion between each ensemble member is the advantage of using the EnKF algorithm, which is different from other data assimilation algorithms. Most of the previous studies used the classical MC method to perform error perturbation on the initial field of soil moisture or forcing data and rarely incorporated the model information into the perturbation. The MC method [40,41] is a purely statistical initial value perturbation method. It is usually assumed that the random error of the initial and background fields obeys the standard Gaussian distribution, but this is only an idealized assumption, not a real situation. Another disadvantage of the MC method is that the model information is not included in the perturbation field and the model error is usually not fully expressed in the initial field.

In view of the aforementioned problems, this study uses the breeding of growing modes (BGM) instead of the MC method as the main perturbation method of data assimilation for the initial field and background field, and proposes a theoretical framework of the EnKF data assimilation based on the BGM perturbation method. The BGM method is one of the main methods for generating initial perturbation in numerical prediction [42,43]. The method generates an initial perturbation field by multiple incubations of the model and the cycle of the perturbation error increases, increasing the proportion of the high-speed growth error continuously, thereby rapidly reaching saturation. The BGM method cultivates the perturbation field through the model, so that the perturbation field would contain the model information, which is more in line with the error distribution state under the real situation. When using the BGM method to generate the perturbation field, determining the breeding length and ensemble size are key points. To generate the perturbation field with different models, different parameters need to be set. Therefore, the BGM method is used in this study to perturb the EnKF data assimilation, and experiments are conducted to evaluate the effects of growth mode breeding length and ensemble size on the EnKF assimilation. In this paper, a theoretical framework of the EnKF data assimilation based on the BGM method is proposed for soil moisture estimation. The framework includes two components: using the BGM method to generate the initial field and the BGM method based on EnKF data assimilation. Firstly, this paper discusses the two ensemble perturbation problems, i.e., the breeding length and the ensemble size in the BGM method. For the study area, a breeding length of 72 h and an ensemble size of 100 were found to be the best solution for data assimilation. On this basis, this study evaluates the impact of the BGM method and the MC method on data assimilation in detail. The Climate Change Initiative (CCI) data and in situ data are used for validating the accuracy and quality of the soil moisture estimated by the EnKF data assimilation, based on the BGM method.

The paper is organized as follows. The theoretical background of the BGM and the EnKF is given in Section 2. Brief introductions to the variable infiltration capacity (VIC) model, the study area, remotely sensed soil moisture products, in situ data, and data pre-processing are provided in Section 3. Details of the experiments and discussion of the results are presented in Sections 4 and 5. Finally, Section 6 summarizes the conclusions and discusses future research directions.

2. Methodology

In this study, we propose an EnKF data assimilation framework based on the BGM perturbation method. The BGM method is used to generate the perturbation field to perturb the initial and background fields. The EnKF assimilation algorithm is then used to perform data assimilation to obtain the analysis value of soil moisture. Figure 1 shows the theoretical framework of the EnKF data assimilation based on the BGM method. The perturbation principle and specific details of the BGM method are introduced in Section 2.1. The EnKF data assimilation is described in Section 2.2.

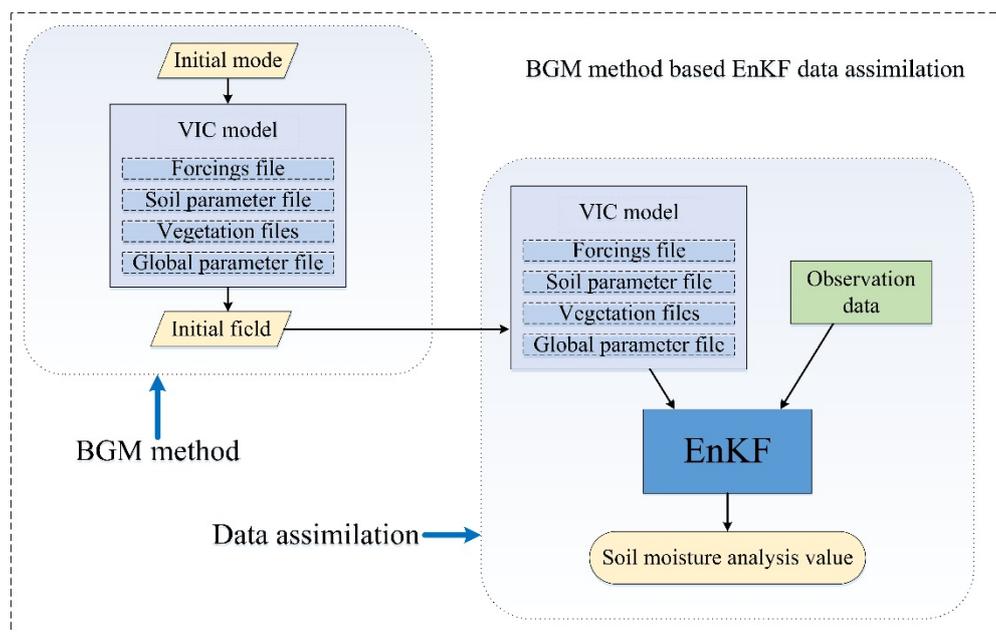


Figure 1. The theoretical framework of the ensemble Kalman filter (EnKF) data assimilation based on the breeding of growing mode (BGM) method. Variable infiltration capacity (VIC) model indicates the variable infiltration capacity model.

2.1. Breeding of Growing Modes

With the initial ensemble perturbations, we must provide an accurate representation of the probability distribution of the true state of the land surface. The shape of this probability distribution will depend on the kind of errors in the true state. If the likelihood of an error pattern is high, we should assign a higher probability to the true state plus or minus that particular error pattern. The BGM method is based on the notion that using the land surface model to dynamically propagate information about the state of the system in space and time will accumulate growing errors by virtue of the perturbation dynamics [42,43]. Compared to other initial distribution methods, this method is more advantageous because of its clear dynamic meaning, simple operation, and low computational cost. In 1992, it was commercialized at the National Centers for Environmental Prediction (NCEP).

Normally, the main sources of model forecast errors can be classified as follows: (1) initial conditions (in this paper, the initial field of soil moisture is specified. Since the true value of soil moisture under natural conditions is unknown, the initial field of soil moisture contains random errors.); (2) boundary conditions (the true value of atmospheric forcing data is also not available. In land surface models and hydrological models, the uncertainty of the forcing data will also bring great uncertainty to the model simulation.); and (3) model equations (since the model uses parameterization and the equation of natural physical processes, the model itself will also bring uncertainty to the estimation of soil moisture). In general, soil moisture estimation using ensembles can be represented by a series of numerical integrals:

$$sm_i(T) = sm_i(0) + \int_{t=0}^T M(sm_i, f, t) dt \quad (1)$$

where $sm_i(T)$ represents the estimated soil moisture model at time T ; $sm_i(0)$ represents the initial field of the soil moisture; M represents the land surface model; and f represents the atmospheric forcing data. Compared with the MC method, the BGM method uses the model integral to cultivate the perturbation field. As the cycle of the perturbation error increases, the proportion of the high-speed growth error increases continuously and later quickly reaches saturation. The BGM method is mainly divided into the following steps:

(1) First, a small random perturbation P_0 is added to the initial field I_0 of the model as the initial mode at the initial time t_0 . The initial mode refers to the growth mode that has not been propagated, and its selection is the first step in implementing the growth mode breeding cycle. There are many ways to select the initial mode. In this study, the initial mode is selected using Equation (2):

$$P(z) = tbE(z) \quad (2)$$

where $P(z)$ is the initial mode of the breeding cycle; t is the adjustment coefficient used to control the initial mode size; b is the Gaussian distribution random number obeying the mean of 0 variance as the model error variance Q ; and $E(z)$ is the root mean square error of soil moisture.

(2) Next, we integrate the original initial field I_0 and the initial field perturbed by the initial mode, i.e. the perturbation field ($I_0 + P_0$), as the initial field of the model and the duration of the integration is recorded as $(t_1 - t_0)$. By estimating the two soil moisture values through the model simulation, we refer to the soil moisture model estimate using the initial field I_0 as the initial field of the model as the control forecast value CF and refer to the soil moisture model estimate using perturbation field $I_0 + P_0$ as the initial field of the model, as the perturbation forecast value PF.

(3) At time t_1 , the difference between the perturbation prediction value and the control forecast value is calculated as $D_1 = PF_1 - CF_1$. To keep the D_1 consistent with the magnitude of the original random perturbation P_0 , the size of D_1 is adjusted. The resized D_1 is taken as the new perturbation P_1 at the next moment. (The process of readjusting the size of D_1 is called 'rescaling').

(4) Adding the newly obtained perturbation P_1 to the model state value I_1 at the corresponding time t_1 and repeating the processes (2)–(4), we use the model to propagate the perturbation field for a period of time. After several breeding cycles, when the growth rate of the perturbation gradually stabilizes at a certain value, we obtain the ideal perturbation field. The process of the BGM method is illustrated in Figure 2.

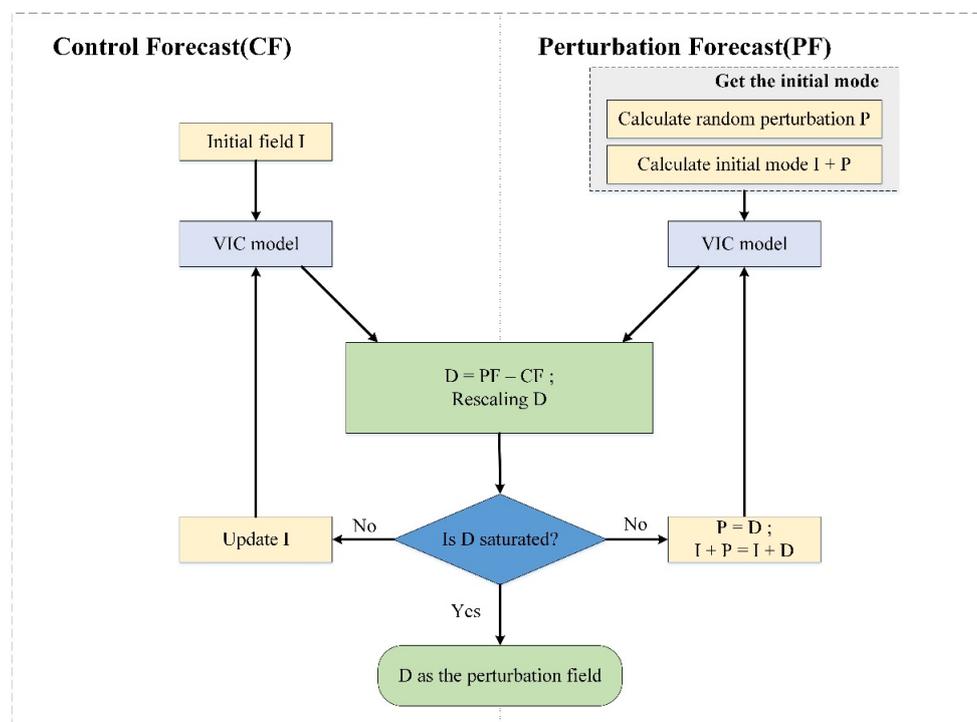


Figure 2. The process of the breeding of growing modes. “I” indicates initial field, “P” indicates random perturbation, “D” indicates the difference between the perturbation prediction value and the control forecast value.

The BGM method simulates the development of growing errors in the model forecast cycle. Compared with the MC method, the BGM method can appropriately adjust the distribution state of the land surface to make its distribution closer to the real state. In data assimilation of the land surface, the BGM method is used to perturb the error of the initial field and the data assimilation model, to permit the physical process of the model to participate in the generation of perturbation. In this way, the perturbation contains not only the statistical information but also the physical information of the error. Compared with the singular vector (SV) method, the BGM method needs less calculation and has higher computational efficiency while attaining the same perturbation effect.

2.2. The BGM Method Based on the EnKF Data Assimilation

The basic idea of the EnKF algorithm is to perturb the state to generate a series of state ensembles, where each ensemble represents a possible value of the state, and the mean value of the ensembles is the optimal estimate of the state. The model is then used to integrate the ensemble forward and predicts the state of each ensemble at the next moment. Each ensemble is updated separately using the observation data. At this point, the mean of all ensembles is the optimal estimation value of the state and the variance of the estimated value could also be calculated through the ensemble. The EnKF algorithm includes two steps, i.e., forecast and update: (1) In the forecast stage, the ensembles of forecast fields are obtained by perturbing a single initial state variable. In this study, the BGM method is used to obtain the ensembles of perturbations. After the ensembles are acquired, the forecast error covariance matrix of the model is calculated using the forecast ensemble. (2) In the update phase, each state variable is updated using the observation value and the observation error covariance matrix, to obtain analysis field ensembles and finally, the mean value of the analysis ensemble is calculated to obtain the analysis value of the model state variable.

2.2.1. Forecast

The forecast process is calculated by the following equation:

$$x_{i,k}^f = M_{k-1,k}(x_{i,k-1}^a) + w_{i,k-1}, (w_{i,k-1} \sim \text{BGM}) \quad (3)$$

where $x_{i,k-1}^a$ is the analysis value of the i -th ensemble member at time $k-1$; $x_{i,k}^f$ is the forecast value of the i -th ensemble member at time k ; $M_{k-1,k}$ is a nonlinear model operator; and $w_{i,k-1}$ is the $k-1$ moment model error. The model error ensemble in this study uses the perturbation field obtained by the BGM method.

2.2.2. Update

In the EnKF, noise that follows the normal distribution is added to the observations, described as follows:

$$y_{i,k}^o = y^o + v_{i,k}, v_{i,k} \sim N(0, R_k) \quad (4)$$

where $y_{i,k}^o$ is the observation corresponding to the i -th ensemble member at time k ; $v_{i,k}$ is the observed error perturbation at time k that obeys a mean of 0; and the covariance matrix is a normal distribution of R_k . The observation error covariance matrix R_k at time k is:

$$R_k = \overline{D_k D_k^T} \quad (5)$$

where $D_k = [v_{1,k}, v_{2,k}, \dots, v_{N,k}]$, $i = 1, 2, \dots, N$ represents the ensemble of observed error perturbations at time k and N represents the ensemble size.

When there are no observations, the model continues to forecast forward and when observations are available the data assimilation updates the state variable of the model:

$$x_{i,k}^a = x_{i,k}^f + K_k(y_{i,k}^o - H_k(x_{i,k}^f)) \quad (6)$$

$$K_k = P_k^f H_k^T (H_k P_k^f H_k^T + R_k^e) \quad (7)$$

where $x_{i,k}^a$ represents the state variable analysis value of the i -th ensemble at time k ; K_k represents the Kalman gain matrix at time k ; H_k represents the observation operator at time k ; and P_k^f represents the forecast error covariance matrix at time k . In contrast to the standard Kalman filtering method [44,45], the EnKF method uses an ensemble of ideas to calculate the error covariance matrix:

$$P_k^f = \frac{1}{N-1} \sum_{i=1}^N (x_{i,k}^f - \bar{x}_k^f) (x_{i,k}^f - \bar{x}_k^f)^T = \frac{1}{N-1} A_k^f (A_k^f)^T \quad (8)$$

$$P_k^f H_k^T = \frac{1}{N-1} \sum_{i=1}^N (x_{i,k}^f - \bar{x}_k^f) \left[H_k(x_{i,k}^f) - H_k(\bar{x}_k^f) \right]^T \quad (9)$$

$$H_k P_k^f H_k^T = \frac{1}{N-1} \sum_{i=1}^N \left[H_k(x_{i,k}^f) - H_k(\bar{x}_k^f) \right] \left[H_k(x_{i,k}^f) - H_k(\bar{x}_k^f) \right]^T \quad (10)$$

$$\bar{x}_k^f = \frac{1}{N} \sum_{i=1}^N x_{i,k}^f \quad (11)$$

$$\bar{x}_k^a = \frac{1}{N} \sum_{i=1}^N x_{i,k}^a \quad (12)$$

where N represents the size of the ensemble; $A_k^f = \left[x_{1,k}^f - \bar{x}_k^f, x_{2,k}^f - \bar{x}_k^f, \dots, x_{N,k}^f - \bar{x}_k^f \right]$ represents the forecast ensemble perturbation; and \bar{x}_k^f represents the mean value of the model state variable forecast at time k .

The main advantage of the EnKF assimilation method is that it does not require the tangent linear mode, the adjoint mode of the computational model, and the observation operator, which greatly reduces the computational complexity. It is suitable for complex and strong nonlinear physical models that have great difficulty in developing tangent linear mode and adjoint mode. However, the EnKF algorithm is susceptible to the size of the ensemble, which would result in faster ensemble propagation and filter divergence. The size of the ensemble is discussed and analyzed in Section 4.2.

3. Datasets

3.1. Study Area and the In Situ Datasets

The Australian Murrumbidgee basin is located in southeastern Australia with a drainage area of 84,000 km², and the study area is located Yanco, in the heart of the Murrumbidgee basin (34.5–35.5° S, 145.5–146.5° E), with an area of 60 km × 60 km (Figure 3). The Yanco area is flat, with few woody vegetation in the study area: rice and barley crops in the west, drought crops in the north, and pastures in the southeast [46]. The annual precipitation in the Yanco region is about 500 mm. The Average Class A pan evaporation varies from less than 1500 mm/year in the southeast to over 2100 mm/year in the west, and this area of arid and semi-humid climates is of particular research value. The OzNet hydrological monitoring network has set up a total of 13 OzNet profiles, widely distributed at various points in the study area to measure the soil moisture. The system observes the soil moisture every 20 min. The daily surface (0–5 cm) soil moisture data obtained were used as verification data to evaluate the data assimilation results.

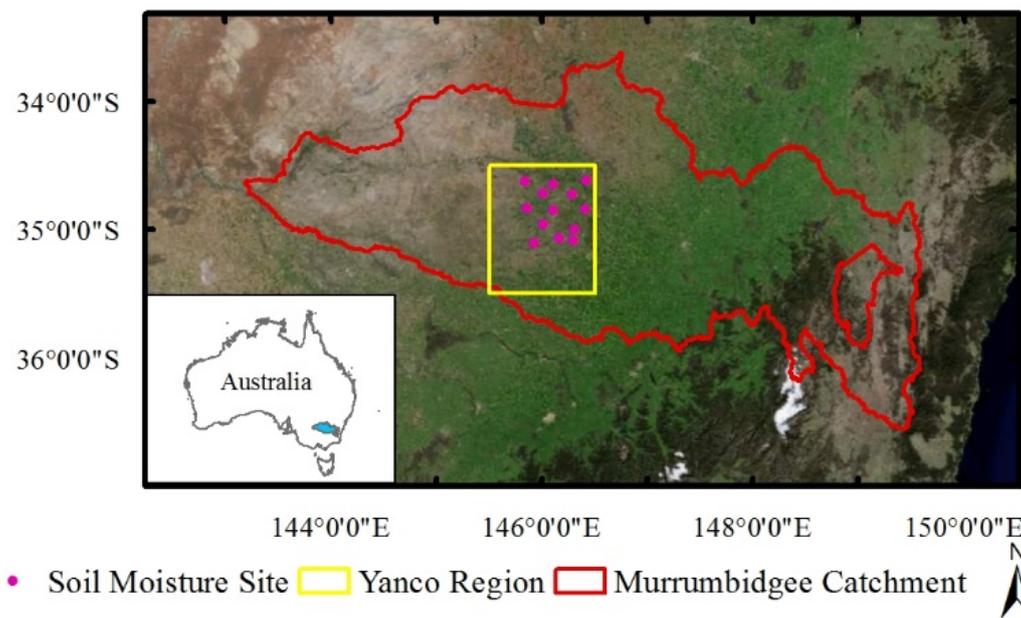


Figure 3. Study area and the spatial distribution of the in situ soil moisture sites.

3.2. Variable Infiltration Capacity Model

The VIC macroscale hydrological model is a large-scale land surface hydrological model developed by researchers at the University of Washington, the University of California at Berkeley, and Princeton University [10,47]. It is based on a spatial distribution network, mainly considering the physical exchange process between atmosphere–vegetation–soil, simulating hydrothermal state changes and water–heat transfer. The model divides the study area into several square grid elements of the same size. The grid size is at least 1 km and the maximum latitude and longitude spacing is 0.5° . The grids are independent of each other. Each grid unit corresponds to its underlying surface parameters and meteorology data, and is calculated independently from each other under the distribution concept. The water volume and energy balance are maintained within each grid and the differences within the subgrid are described statistically.

The VIC model requires five input files: meteorological forcing file, soil parameter file, vegetation parameter file, vegetation library file, and global parameter file. The meteorological forcing file for this experiment uses the ERA-Interim reanalysis data from The European Centre for Medium-Range Weather Forecasts (ECMWF) from 1 July 2015 to 31 December 2016. The data was obtained from <https://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/>. The forcing data include seven forcing variables such as average air temperature, total precipitation (rain and snow), atmospheric pressure, incoming shortwave radiation, incoming longwave radiation, vapor pressure, and wind speed. The maximum spatial resolution of the ERA-Interim reanalysis data is $0.125^\circ \times 0.125^\circ$. To accurately describe the soil moisture value of the study area and its variation, we divided the Yanco experimental area into 19 rows and 19 columns, with a total of 361 independent grids, and set the spatial and temporal resolution of the soil moisture simulated by the VIC respectively to $0.05^\circ \times 0.05^\circ$ and 6 h. Therefore, the spatial resolution of the forcing data is also resampled to $0.05^\circ \times 0.05^\circ$. The soil parameter file is used to describe the spatial heterogeneity of the soil, and the global 5' soil texture classification data provided by the National Oceanic and Atmospheric Administration (NOAA) are used. The file contains parameters such as the main soil type and mean elevation for each grid, some of which need to be graded to fit the study area. The vegetation database comes from the sample files from the VIC official website, storing parameters for each type of land cover, including leaf area index (LAI), structural resistance, zero plane displacement, etc. The vegetation parameter file provides information about multiple vegetation types, arm strengths, and leaf area indices contained in each grid. The vegetation

parameter file was obtained based on the 1 km global vegetation distribution data published by the University of Maryland. The document classifies the land cover types into 14 categories and the details are presented in Table 1. The global parameter file is the interface of the VIC model. It describes the time step of the model runs, the start and end dates of the simulation period, the paths of the five input files and the simulation results, and the log of the frozen soil model runs. It plays the role of guiding and function setting for the operation of the model. Our profile runs the VIC model and obtains the VIC soil moisture simulation data every 6 h from 1 January 2016 to 31 December 2017, and uses the VIC model simulation data and remote sensing observation data for data assimilation.

Table 1. Description of global land cover types developed by the University of Maryland.

Classification Number	Vegetation Classification
0	Water
1	Evergreen needleleaf forest
2	Evergreen broadleaf forest
3	Deciduous needleleaf forest
4	Deciduous broadleaf forest
5	Mixed forest
6	Woodland
7	Woodland grasslands
8	Closed shrublands
9	Open shrublands
10	Grasslands
11	Crop land
12	Bare ground
13	Urban and built-up

3.3. SMAP L3 Soil Moisture Data

The SMAP satellite is a soil moisture satellite launched by NASA on 31 January 2015. The instruments carried by the SMAP include L-band radar and L-band radiometers that provide four remote sensing soil moisture products, namely passive, active, active–passive, and enhanced passive soil moisture. The SMAP radar was discontinued in July 2015; hence, we chose the SMAP L3 passive soil moisture product [48]. This product could be used to observe the soil moisture of the surface at a depth of 0–5 cm. In this study, the SMAP passive level 3 products with a spatial resolution of 36 km generated on an equal-area multi-resolution earth grid are used, and the SMAP soil moisture data are estimated directly through the remote sensing data and the ground station. The SMAP soil moisture data are available freely for users through this link: <https://nsidc.org/data/SPL3SMP>.

3.4. European Space Agency Climate Change Initiative

The CCI soil moisture product dataset published by the European Space Agency (ESA) contains long-term sequence datasets of active microwave remote sensing datasets, passive microwave remote sensing datasets, and active and passive synthetic microwave remote sensing datasets, which provide soil moisture data at a global scale. The spatial resolution is $0.25^\circ \times 0.25^\circ$, the data unit is volumetric water content m^3/m^3 , and the time resolution on average is 1 day. The soil moisture data used in this study are the soil moisture data product synthesized by the CCI v4.2 active and passive microwave soil moisture remote sensing data. The product is available from 1978 to 2016, based on active and passive soil moisture products. More details of the CCI soil moisture products are available in [49,50].

3.5. Data Pre-Processing

The four soil moisture datasets mentioned in the above section have different spatial resolutions, observed soil depths, and soil moisture units, and hence they had to be unified at the pre-processing level. The original information of the data is presented in Table 2.

Table 2. Spatial resolutions, measure soil depths, and soil moisture units of the four soil moisture datasets.

Data	Spatial Resolution	Soil Depth	Soil Moisture Unit
VIC simulation	0.05° × 0.05°	0–10 cm	mm
In situ dataset	Point scale	0–5 cm	m ³ /m ³
Soil moisture active passive (SMAP)	36 km × 36 km	0–5 cm	m ³ /m ³
European Space Agency (ESA) Climate Change Initiative (CCI)	0.25° × 0.25°	0.5–2 cm	m ³ /m ³

To ensure the same spatial resolutions for all the datasets, the nearest neighbor interpolation method was used to resample the spatial resolutions of the SMAP soil moisture data and the ESA CCI soil moisture data to 0.05° × 0.05°. As the soil moisture of the surface layer of 0–10 cm is usually closely related, the observed soil depths from the datasets were regarded as the observation depth of 0–10 cm [51]. The unit of soil moisture in the in situ dataset, the SMAP, and the ESA CCI data were changed to mm, to prepare them for the subsequent data assimilation. Locally, the SMAP, CCI, and the in situ datasets have different climatology, which necessitates a bias correction step, in order to comply with the assumptions of the data assimilation system. The bias correction is implemented by matching the cumulative distribution function (CDF) of the soil moisture products of the SMAP computed over the entire study period to that of the CCI and in situ data at each grid point (CDF-matching) [52,53].

4. Results

In this study, we use the BGM method to perturb the initial field of data assimilation, generate the initial ensemble of the EnKF data assimilation, and use the EnKF method for data assimilation. The breeding length and ensemble size of the BGM method are discussed in Sections 4.1 and 4.2, respectively, through experiments: In Section 4.1, we first superimpose the initial mode of the soil moisture initial field of the VIC model at 0:00 on 2 January 2016 and carry out the BGM method experiment with breeding lengths of 24 h, 48 h, and 72 h. The SMAP L3 soil moisture data were then used to assimilate the experimental area and the data assimilation time ranged from 1 March 2016 to 31 December 2016, for a total of 305 days. In Section 4.2, we take the data assimilation with ensemble sizes of 20, 50, 100, 200, and 500; for the same time period. Section 4.3 performs the data assimilation experiments of the BGM method and the MC method, and evaluates the soil moisture estimation accuracy of the two methods.

For the verification metric of soil moisture data obtained by data assimilation, the root mean square error (RMSE) and the Pearson correlation coefficient R were calculated as follows:

$$RMSE = \sqrt{E[(SM_{estimate} - SM_{truth})^2]} \quad (13)$$

$$R = \frac{E[(SM_{estimate} - E(SM_{estimate}))(SM_{truth} - E(SM_{truth}))]}{\sigma_{estimate}\sigma_{truth}} \quad (14)$$

where $E[\cdot]$ or $E(\cdot)$ represents the mean operator; $SM_{estimate}$ represents the soil moisture data estimate obtained by the VIC simulation or data assimilation; SM_{truth} represents the true value of the soil moisture (the absolute true value of soil moisture could be obtained in the real environment. Here, the soil moisture data verified by comparison represent the true value); $cov(\cdot)$ represents the covariance operation; $\sigma_{estimate}$ and σ_{truth} represent the standard deviations of the estimated soil moisture data and the truth soil moisture data, respectively.

4.1. Influence of Different Mode Breeding Length Perturbation on the EnKF Data Assimilation

The breeding length is a problem worth studying in the BGM method. The breeding length is the total time of propagation of the initial perturbation in the model. There have been some studies on

breeding length in the field of ensemble numerical prediction, but there are few studies on its relation to data assimilation in a land surface model [42]. In this study, with reference to the relevant research results in the field of ensemble numerical prediction, the specific problems of the effects of different reproductive lengths on the land surface data assimilation are tested and discussed.

For the study region, we set the coefficient t in the Equation (2) to be 1 and the ensemble size to be 20. We superimposed the initial mode of the soil moisture initial field of the VIC model at 0:00 on 1 January 2016 and carried out the BGM experiment with the breeding lengths of 24 h, 48 h, and 72 h, and obtained the perturbation fields. The data assimilation initial field was perturbed by the 24 h, 48 h, and 72 h perturbation fields, obtained by the BGM method, and the data of the study area were assimilated using the SMAP L3 soil moisture data. The data assimilation time ranges from 1 March 2016 to 31 December 2016 (305 days). After the data assimilation experiment, we used the CCI data to verify the soil moisture results obtained by the VIC model simulation and the soil moisture analysis obtained by different breeding length data assimilations.

Figure 4 shows the RMSEs of the VIC simulation soil moisture values and the soil moisture analysis values obtained from the data assimilation experiments of different breeding lengths, using the CCI data for validating the results. Data assimilation using the BGM method could effectively improve the accuracy of estimation of the soil moisture and the RMSE values of data assimilation using different model breeding lengths are similar (Figure 4). From the above experimental results, we found that although there are differences between the results of data assimilation experiments at different breeding lengths, the 72 h breeding length has a better perturbation effect on the EnKF data assimilation and the generated soil moisture estimation data had the smallest value of RMSE.

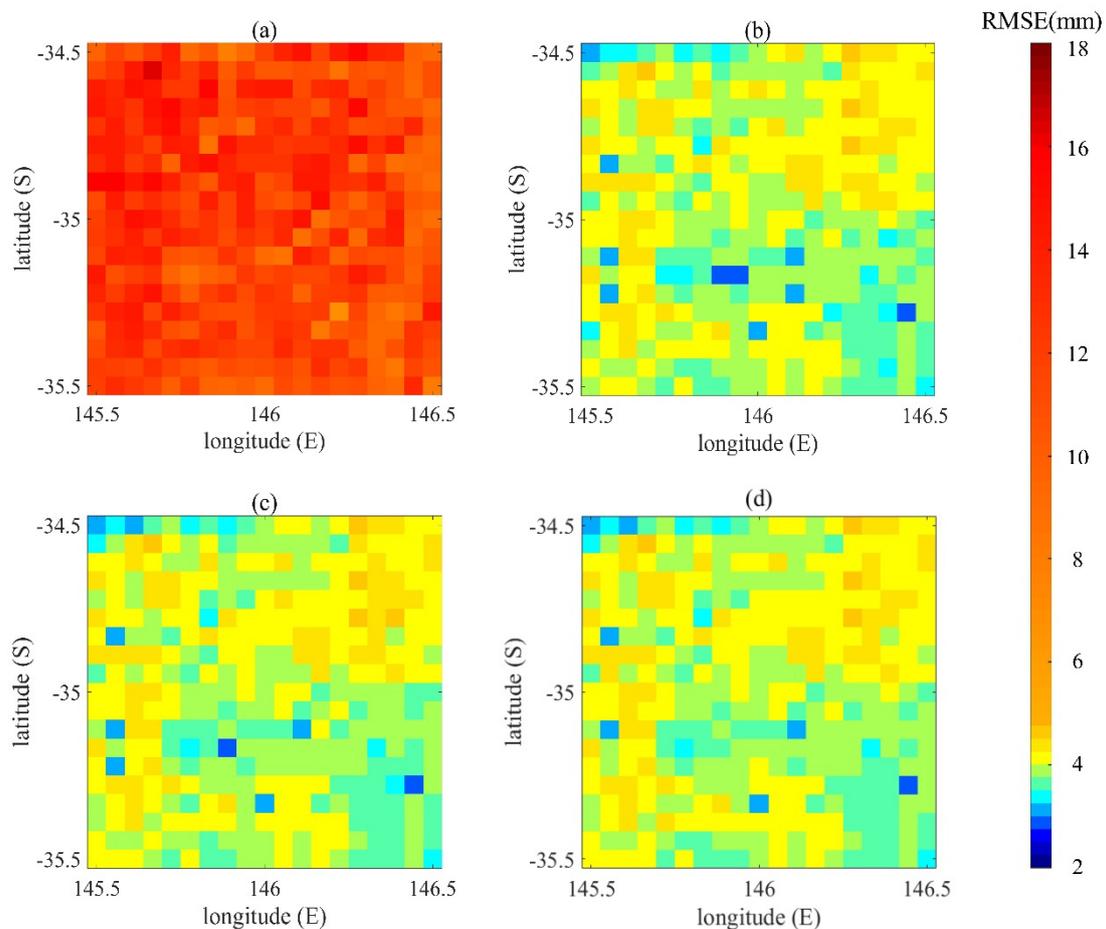


Figure 4. The root mean square error (RMSE) values between the CCI data and: (a) VIC simulation; the BGM method using the breeding lengths of (b) 24 h; (c) 48 h; and (d) 72 h.

Figure 5 shows the correlation coefficient R between the data assimilated soil moisture analysis values with different breeding lengths and the CCI data. Table 3 shows the accuracy of the soil moisture estimation results of the four sets of experiments using the CCI data. From Figure 5, we find that the correlation coefficient of the southern region is slightly higher than that of the northern region, indicating that the data assimilation results of the study area are slightly better than that of the northern region. The highest value of the RMSE of the four experimental sets is the RMSE mean of the VIC simulation experiment, which is 11.9075. Compared with the VIC simulation experiment, the RMSE means of the three sets of data assimilation experiments have significantly improved, which proves that data assimilation using the BGM method could improve the estimation accuracy of the soil moisture. Among the four sets of experiments, the mean and variance of the RMSE of the data assimilation experiment with the breeding length of 72 h are the smallest, indicating that this result was more accurate. The mean values of R for all the experiments are above 0.8. The mean and variance of R of the three groups of data assimilation experiments are very close.

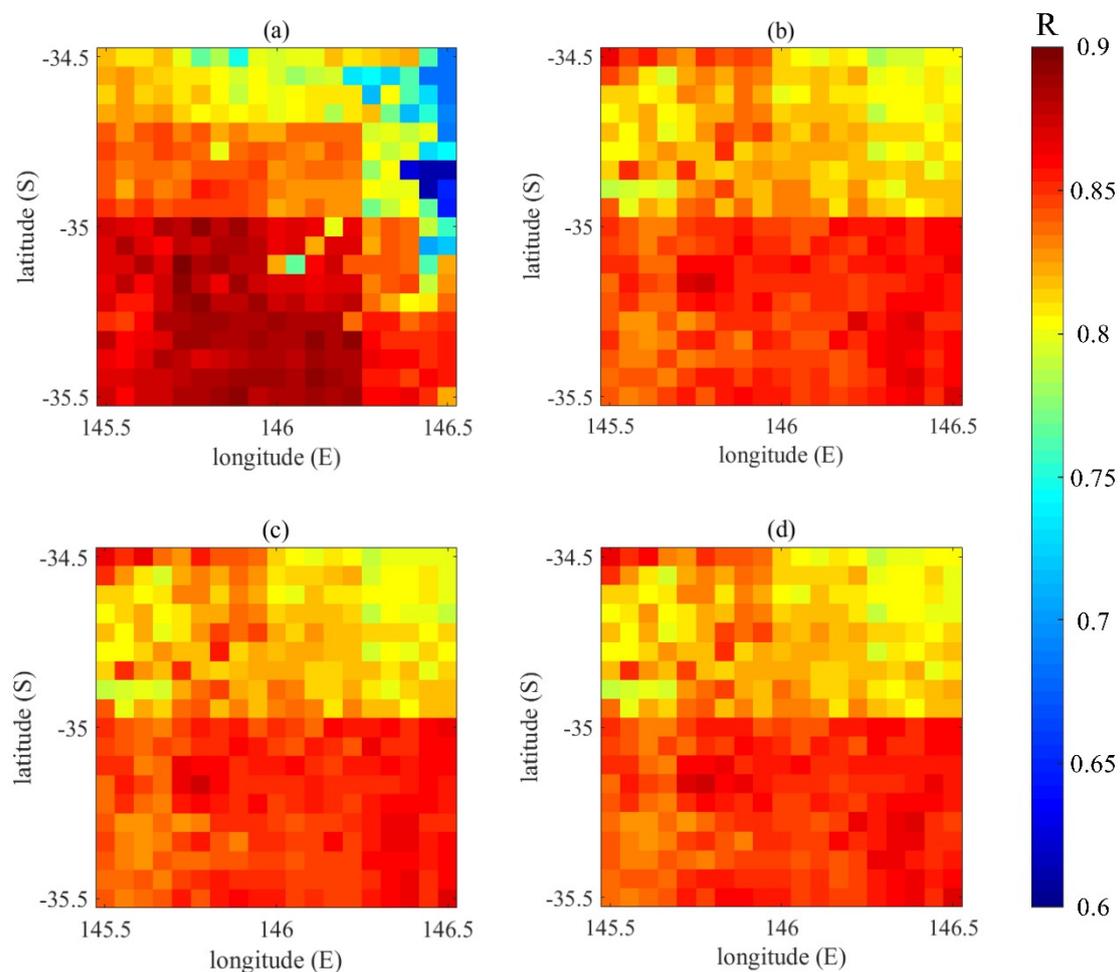


Figure 5. The different correlation coefficient (R) values between the CCI data and: (a) VIC simulation; the BGM method using the breeding lengths of (b) 24 h; (c) 48 h; and (d) 72 h.

Table 3. The mean and variance of the RMSE and R between the model and the data assimilation experiments with different breeding time lengths.

Data	RMSE (mm)		R	
	Mean	Variance	Mean	Variance
VIC Simulation	11.9075	3.1163	0.8041	0.0028
DA _{24h} breeding time	3.9714	0.0888	0.8344	0.00041143
DA _{48h} breeding time	3.9665	0.0819	0.8344	0.00041223
DA _{72h} breeding time	3.9579	0.0765	0.8345	0.00040387

Through the above experiments, the soil moisture analysis values obtained from the initial field generated by the BGM method with different breeding lengths were analyzed and evaluated. It was found that compared with the 24 h and 48 h breeding time, the perturbation effect of the EnKF data assimilation with 72 h breeding time is better and the generated soil moisture estimation data have smaller RMSE and larger R, which is more consistent with the verification data.

4.2. The Effect of Ensemble Sizes on Data Assimilation

The EnKF performs the estimation of the background error covariance matrix by the statistical simulation of a finite ensemble of members; hence, different ensemble sizes have a significant influence on the data assimilation results. If the size of the ensemble members is too small, the background error covariance will be unrealistically small and eventually lead to filter divergence, which makes data assimilation impossible; if it is too large, the efficiency of data assimilation will be significantly affected. Therefore, the size of the ensemble plays an important part in the EnKF perturbation. We take data assimilation with ensemble sizes of 20, 50, 100, 200, and 500 to evaluate the soil moisture analysis values obtained by using different ensemble sizes for data assimilation.

Based on the conclusions of the experiments on the effects of different breeding length perturbations on the EnKF data assimilation (see Section 4.1), we uniformly set the breeding length to 72 h. The BGM method was used to generate a set of ensemble sizes of 20, 50, 100, 200, and 500, and data assimilation experiments were performed. The experimental results were validated using the CCI observation data. The data assimilation time range is from 1 March 2016 to 31 December 2016 (305 days). Figure 6 shows the RMSE of the soil moisture analysis values obtained from the data assimilation experiments of different ensemble sizes, compared with the CCI data. The correlations (R) between the observations and the results from data assimilation with different ensemble sizes are shown in Figure 7. Table 4 shows the accuracy of soil moisture estimation results for the five sets of experiments with different ensemble sizes. We found that using the BGM method to generate the initial field, even for small ensembles, could still obtain good soil moisture estimates. The appropriate ensemble size does not affect the efficiency of the data assimilation procedure. It could express the background field error well and could prevent the filter from diverging. When the ensemble size is 100, the RMSE mean is small and the variance is comparable to the data assimilation experiments of the other ensemble sizes. Through the mean and variance of RMSE and R, we found that when using the BGM method to generate ensembles of different sizes for data assimilation, the obtained data assimilation results are different, but those differences are not very big. Therefore, we chose the ensemble size 100 for data assimilation to ensure the efficiency and accuracy of the soil moisture estimation for data assimilation.

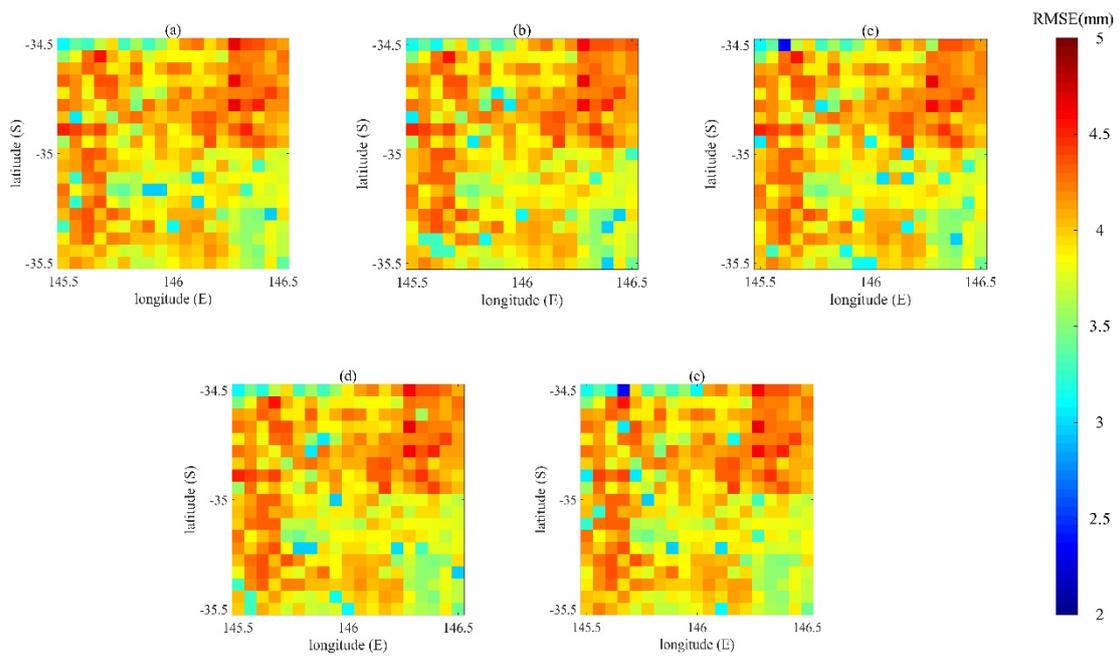


Figure 6. The RMSE values between the CCI data and the assimilated data with ensemble sizes of: (a) 20; (b) 50; (c) 100; (d) 200; and (e) 500.

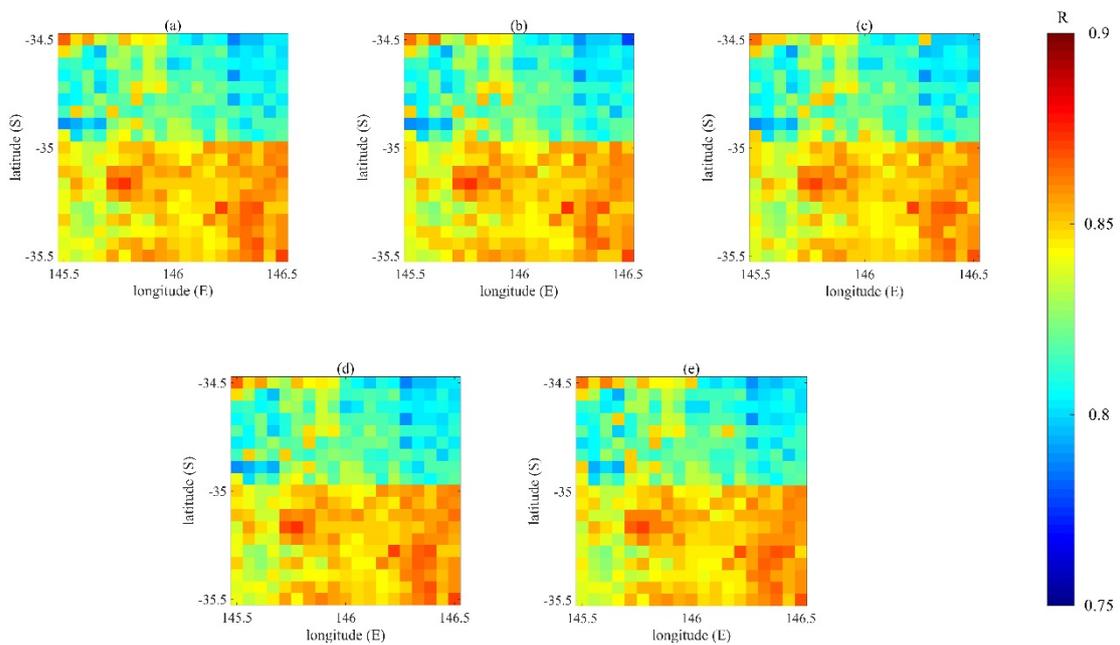


Figure 7. The R values between the CCI data and the assimilated data with ensemble sizes of: (a) 20; (b) 50; (c) 100; (d) 200; and (e) 500.

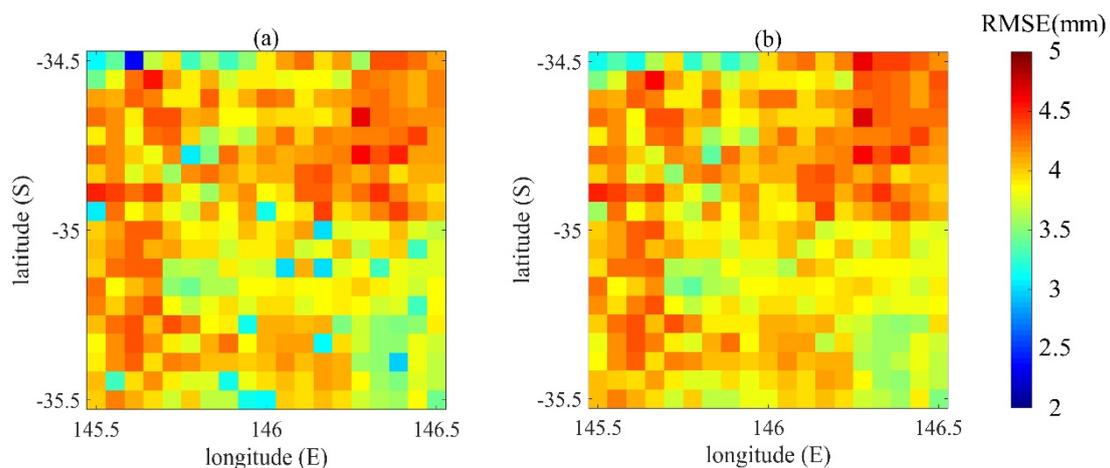
Table 4. The mean and variance of the RMSE and R between the CCI data and the data assimilation experiments with different ensemble sizes.

Ensemble Size	RMSE (mm)		R	
	Mean	Variance	Mean	Variance
20	3.9579	0.0988	0.8345	0.00041143
50	3.9515	0.0920	0.8343	0.00041474
100	3.9458	0.0966	0.8343	0.00040132
200	3.9528	0.0907	0.8341	0.00040602
500	3.9436	0.0968	0.8345	0.00040283

4.3. Comparison of Data Assimilation Experiments

As shown in Sections 4.1 and 4.2, we have determined the breeding length and initial field ensemble size of the BGM method. Now, we use these values to conduct experiments using the BGM and the MC method and compare the accuracy of the soil moisture estimation between the two methods. The CCI observation data and in situ data from 13 locations in the study area were used to validate the data assimilation results. The RMSE and R were calculated to quantitatively evaluate the different data assimilation results.

Figure 8 shows the RMSE and Figure 9 shows the R between the CCI data and the soil moisture analysis values obtained by the data assimilation experiment, using different perturbation methods. The RMSE mean of the soil moisture estimation results obtained by data assimilation using the BGM method is 3.9458 and that using the MC method is 3.9911. The R mean of soil moisture estimation results obtained by data assimilation using the BGM method is 0.8343 and using the MC method is 0.8272. The accuracy of the soil moisture analysis values obtained from data assimilation using the BGM method is better than that obtained using the MC method. To accurately characterize the influence of different perturbation methods on the estimation accuracy, we calculated the deviation between the RMSEs of the two assimilation methods, that is, the RMSE of the data assimilation analysis value and the true value of the BGM method, minus the RMSE of the data assimilation analysis value and the true value of the MC method, $\delta_{\text{RMSE}} = \text{RMSE}_{\text{DA(BGM)}} - \text{RMSE}_{\text{DA(MC)}}$. If δ_{RMSE} is negative, the estimation accuracy of the BGM method is higher than that of the MC method; if δ_{RMSE} is positive, it is proved that the estimation accuracy of the MC method is higher than that of the BGM method. The specific results are shown in Figure 10.

**Figure 8.** The RMSE between the CCI data and assimilated data using: (a) the BGM method; and (b) the Monte Carlo (MC) method.

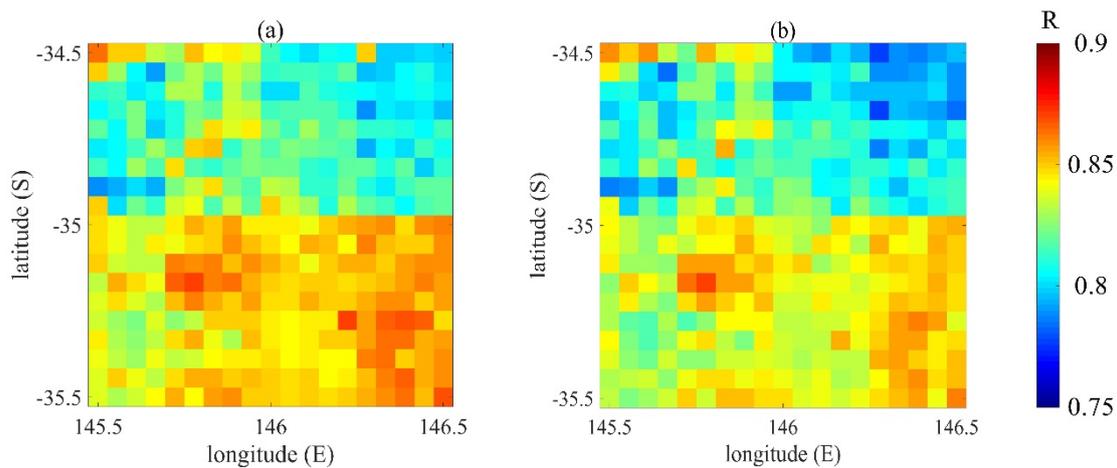


Figure 9. The R between the CCI data and assimilated data using: (a) the BGM method; and (b) the MC method.

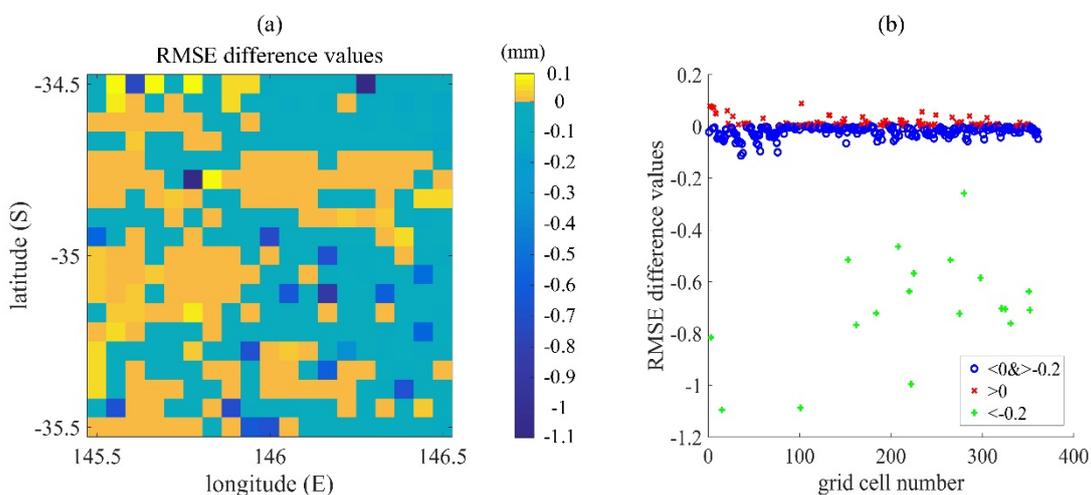


Figure 10. The deviation in RMSEs between the data assimilations using the BGM and the MC methods: (a) Spatial distribution of RMSE difference values over study area; (b) The values of RMSE difference values with grid cell number.

The blue area shows negative values and the yellow area shows positive values. The number of grids in the blue region is 248, accounting for 68.7% of the total number of grids. The minimum value in the blue region is -1.0952 mm, which proves that the RMSE of the BGM method is smaller than the RMSE of the MC method by 1.0952 mm. The number of grids in the yellow area is 113, accounting for 31.3% of the total grid number and the maximum value of the yellow area is 0.0885 mm, which proves that the RMSE of the MC method is smaller than the RMSE of the BGM method by 0.0885 mm. In Figure 10b, the red point is positive and the blue and green points are negative. Most of the points are distributed around the value of 0, where there are more blue points than red, which proves that the BGM method is better than the MC method in most of the grids. The green dot represents a grid point with $\delta_{RMSE} < -0.2$ mm and has a total of 19 grid points, accounting for about 0.05% of the total grid points. These points represent a significant improvement in the data assimilation ability of the BGM method compared with the MC method. It is proved by experiments that the BGM method is better than the MC method in data assimilation perturbation and could improve the accuracy of the data assimilation for soil moisture estimation.

In addition to the validation with the CCI observation data, we also use data from 13 in situ stations in the study area. Since the Y3 site had no soil moisture data in 2016, we verified the experimental results using the remaining 12 sites excluding the Y3. Site data can only represent point-scale soil

moisture observations, so we used the site data to verify the soil moisture analysis values of the grid where the site is located.

Figure 11 and Table 5 clearly show that the data assimilation could significantly improve the accuracy of the soil moisture estimation compared with the soil moisture simulated by the VIC model. Among them, the soil moisture analysis value obtained by data assimilation using the BGM method is more accurate than the MC method data assimilation. Based on the data assimilation experiments corresponding to the 12 in situ sites, the RMSE of the analysis values using the BGM method at the 10 sites (Y1, Y2, Y4, Y5, Y6, Y7, Y10, Y11, Y12, and Y13) are better than the RMSE of the data assimilated values obtained using the MC method. The two assimilated values at the Y6 site had the largest RMSE deviation of 0.8500. The RMSE of the MC method data assimilation analysis values at the Y8 and Y9 sites is better than the RMSE of the values obtained using the BGM method. In the 12 sites of the study area, the average of the RMSE of the BGM method analysis values was reduced by 0.1015, compared with the average RMSE of the MC method values. For the grids corresponding to the 12 sites, the R of the BGM method analysis value is greater than the R of the MC method value. The average value of R of the BGM method data assimilation analysis value increased by 0.0071 compared with the average value of R of the MC method data assimilation analysis value. The experiment proves that the accuracy of the data assimilation analysis value obtained using the BGM method is higher than obtained using the MC method.

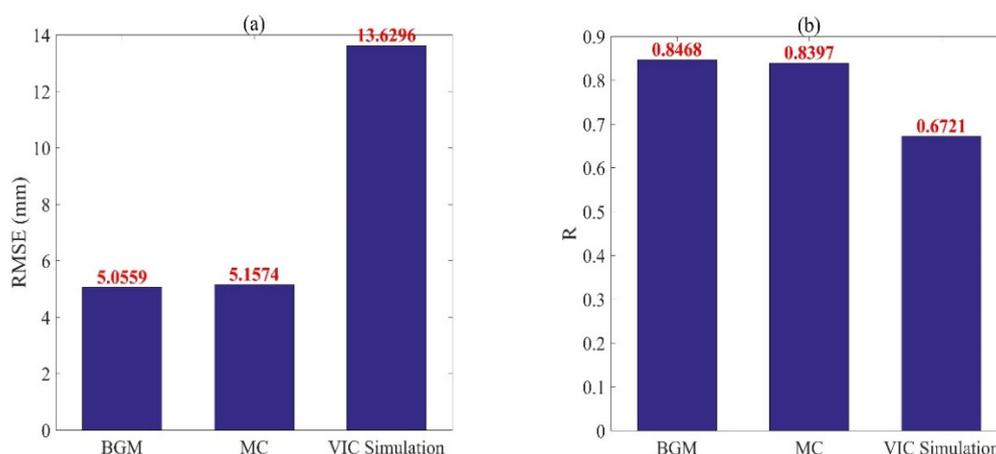


Figure 11. The RMSE (a) and R (b) of the data assimilation using the BGM method, the MC method and the VIC simulation experiments.

Table 5. The RMSE and R of the assimilated values (using the BGM and MC methods) and the simulation experiment (VIC) spatially averaged over the grids where the in situ sites are located.

In Situ Data	BGM		MC		VIC Simulation	
	RMSE	R	RMSE	R	RMSE	R
Y1	2.5175	0.7120	2.5205	0.7118	21.4296	0.7887
Y2	4.3142	0.9401	4.3373	0.9305	7.4960	0.7810
Y4	1.8762	0.8656	1.8896	0.8643	13.3195	0.7066
Y5	5.5269	0.8516	5.5389	0.8512	13.8519	0.6628
Y6	9.1090	0.8464	9.9590	0.8401	14.8214	0.5254
Y7	5.3233	0.7578	5.3732	0.7539	15.5001	0.6542
Y8	4.0231	0.9061	3.9824	0.9036	9.9349	0.6489
Y9	6.0570	0.9177	6.0390	0.9132	10.2563	0.7050
Y10	8.7091	0.8398	8.7140	0.8206	9.6082	0.7783
Y11	1.9072	0.8521	1.9127	0.8504	14.8497	0.7075
Y12	1.9132	0.8941	1.9352	0.8639	14.8263	0.6120
Y13	9.3942	0.7788	9.6872	0.7731	17.6611	0.4949
Mean	5.0559	0.8468	5.1574	0.8397	13.6296	0.6721

5. Discussion

We argue that a more accurate estimate of the soil moisture could be obtained using the BGM method to perturb data assimilation compared to the MC method. The BGM method uses the model loop to describe the growth of error in the model, so that the perturbation field contains not only the initial field error information, but also the model error information. As a purely statistical initial field perturbation method, the MC method could only generate random numbers obeying the standard distribution to represent the error. Therefore, the MC method usually assumes that the random errors of the initial field and the background field obey the standard Gaussian distribution. The BGM method uses the model to breed the error and uses the model to appropriately adjust the form of the error distribution, so that the error distribution is more in line with the true error distribution state.

As pointed out by Renzullo et al. [31], implementation of the EnKF does not explicitly represent the model error. The EnKF data assimilation based on the BGM method proposed in this study provides a new idea to consider the model error during error estimation of data assimilation. The BGM method allows the model to participate in the perturbation of the initial field and the error term w so that the model information is completely absorbed by the ensemble, and the EnKF data assimilation completes the expression of the model error. We used two sets of experiments to address the two key issues: the breeding length and ensemble size in the ensemble perturbation and data assimilation. We performed a quantitative analysis of the data assimilation experiments using different breeding lengths (Section 4.1), and found that a breeding time of 72 h produced the most accurate soil moisture estimates. When the breeding time is too long, the accuracy of the data assimilation is not significantly improved; however, longer breeding times have a significant impact on experimental efficiency. We also conducted experiments with the BGM method using different ensemble sizes (Section 4.2). Different ensemble sizes lead to differences in the data assimilation results, but this difference is not significant, because the perturbation field generated by the BGM method could completely express the random error and the model error of the initial state of the soil moisture. The choice of the ensemble size should not only be considered in terms of the accuracy of the soil moisture estimation, but also the efficiency of data assimilation. The expansion of the ensemble size will reduce the efficiency of data assimilation and increase the running time. Therefore, on the basis of balancing operational efficiency, we chose an ensemble size of 100 as the ensemble size option for the BGM method experiment.

In order to fully verify the accuracy of the EnKF data assimilation framework based on the BGM method, we used in situ data from 13 locations in the study area as the main verification data to validate the data assimilation of two different perturbation methods. The in situ stations in the study area were located sparsely, which limited the verification of large-scale soil moisture data. Therefore, we set the spatial resolution of the data assimilation analysis value as 0.05° , so the data assimilation analysis value would be closer in spatial resolution with data from sites, and the verification result of the site data for soil moisture could be more accurate and reasonable. The in situ data cannot afford the verification for the data assimilation analysis value in all the study areas, so we use CCI data as auxiliary data. The data precision of CCI data in Yanco area is high enough to support verification [18]. In 2016, the overall coverage rate of CCI data in Yanco area was 94.30%, which provides enough samples as validation data. The difference in spatial resolution between datasets would result in uncertainty, but giving that the soil moisture analyzed value validation we performed is between the site and CCI data, the verification results of this paper are still objective and reliable.

The experimental comparison of the data assimilation using two different sets of perturbation methods (see Section 4.3) shows that the initial perturbation obtained by the BGM method is more suitable for the initial field and error term w of the disturbance data assimilation, compared with the initial perturbation obtained by the MC method. The accuracy of the soil moisture analysis value estimated by the BGM method is higher than that estimated by the MC method. The EnKF data assimilation theory framework based on the BGM method proposed in this study could improve the estimation of soil moisture. The theoretical framework uses the BGM method to breed the perturbation field so that the initial field and the error term w contain the model information. The perturbation field

bred by the BGM method is used to perturb the initial field error and the background field error of the EnKF so that the soil moisture estimation of the EnKF data assimilation could be accurately adjusted.

6. Summary and Conclusions

This study proposes using the BGM method for the EnKF data assimilation in relation to the perturbation of land surface data assimilation. The BGM method is widely used in meteorological numerical prediction and atmospheric data assimilation. This paper introduces the BGM method and applies it to land surface data assimilation, using the BGM method to generate the initial field of the EnKF and the error term w of the prediction step in the EnKF, so that the perturbation ensemble completes the error expression in the EnKF data assimilation. The advantages of using the BGM method are: (1) The BGM method uses the model in breeding the mode so that the initial field and error term w of the data assimilation not only represent the initial field error, but also contain relevant information of the model error; (2) Adjust the ensemble distribution pattern of the initial field and the error term w to make it more in line with the error distribution state under real conditions (the soil moisture error distribution in the actual state does not necessarily strictly obey the standard Gaussian distribution). In this study, the BGM method is used to perform the EnKF data assimilation and the problem of the BGM perturbation mode breeding length and ensemble size is experimented. Finally, the experimental comparison with the MC method proves the applicability and superiority of the BGM method in land surface data assimilation.

The results of the validations done using the CCI remote sensing data and in situ data are different because the accuracy of in situ data is high, but its spatial resolution is point scale and the spatial resolution of the analysis values is different. The CCI data are consistent with the spatial resolution of the analysis values, but the accuracy of the CCI data is not as good as the in situ data. The data assimilation results of the two perturbation methods used in this experiment have significantly improved the accuracy of the soil moisture estimates compared to the results from the VIC model simulation. The data assimilation uses accurate observation data to correct the simulated values of the model; therefore, the data assimilation results are more accurate. Even when the accuracy of the observation data is not ideal, the accuracy of the analysis value of the BGM method EnKF data assimilation is still reliable and needs to be further explored in future research.

In this study, we propose an EnKF data assimilation framework based on the BGM method. It is experimentally proven that the BGM method could effectively generate the data assimilation initial field, perturbation error term w , and improve the estimation accuracy of the data assimilation of soil moisture. In future research, we plan to extend the EnKF data assimilation framework based on the BGM method in the following ways: (1) discuss the influence of the framework on the unbiased of the EnKF state estimator. (2) evaluate whether the BGM method is applicable to data assimilation on a global scale and (3) analyze the applicability of the BGM method to other Kalman filter data assimilation methods.

Author Contributions: Conceptualization, Y.L.; Data curation, Y.L.; Formal analysis, Y.L., H.S. and Z.J.; Investigation, Y.L.; Methodology, Y.L. and B.G.M.; Resources, Y.L.; Software, Y.L.; Supervision, Y.L.; Validation, Y.L.; Visualization, Y.L.; Writing—original draft, Y.L.; Writing—review & editing, Y.L. and B.G.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China: 61971316; the National Key Research Development Program of China: 2017YFB0503604; the National Key Research Development Program of China: 2016YFB0502204.

Acknowledgments: We are grateful to the science teams for providing excellent and accessible SMAP (smap.jpl.nasa.gov/data/), CCI (www.esa-soilmoisture-cci.org/), OzNet (www.oznet.org.au/murrumbidgeesm.html), and ECMWF (www.ecmwf.int) data, and the VIC (vic.readthedocs.io/en/master/) model.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Entekhabi, D.; Rodriguez-Iturbe, I.; Castelli, F. Mutual interaction of soil moisture state and atmospheric processes. *J. Hydrol.* **1996**, *184*, 3–17. [[CrossRef](#)]
2. Western, A.W.; Grayson, R.B.; Blöschl, G. Scaling of soil moisture: A hydrologic perspective. *Annu. Rev. Earth Planet. Sci.* **2002**, *30*, 149–180. [[CrossRef](#)]
3. Aubert, D.; Loumagne, C.; Oudin, L. Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall-runoff model. *J. Hydrol.* **2003**, *280*, 145–161. [[CrossRef](#)]
4. Seneviratne, S.I.; Luthi, D.; Litschi, M.; Schar, C. Land-atmosphere coupling and climate change in Europe. *Nature* **2006**, *443*, 205–209. [[CrossRef](#)]
5. Gentile, P.; Polcher, J.; Entekhabi, D. Harmonic propagation of variability in surface energy balance within a coupled soil-vegetation-atmosphere system. *Water Resour. Res.* **2011**, *47*, 21. [[CrossRef](#)]
6. Seneviratne, S.I.; Corti, T.; Davin, E.L.; Hirschi, M.; Jaeger, E.B.; Lehner, I.; Orlowsky, B.; Teuling, A.J. Investigating soil moisture-climate interactions in a changing climate: A review. *Earth Sci. Rev.* **2010**, *99*, 125–161. [[CrossRef](#)]
7. McColl, K.A.; Alemohammad, S.H.; Akbar, R.; Konings, A.G.; Yueh, S.; Entekhabi, D. The global distribution and dynamics of surface soil moisture. *Nat. Geosci.* **2017**, *10*. [[CrossRef](#)]
8. Assouline, S. Infiltration into soils: Conceptual approaches and solutions. *Water Resour. Res.* **2013**, *49*, 1755–1772. [[CrossRef](#)]
9. Findell, K.L.; Eltahir, E.A.B. Atmospheric controls on soil moisture-boundary layer interactions. Part II: Feedbacks within the continental United States. *J. Hydrometeorol.* **2003**, *4*, 570–583. [[CrossRef](#)]
10. Liang, X.; Lettenmaier, D.P.; Wood, E.F.; Burges, S.J. A simple hydrologically based model of land-surface water and energy fluxes for general-circulation models. *J. Geophys. Res. Atmos.* **1994**, *99*, 14415–14428. [[CrossRef](#)]
11. Hamman, J.J.; Nijssen, B.; Bohn, T.J.; Gergel, D.R.; Mao, Y.X. The Variable Infiltration Capacity model version 5 (VIC-5): Infrastructure improvements for new applications and reproducibility. *Geosci. Model. Dev.* **2018**, *11*, 3481–3496. [[CrossRef](#)]
12. Merwade, V.; Baffaut, C.; Bieger, K.; Boithias, L.; Rathjens, H. Featured Series Introduction: SWAT Applications for Emerging Hydrologic and Water Quality Challenges. *J. Am. Water Resour. Assoc.* **2017**, *53*, 67–68. [[CrossRef](#)]
13. Smith, A.B.; Walker, J.P.; Western, A.W.; Young, R.I.; Ellett, K.M.; Pipunic, R.C.; Grayson, R.B.; Siriwardena, L.; Chiew, F.H.S.; Richter, H. The Murrumbidgee soil moisture monitoring network data set. *Water Resour. Res.* **2012**, *48*, 6. [[CrossRef](#)]
14. Karthikeyan, L.; Pan, M.; Wanders, N.; Kumar, D.N.; Wood, E.F. Four decades of microwave satellite soil moisture observations: Part 1. A review of retrieval algorithms. *Adv. Water Resour.* **2017**, *109*, 106–120. [[CrossRef](#)]
15. Karthikeyan, L.; Pan, M.; Wanders, N.; Kumar, D.N.; Wood, E.F. Four decades of microwave satellite soil moisture observations: Part 2. Product validation and inter-satellite comparisons. *Adv. Water Resour.* **2017**, *109*, 236–252. [[CrossRef](#)]
16. Zeng, J.Y.; Li, Z.; Chen, Q.; Bi, H.Y.; Qiu, J.X.; Zou, P.F. Evaluation of remotely sensed and reanalysis soil moisture products over the Tibetan Plateau using in-situ observations. *Remote Sens. Environ.* **2015**, *163*, 91–110. [[CrossRef](#)]
17. Cui, C.Y.; Xu, J.; Zeng, J.Y.; Chen, K.S.; Bai, X.J.; Lu, H.; Chen, Q.; Zhao, T.J. Soil Moisture Mapping from Satellites: An Intercomparison of SMAP, SMOS, FY3B, AMSR2, and ESA CCI over Two Dense Network Regions at Different Spatial Scales. *Remote Sens.* **2018**, *10*, 33. [[CrossRef](#)]
18. Ma, H.L.; Zeng, J.Y.; Chen, N.C.; Zhang, X.; Cosh, M.H.; Wang, W. Satellite surface soil moisture from SMAP, SMOS, AMSR2 and ESA CCI: A comprehensive assessment using global ground-based observations. *Remote Sens. Environ.* **2019**, *231*, 14. [[CrossRef](#)]
19. Reichle, R.H. Data assimilation methods in the Earth sciences. *Adv. Water Resour.* **2008**, *31*, 1411–1418. [[CrossRef](#)]
20. Kalnay, E.; Kanamitsu, M.; Kistler, R.; Collins, W.; Deaven, D.; Gandin, L.; Iredell, M.; Saha, S.; White, G.; Woollen, J.; et al. The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **1996**, *77*, 437–471. [[CrossRef](#)]

21. Drusch, M. Initializing numerical weather prediction models with satellite-derived surface soil moisture: Data assimilation experiments with ECMWF's Integrated Forecast System and the TMI soil moisture data set. *J. Geophys. Res. Atmos.* **2007**, *112*, 14. [[CrossRef](#)]
22. Jones, A.S.; Cogan, J.; Mason, G.; McWilliams, G. Design of a temporal variational data assimilation method suitable for deep soil moisture retrievals using passive Microwave radiometer data. In Proceedings of the 2010 11th Specialist Meeting on Microwave Radiometry and Remote Sensing of the Environment, Washington, DC, USA, 1–4 March 2010; pp. 60–62.
23. Tian, X.J.; Xie, Z.H.; Dai, A.; Shi, C.X.; Jia, B.H.; Chen, F.; Yang, K. A dual-pass variational data assimilation framework for estimating soil moisture profiles from AMSR-E microwave brightness temperature. *J. Geophys. Res. Atmos.* **2009**, *114*, 12. [[CrossRef](#)]
24. Reichle, R.H.; Koster, R.D. Assessing the impact of horizontal error correlations in background fields on soil moisture estimation. *J. Hydrometeorol.* **2003**, *4*, 1229–1242. [[CrossRef](#)]
25. Crow, W.T.; Chen, F.; Reichle, R.H.; Liu, Q. L band microwave remote sensing and land data assimilation improve the representation of prestorm soil moisture conditions for hydrologic forecasting. *Geophys. Res. Lett.* **2017**, *44*, 5495–5503. [[CrossRef](#)] [[PubMed](#)]
26. Lievens, H.; Reichle, R.H.; Liu, Q.; De Lannoy, G.J.M.; Dunbar, R.S.; Kim, S.B.; Das, N.N.; Cosh, M.; Walker, J.P.; Wagner, W. Joint Sentinel-1 and SMAP data assimilation to improve soil moisture estimates. *Geophys. Res. Lett.* **2017**, *44*, 6145–6153. [[CrossRef](#)] [[PubMed](#)]
27. Evensen, G. sequential data assimilation with a nonlinear quasi-geostrophic model using monte-carlo methods to forecast error statistics. *J. Geophys. Res. Ocean* **1994**, *99*, 10143–10162. [[CrossRef](#)]
28. Evensen, G. The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean. Dyn.* **2003**, *53*, 343–367. [[CrossRef](#)]
29. Reichle, R.H.; Koster, R.D. Global assimilation of satellite surface soil moisture retrievals into the NASA Catchment land surface model. *Geophys. Res. Lett.* **2005**, *32*, 4. [[CrossRef](#)]
30. Gruber, A.; Crow, W.T.; Dorigo, W.A. Assimilation of Spatially Sparse In Situ Soil Moisture Networks into a Continuous Model Domain. *Water Resour. Res.* **2018**, *54*, 1353–1367. [[CrossRef](#)]
31. Renzullo, L.J.; van Dijk, A.; Perraud, J.M.; Collins, D.; Henderson, B.; Jin, H.; Smith, A.B.; McJannet, D.L. Continental satellite soil moisture data assimilation improves root-zone moisture analysis for water resources assessment. *J. Hydrol.* **2014**, *519*, 2747–2762. [[CrossRef](#)]
32. Kolassa, J.; Reichle, R.H.; Liu, Q.; Cosh, M.; Bosch, D.D.; Caldwell, T.G.; Colliander, A.; Collins, C.H.; Jackson, T.J.; Livingston, S.J.; et al. Data Assimilation to Extract Soil Moisture Information from SMAP Observations. *Remote Sens.* **2017**, *9*, 1179. [[CrossRef](#)]
33. Zupanski, D.; Zupanski, M. Model error estimation employing an ensemble data assimilation approach. *Mon. Weather Rev.* **2006**, *134*, 1337–1354. [[CrossRef](#)]
34. Turner, M.R.J.; Walker, J.P.; Oke, P.R. Ensemble member generation for sequential data assimilation. *Remote Sens. Environ.* **2008**, *112*, 1421–1433. [[CrossRef](#)]
35. Reichle, R.H.; McLaughlin, D.B.; Entekhabi, D. Hydrologic data assimilation with the ensemble Kalman filter. *Mon. Weather Rev.* **2002**, *130*, 103–114. [[CrossRef](#)]
36. Wu, K.; Shu, H.; Nie, L.; Jiao, Z.H. Triple collocation-based estimation of spatially correlated observation error covariance in remote sensing soil moisture data assimilation. *J. Appl. Remote Sens.* **2018**, *12*, 19. [[CrossRef](#)]
37. Reichle, R.H.; Koster, R.D.; Liu, P.; Mahanama, S.P.P.; Njoku, E.G.; Owe, M. Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR). *J. Geophys. Res. Atmos.* **2007**, *112*, 14. [[CrossRef](#)]
38. Kumar, S.V.; Reichle, R.H.; Koster, R.D.; Crow, W.T.; Peters-Lidard, C.D. Role of Subsurface Physics in the Assimilation of Surface Soil Moisture Observations. *J. Hydrometeorol.* **2009**, *10*, 1534–1547. [[CrossRef](#)]
39. Kumar, S.V.; Reichle, R.H.; Harrison, K.W.; Peters-Lidard, C.D.; Yatheendradas, S.; Santanello, J.A. A comparison of methods for a priori bias correction in soil moisture data assimilation. *Water Resour. Res.* **2012**, *48*, 16. [[CrossRef](#)]
40. Leith, C.E. Theoretical Skill of Monte Carlo Forecasts. *Mon. Weather Rev.* **1974**, *102*, 409. [[CrossRef](#)]
41. Palmer, T.; Mureau, R.; Molteni, F.J.W. The monte carlo forecast. *Weather* **1990**, *45*, 198–207. [[CrossRef](#)]
42. Toth, Z.; Kalnay, E. ensemble forecasting at nmc—The generation of perturbations. *Bull. Am. Meteorol. Soc.* **1993**, *74*, 2317–2330. [[CrossRef](#)]

43. Toth, Z.; Kalnay, E. Ensemble forecasting at NCEP and the breeding method. *Mon. Weather Rev.* **1997**, *125*, 3297–3319. [[CrossRef](#)]
44. Kalman, R.E. A New Approach to Linear Filtering and Prediction Problems. *J. Basic Eng.* **1960**, *82*, 35–45. [[CrossRef](#)]
45. Dee, D.P. Simplification of the kalman filter for meteorological data assimilation. *Q. J. R. Meteorol. Soc.* **1991**, *117*, 365–384. [[CrossRef](#)]
46. Panciera, R.; Walker, J.P.; Jackson, T.J.; Gray, D.A.; Tanase, M.A.; Ryu, D.; Monerris, A.; Yardley, H.; Rudiger, C.; Wu, X.L.; et al. The Soil Moisture Active Passive Experiments (SMAPEX): Toward Soil Moisture Retrieval From the SMAP Mission. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 490–507. [[CrossRef](#)]
47. Wood, E.F. Global scale hydrology—Advances in land surface modeling. *Rev. Geophys.* **1991**, *29*, 193–201. [[CrossRef](#)]
48. O'Neill, P.; Chan, S.; Njoku, E.; Jackson, T.; Bindlish, R.J.V. *SMAP L3 Radiometer Global Daily 36 km EASE-Grid Soil Moisture, Version 6*; NASA National Snow and Ice Data Center Distributed Active Archive Center: Boulder, CO, USA, 2019.
49. Dorigo, W.; Wagner, W.; Albergel, C.; Albrecht, F.; Balsamo, G.; Brocca, L.; Chung, D.; Ertl, M.; Forkel, M.; Gruber, A.; et al. ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sens. Environ.* **2017**, *203*, 185–215. [[CrossRef](#)]
50. Liu, Y.Y.; Dorigo, W.A.; Parinussa, R.M.; de Jeu, R.A.M.; Wagner, W.; McCabe, M.F.; Evans, J.P.; van Dijk, A. Trend-preserving blending of passive and active microwave soil moisture retrievals. *Remote Sens. Environ.* **2012**, *123*, 280–297. [[CrossRef](#)]
51. Albergel, C.; Rudiger, C.; Pellarin, T.; Calvet, J.C.; Fritz, N.; Froissard, F.; Suquia, D.; Petitpa, A.; Pignatelli, B.; Martin, E. From near-surface to root-zone soil moisture using an exponential filter: An assessment of the method based on in-situ observations and model simulations. *Hydrol. Earth Syst. Sci.* **2008**, *12*, 1323–1337. [[CrossRef](#)]
52. Reichle, R.H. Bias reduction in short records of satellite soil moisture. *Geophys. Res. Lett.* **2004**, *31*. [[CrossRef](#)]
53. Calheiros, R.V.; Zawadzki, I. Reflectivity-Rain Rate Relationships for Radar Hydrology in Brazil. *J. Clim. Appl. Meteorol.* **1987**, *26*, 118–132. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).