



## Article

# Improving Soil Water Content and Surface Flux Estimation Based on Data Assimilation Technique

He Chen <sup>†</sup>, Rencai Lin <sup>†</sup>, Baozhong Zhang <sup>\*</sup> and Zheng Wei

State Key Laboratory of Watershed Water Cycle Simulation and Regulation,  
China Institute of Water Resources and Hydropower Research, Beijing 100048, China;  
chenhe@iwhr.com (H.C.); rencaihrs@iwhr.com (R.L.); weizheng@iwhr.com (Z.W.)

\* Correspondence: zhangbz@iwhr.com; Tel.: +86-138-1197-0596

<sup>†</sup> These authors contributed equally to this work.

**Abstract:** Land surface model is a powerful tool for estimating continuous soil water content (SWC) and surface fluxes. However, simulation error tends to accumulate in the process of model simulation due to the inevitable uncertainties of forcing data and the intrinsic model errors. Data assimilation techniques consider the uncertainty of the model, update model states during the simulation period, and therefore improve the accuracy of SWC and surface fluxes estimation. In this study, an Ensemble Kalman Filter (EnKF) technique was coupled to a Hydrologically Enhanced Land Process (HELP) model to update model states, including SWC and surface temperature (Ts). The remotely sensed latent heat flux (LE) estimated by Surface Energy Balance System (SEBS) was used as the observation value in the data assimilation system to update the model states such as SWC and Ts, etc. The model was validated by the observation data in 2006 at the Weishan flux station, where the open-loop estimation without state updating was treated as the benchmark run. Results showed that the root mean square error (RMSE) of SWC was reduced by 30~50% compared to the benchmark run. Meanwhile, the surface fluxes also had significant improvement to different extents, among which the RMSE of LE estimation from the wheat season and maize season reduced by 33% and 44%, respectively. The application of the data assimilation technique can substantially improve the estimation of surface fluxes and SWC states. It is suggested that the data assimilation system has great potential to be used in the application of land surface models in agriculture and water management.

**Keywords:** HELP model; EnKF; soil water content; surface flux; surface temperature; SEBS model



**Citation:** Chen, H.; Lin, R.; Zhang, B.; Wei, Z. Improving Soil Water Content and Surface Flux Estimation Based on Data Assimilation Technique.

*Remote Sens.* **2022**, *14*, 3183.

<https://doi.org/10.3390/rs14133183>

Academic Editor: Gabriel Senay

Received: 11 May 2022

Accepted: 27 June 2022

Published: 2 July 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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 (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Land surface flux is an important index for evapotranspiration (ET) calculation, which is an important parameter of the hydrological cycle [1]. Land surface flux includes surface temperature (Ts), latent heat flux (LE), net radiation (Rn), soil heat flux (G<sub>0</sub>), and sensible heat flux (H). Soil water content (SWC) is a key parameter of the land surface process and plays an important role in the exchange of matter and energy between different components in the earth system [2,3]. At the same time, SWC has a strong effect on ET, which is an important index to measure the degree of soil drought. Timely and accurate estimation of surface energy and SWC is of great significance for drought monitoring, agricultural water resources management, and crop yield prediction.

The land surface model is an effective tool for continuous simulation of SWC and surface fluxes. However, simulation errors tend to accumulate during the operation of the model caused by the input uncertainty as well as continuous model states estimation, and the heterogeneity of the underlying surface also limits the application of the model on the regional scale. With the rapid development of remote sensing technology, a large number of remote sensing observation data have been widely used in the model research

of various fields and disciplines such as atmosphere [4], hydrology [5], agriculture [6], and geology [7]. Remote sensing data have the advantage of continuity in space and dynamic change in time. On the one hand, remote sensing data can make up for the shortcomings of few meteorological observation stations spread, large spatial interpolation error generated, and difficulties in mountainous areas observation. On the other hand, remote sensing data can be used to estimate vegetation, soil, and other parameter information, which can save manpower, money and time spent in carrying out experiments. However, remote sensing observation can only provide these elements at satellite over passing instantaneous time, which limits its usage scenarios. The combination of model estimation and remote sensing observation, i.e., multi-source data fusion is one of the hot issues in the research of remote sensing technology, and the key is how to take advantages of both strategies.

The uncertainty caused by multi-source data fusion is also an important factor affecting the application of remote sensing data in SWC simulation. Many pieces of research focused on reducing the uncertainty of multi-source data, such as data fusion [5], data downscaling [8], sensitivity analysis [9], variability analysis [10–12], and data assimilation [13]. Zhao et al. [14] proposed the SWC downscaling method based on random forest, which is helpful in improving the spatial resolution of passive microwave SWC products and promotes its application at the regional scale. Bai et al. [15] used the enhanced spatial and temporal adaptive reflectance fusion model (ESTARFM) and machine learning algorithm to generate surface SWC at 30 m spatial resolution in Haihe Basin from 2015 to 2017, which is valuable for agricultural water resources management at the field scale. There are many pieces of research on surface SWC estimating using remote sensing technology [2,16–18]. However, in addition to surface SWC, root zone and deep SWC information is also very important for agricultural management. Numerous studies have extended the surface SWC estimated by remote sensing to the deep SWC. It is necessary to establish some hypotheses about vertical distribution of SWC in soil profile. Many studies have used the one-dimensional Richards equation [19]. The prediction of root SWC is greatly affected by the covariance error in the framework of data assimilation, so the assimilation can be successful only when the observation error and model error are both small enough. Sabater et al. [20] compared the ability of two different Kalman filtering methods and two different variational methods to predict root SWC from surface SWC. Baldwin et al. [21] used the SMAR-EnKF method to combine the dynamic process model with SMAP products to obtain SWC of the vertical soil profile by assimilation. These studies provided a good idea for a root or deep SWC estimation. However, due to spatial heterogeneity, it is hard to say whether the formula is effective in the whole region from surface to deep. Thus, it is necessary to use the land surface model to simulate the root zone and deep SWC. If the land surface model is used only, the errors will still accumulate over time, because SWC is not only the output of the model, but also the key state variable and one of the initial conditions affecting the next time step. Therefore, it is necessary to use data assimilation to improve SWC, especially in the deep and root layers.

Data assimilation gives us a tool to optimize model states and thus reduce simulation errors. The essence of data assimilation is error estimation and error simulation, that is, to correct the model simulation process by introducing observation data of time series so as to improve model simulation accuracy and reduce model uncertainty [22]. In recent years, data assimilation technology has been developed rapidly and gradually introduced into the research of hydrology, such as SWC assimilation [19,23] and Ts assimilation [24]. These studies proved that data assimilation technology has great potential to improve the accuracy of hydrological prediction. At the same time, due to the extensive application of remote sensing technology, more and more surface parameters can be obtained from remote sensing data, which provides data support for the application of data assimilation in hydrological models. Data assimilation methods mainly include variational method and filtering method (Table 1). The variational method is widely used in meteorology, oceanography and other fields because it does not need preheating and has good assimilation results in a short computing time. Filtering method can take

into account the uncertainties of different source data and models, and the errors can propagate over time, which widely used in the field of hydrology.

**Table 1.** Comparison of commonly used data assimilation methods.

Method	Variational Method	Filtering Method
Fundamentals	Bayesian theory	Best-fitting regression
Objective function	The posterior probability density is the highest	The error covariance matrix is minimal
Basic assumption	Model errors do not propagate over time	The uncertainties of the observed values are known
Advantage	No need to warm up; accurate calculation in complex systems	The calculation result is accurate after preheating; easy Integration
Disadvantage	Heavy calculation burden	It cannot be solved under ultra-high nonlinear condition.

Kalman filter algorithm was proposed by Kalman [25] in 1960. Its basic idea is that the model state is predicted first, then the observation data are introduced, and the model state is re-analyzed according to the observation data at the end. Komma et al. [26] integrated the EnKF method into a conceptual hydrological model and applied it to small watersheds. The prediction accuracy of the hydrological model is improved by optimizing the initial state variables of the hydrological model. Thirel et al. [27] built a data assimilation system in France. By optimizing the state variables of the previous hydrological model, the optimized state variables are used as the initial state variables of the forecast model. In terms of SWC assimilation, Walker et al. [28] assimilated surface soil moisture, indicating that data assimilation could improve the simulation accuracy of SWC when the initial value of the model deviated greatly from the measured value. Crow et al. [29] showed that assimilation of surface soil moisture could improve the simulation accuracy of soil moisture and surface latent heat flux when precipitation data error is large. Chen et al. [30] conducted a three-layer SWC assimilation in the SWAT model, and the results showed that surface soil moisture assimilation had the best performance. In terms of ET assimilation, Qin et al. [31] used LE simulated by the SEBS model as an observation value to assimilate the WEP-L distributed hydrological model, however, it was difficult to be verified. In addition to the single-objective assimilation mentioned above, there is also multi-objective joint assimilation. Pan et al. [32] used the Constrained EnKF (CoEnKF) method to assimilate surface and root soil moisture, latent heat, and runoff in variable infiltration capacity (VIC) model. According to the correlation between the error level of the water balance items and the non-closure terms, the non-closure terms are allocated. The closure of the hydrological process is realized.

It is found that there is no unified evaluation method for model improvement in existing studies. In the simulation of different hydrological factors, the stability of the assimilation method is different. At the same time, it is also worth discussing whether the model simulation for long time series is effective.

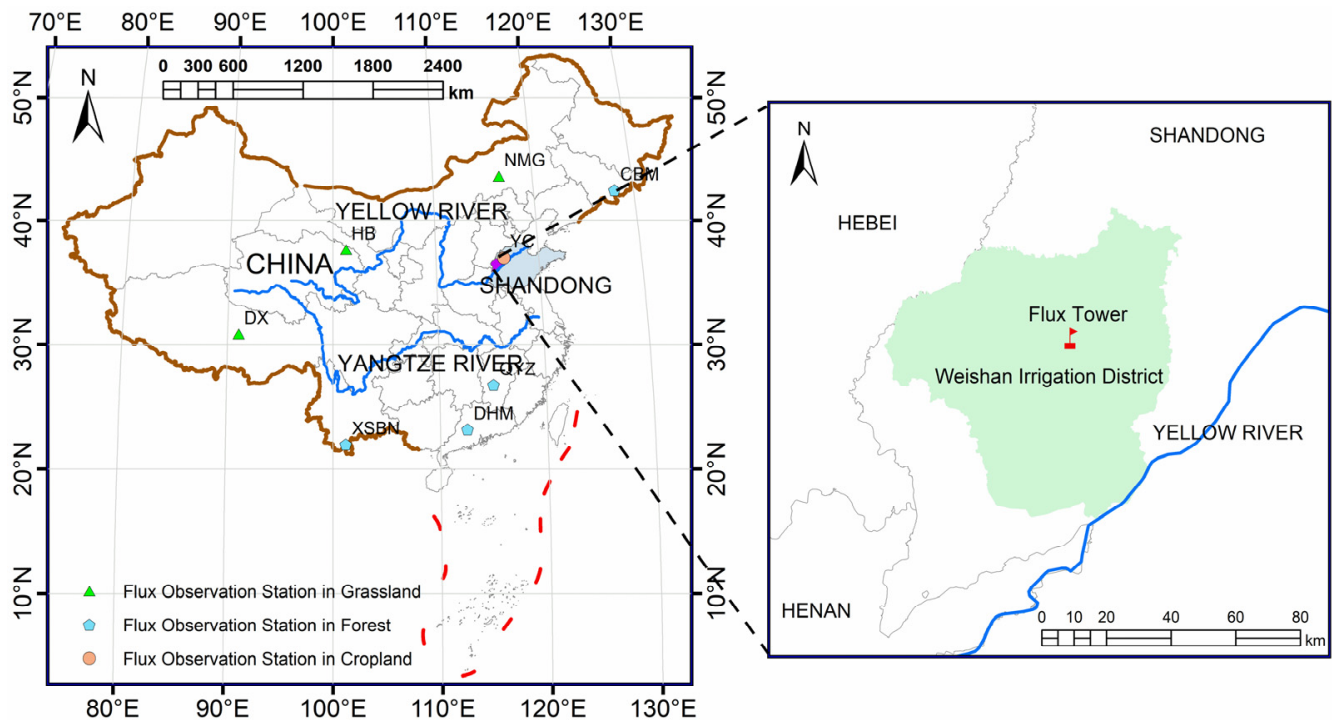
In this study, an Ensemble Kalman Filter (EnKF) technique was coupled to a Hydrologically Enhanced Land Process (HELP) model to update model states, including SWC and Ts. The remotely sensed LE estimated by Surface Energy Balance System (SEBS) was used as the observation value in the data assimilation system to update the model states such as SWC and Ts, etc. The model was validated by the observation data in 2006 at the Weishan flux station, where the open-loop estimation without state updating was treated as the benchmark run.

## 2. Materials and Methodology

### 2.1. Study Site

Weishan Irrigation District (36°8′N~37°1′N, 115°25′E~116°31′E) is located in the lower reaches of the Yellow River and in the middle of the North China Plain. It is the fifth largest irrigation district in China, with an irrigation area of about 3600 km<sup>2</sup>. The irrigation district

belongs to the temperate monsoon sub-humid climate area. More than 80% of the land use in the irrigation district is farmland, followed by construction, with a small amount of forest and water. The performance of the data assimilation system was evaluated at Weishan flux station ( $36^{\circ}38'55.5''\text{N}$ ,  $116^{\circ}3'15.3''\text{E}$ ) located in the irrigation district, which is representative of this area (Figure 1).



**Figure 1.** The geographical location of the study area. Changbai Mountain (CBM); Dinghu Mountain (DHM); Dangxiang (DX); Haibei (HB); Neimenggu (NMG); Qianyanzhou (QYZ); Xishuangbanna (XSBN); Yucheng (YC).

The annual pan evaporation (1950 mm measured by 20 cm diameter evaporation pan) significantly exceeds the annual precipitation (532 mm), and the annual mean air temperature ( $T_a$ ) is  $13.3\text{ }^{\circ}\text{C}$ . The dominant crops are winter wheat and summer maize planted in rotation, representing the main cropping system in this region. The rainfall is insufficient to support the winter wheat production, as about two-thirds of the annual rainfall occurs from July to September. Thus, agricultural production of the wheat-growing season relies primarily on surface ditch irrigation, with the irrigation water drawn mainly from the Yellow River [33].

## 2.2. Data

The data used in our study include ground measurements and remote sensing data (MODIS). Ground measurements mainly include observation data of flux station and meteorological station, including flux,  $T_s$ , SWC,  $T_a$ , air pressure, rainfall, etc. The remote sensing data used included MOD09GA, MOD13Q1, and MOD11A1.

### 2.2.1. Ground Measurements

For ground measurements, a 10 m flux tower was set, and measurement instruments were equipped on and around the tower. The model input and validation data needed in the study, including flux data, soil data, meteorological data, and vegetation parameters, were observed. The LE and sensible heat flux ( $H$ ) was measured by an eddy-covariance system at the height of 3.7 m. Downwelling shortwave radiation ( $R_{s\_down}$ ), downwelling longwave radiation ( $R_{l\_down}$ ), and upwelling longwave radiation ( $R_{l\_up}$ ) were measured by a radiometer at the height of 3.5 m. Soil heat flux ( $G$ ) was measured by



two heat transducers on opposite side of the tower at a depth of 0.03 m. Soil temperature ( $T_{\text{soil}}$ ) and SWC were measured at the depth of 0.05, 0.10, 0.20, 0.40, 0.80, and 1.6 m. Meteorological data used in this study were recorded at 10 min intervals, including  $T_a$  and relative humidity (RH), wind speed (WS), and precipitation (P). Skin temperature was measured with an infrared thermometer. The canopy height and leaf area index (LAI) were measured twice per month at several locations surrounding the tower by randomly selecting samples within the cropland. Table 2 shows the observation projects of observation station.

**Table 2.** Observation projects of flux station.

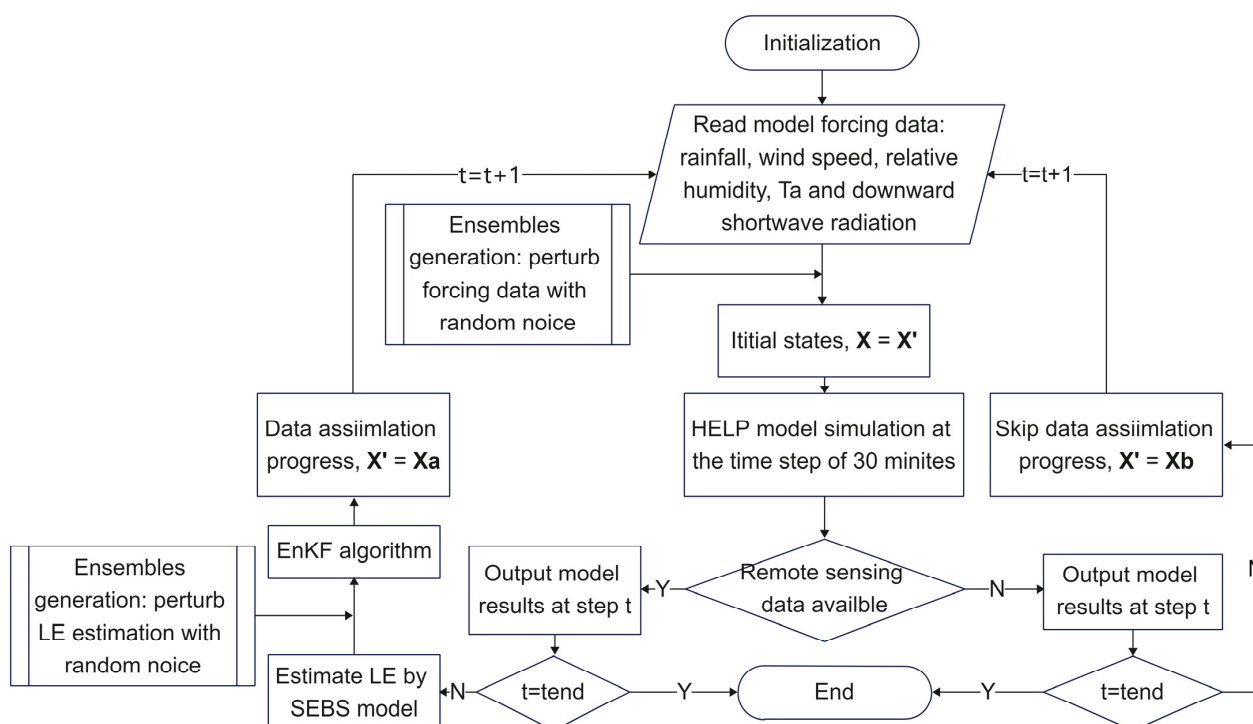
Observation Parameters	Instrument	Height/Depth (m)	Frequency of Storage (min)
LE	CSAT3, Campbell Scientific, Logan, UT, USA	3.7	30
H	LI7500, LI-COR, Lincoln, NE, USA	3.7	30
$R_{s\_down}$	CNR-1, Kipp & Zonen, Delft, The Netherlands	3.5	10
$R_{l\_down}$	CNR-1, Kipp & Zonen, Delft, The Netherlands	3.5	10
$R_{l\_up}$	CNR-1, Kipp & Zonen, Delft, The Netherlands	3.5	10
G	HFP01SC, Hukseflux, The Netherlands	0.03	10
$T_{\text{soil}}$	Campbell-107, Campbell Scientific Inc., Logan, UT, USA	−0.05, −0.1, −0.2, −0.4, −0.8, −1.6	10
SWC	TRIME-EZ/IT, IMKO, Ettlingen, Germany	−0.05, −0.1, −0.2, −0.4, −0.8, −1.6	10

### 2.2.2. Remote Sensing Data

For remote sensing data, land products of MODIS on-boarded Terra and Aqua satellite were used in this study, and the over-passing time of each satellite is around 10:30 am and 13:30 pm. (<http://reverb.echo.nasa.gov/reverb/>, accessed on 10 May 2022). Collection 5 data onboard the Terra satellite and corresponding products onboard the Aqua satellite was used in this study [34]; broadband surface albedo [35],  $T_s$  and emissivity, and NDVI [36] were thus derived. Detailed introduction of above could refer to the website (<http://modis-land.gsfc.nasa.gov/>, accessed on 10 May 2022). All the data were re-projected from the Sinusoidal (SIN) projection into the Universal Transverse Mercator (UTM) projection and resampled to 250 m.

### 2.3. Methodology

The core component of land surface process data assimilation system is hydrological model or land surface model, data assimilation algorithm, and an independent source of observation to provide extra information. In this study, the Ensemble Kalman Filter (EnKF) algorithm was embodied in the HELP model. The estimated LE based on remote sensing using Surface Energy Balance System (SEBS) was used as an independent observation source. The framework of the scheme is shown in Figure 2.



**Figure 2.** Workflow of this research.

### 2.3.1. Hydrologically Enhanced Land Process (HELP) Model

HELP model is based on Simple Biosphere Model 2 (SiB2), which enhances runoff process and other hydrological processes modeling. SiB2 model was introduced by Sellers et al. [37], based on the Simple Biosphere Model (SiB), into the photosynthesis-stomatal conductance sub-module. Thus, the surface energy and water balance can be better simulated. SiB2 consists of one vegetation layer and three soil layers, mainly predicting three layers of SWC and temperature, sensible and latent heat fluxes.

In HELP model, radiation transfer, energy, and carbon exchange between land surface and atmosphere is the same as SiB2, while below is the improvement in HELP model: (1) soil in unsaturated zone was divided into 0.1 m layer, and SWC exchange in each layer was described by one-dimensional Richards equation; (2) relationship between unsaturated and saturated hydraulic conductivity was described using van Genuchten equation instead of Cambell/Clapp-Hornberger equation; (3) interflow and groundwater discharge and water exchange between SWC and river way were added to the surface runoff model, described by mass conservation equation and Darcy's law.

In order to enhance the simulation of SWC, unsaturated soil layer was further refined to 10 layers. Thus, the model has 17 state variables in total as: 0~10 cm layer SWC, canopy interception, land surface precipitation storage, temperature at canopy, land surface and deep soil layer, groundwater table, and river water level.

Except for meteorological input data derived from site measurement, HELP model had a set of parameters to describe the physical characteristics of soil and physiological property. Vegetation parameters included vegetation form, optical parameters, and physiological parameters, which were determined by crop time and invariant with time. The vegetation parameters were derived by field experiment at Weishan flux station. The soil parameters are listed in Table 3.

**Table 3.** The characteristic parameters of different soil layers at Weishan flux station.

Soil Layer (cm)	Saturation Capacity ( $\text{m}^3 \cdot \text{m}^{-3}$ )	Residual Humidity ( $\text{m}^3 \cdot \text{m}^{-3}$ )	Saturation Hydraulic Conductivity ( $\text{mm} \cdot \text{h}^{-1}$ )	Van Genuchten Parameter, $\alpha$ ( $\text{cm}^{-1}$ )	Van Genuchten Parameter, $n$
0–10	0.424	0.033	55.0	$1.74 \times 10^{-2}$	1.409
20–30	0.393	0.080	-	$6.7 \times 10^{-3}$	1.769
100–110	0.435	0.148	3.2	$5.3 \times 10^{-3}$	2.539

### 2.3.2. Ensemble Kalman Filter (EnKF)

The data assimilation approach used in this study was EnKF, proposed by Evensen [38]. The basic idea of EnKF incorporates an ensemble of forecasts to estimate background error covariance, which is a Monte Carlo approximation to the traditional Kalman filter [39]. Data assimilation using EnKF method consists of two steps: (1) ensemble forecast; (2) model states update. At each time step  $t$ , a set of ensemble samples of model state variables is generated using Monte Carlo approach, noted as  $\mathbf{x}_{i,t}^b$ , where subscript  $i$  indicates the  $i$ th sample, superscript  $b$  indicates states before update. The dimension of vector  $\mathbf{x}$  is  $m$ , indicated the number of model states is  $m$ . Model forecast at initial time can be expressed as:

$$\mathbf{x}_{i,t=1}^b = M(\mathbf{x}_{i,0}), \quad i = 1, \dots, n \quad (1)$$

where  $\mathbf{x}_{i,0}$  is the initial states of the model;  $M$  indicates the modeling operator;  $n$  is the ensemble size. The update equation for each ensemble member of EnKF is:

$$\mathbf{x}_i^a = \mathbf{x}_i^b + \mathbf{K}(\mathbf{y}_i - \mathbf{H}\mathbf{x}_i^b), \quad i = 1, \dots, n \quad (2)$$

$$\mathbf{K} = \mathbf{P}^b \mathbf{H}^T (\mathbf{H} \mathbf{P}^b \mathbf{H}^T + \mathbf{R})^{-1} \quad (3)$$

$$\mathbf{P}^b = \frac{1}{n-1} \mathbf{x}'^b (\mathbf{x}'^b)^T \quad (4)$$

$$\mathbf{x}'^b = (\mathbf{x}'_1^b, \dots, \mathbf{x}'_n^b) \quad (5)$$

where  $\mathbf{K}$  is the Kalman gain;  $\mathbf{x}^a$  is the analysis after the update;  $\mathbf{y}_i$  is the vector of the observations with the dimension of  $p$ ;  $\mathbf{P}^b$  is the background error covariance matrix;  $\mathbf{H}$  is the operator that converts the model states to the observation space with the dimension of  $p \times n$ ;  $\mathbf{R}$  is the observation error covariance matrix with the dimension of  $p \times p$ ;  $\mathbf{x}'_i^b$  is the ensemble anomaly for the  $i$ th ensemble member.

At model state update stage, each  $i$  of the  $m$  ensemble members is updated by Equation (2). For each ensemble member, the  $p$  dimensional observation  $\mathbf{y}_i$  is sampled from a distribution with the mean equal to the observation and variance of  $\mathbf{R}$ .

### 2.3.3. Surface Energy Balance System (SEBS) Model

Instantaneous LE at satellite over-passing time estimated by Surface Energy Balance System (SEBS) was treated as the observation of the data assimilation system, i.e., the vector  $\mathbf{y}$  in Equation (2). SEBS is a single layer model based on energy balance theory that uses an excess resistance term which accounts for the difference of the roughness lengths for momentum and heat between vegetation and surface [40], which consists of four components: (1) land surface physical parameters from spectral reflectance and radiance; (2) roughness length for heat transfer by extended model; (3)  $H$  iterated from Monin-Obukhov similarity theory; (4) LE determined by evaporative fraction based on energy balance equation at limiting cases [41]. Compared with other single layer models based on energy balance theory, SEBS has an advantage in the independent calculation of each pixel, i.e., even if partial pixels are contaminated because of cloud or rainfall on one day, other pixels exposed under clear day still can be used to calculate energy fluxes. The

merit of SEBS can maximize the usage of remote sensing data and will provide observation data to data assimilation system to the greatest extent.

The surface energy balance equation [42] can be written as:

$$R_n = G_0 + H + LE \quad (6)$$

where  $R_n$  is the net radiation,  $W \cdot m^{-2}$ ;  $G_0$  is the soil heat flux,  $W \cdot m^{-2}$ ;  $H$  is the sensible heat flux,  $W \cdot m^{-2}$ ;  $LE$  is the latent heat flux,  $W \cdot m^{-2}$ . The parameterization scheme of roughness length of momentum transfer ( $z_{0m}$ ) and zero displacement height ( $d$ ) were derived from empirical relationship given by Shaw et al. [43] from the second-order closure theory [44] to suit the situation in China.

$$z_{0m} = \begin{cases} z_{0s} + 0.3hZ^{1/2}, & 0 \leq Z \leq 0.2 \\ 0.3h(1 - d/h), & 0.2 \leq Z < 1.5 \end{cases} \quad Z = C_d LAI \quad (7)$$

$$d = 1.1h \ln(1 + Z^{1/4}) \quad (8)$$

where  $h$  is the canopy height; LAI is the leaf area index;  $C_d$  is the drag coefficient of the foliage elements (set to 0.2);  $z_{0s}$  is the substrate roughness height (set to 0.01 m for bare soil) [45].

The model was already estimated using MODIS data at the study site and validated by in situ measurement of eddy covariance system. The results showed that relative error and root mean square error of LE in both wheat and maize season was within 20%, indicating SEBS model using remote sensing data had good accuracy in estimating LE. Detailed results of validation and discussion could refer to Yang et al. [46].

#### 2.3.4. Implementation of Data Assimilation

A key issue in implementing data assimilation method is quantifying the covariance error matrices. For EnKF, variance between multiple ensemble members is the way to quantify the model error. In this study, Monte Carlo sampling through assigned random noise in the forcing data and observation were derived from emulating the model error, which is proved to be an easy and robust way; moreover, the physical basis and interpretation of individual error sources remain clear [26].

The forcing data of HELP model include P, WS, RH, Ta, and  $R_{s\_down}$ . The ensembles of forcing data were emulated by perturbing each forcing data at each time step, which were generated as Clark et al. [47]:

$$f' = f(1 + \gamma_f), \quad \gamma_f \sim N(0, \omega_f) \quad (9)$$

where  $\gamma_f$  is the random noise with a covariance of  $\omega_f$ , i.e.,  $\gamma_f$  is a realization from a normal distribution ranging from  $-\omega_f$  to  $\omega_f$ , in which the subscript of  $f$  indicates different sources of forcing data.

Similarly, the data assimilation approach implicates an assumption of knowing or at least accurately estimating observation error. The ensembles of observation were also generated as follows:

$$y' = y(1 + \gamma_y), \quad \gamma_y \sim N(0, \omega_y) \quad (10)$$

where the symbols have the same meaning as in Equation (8), in which the subscript of  $y$  indicated observation in the data assimilation system, i.e., the LE estimated by SEBS. By generating each ensemble of forcing data and observation, a set of ensemble model states were calculated by HELP model, and the model error covariance and observation error covariance ( $P^b$  and  $R$  in Equation (3)) were thus quantified. Considering the actual error of in situ measurements and SEBS simulation, random distribution of forcing data and observation data in HELP model is listed in Table 4.

**Table 4.** Random distribution of forcing and observation data of data assimilation system.

Input	Forcing Data					Observation
	P (mm)	WS (m/s)	RH	Ta (°C)	R <sub>s_down</sub>	
$\omega$	0.2	0.3	0.1	0.15	0.1	0.2

The HELP model has 17 model states, in which the canopy interception and surface interception are the states only relative to rainfall, which means the simulation error of these two states does not accumulate with time. Moreover, groundwater level and river water level are also influenced by neighboring pixels, which can hardly be accurately simulated by one-dimensional vertical model. Therefore, in the data assimilation system, only 10 layers of SWC and three layers of temperature were updated. According to previous studies, comprehensively considering the computational requirements and data assimilation performance, the recommended ensemble size is 10 times of model states [32,48], i.e., 130 in this study. The soil depth of surface, root, and deep layer was 0–10, 20–30, and 100–110 cm, respectively. For the convenience of description, SWC corresponding to surface, root, and deep layer are represented by SWCs, SWCr, and SWCd, respectively.

The time step of HELP model is 30 min, and the time step of SEBS is two times per day at 10:30 and 13:30 under clear day, which means the observation is not available at each time step of HELP model. In this study, the model states were updated by EnKF when LE estimation was available; otherwise, the model was free running. The HELP model ran from 1 October 2005 to 31 December 2005 freely with initial SWC, and temperature values originated from ground measurements to warm up the model, then the model assembled data assimilation system ran from 1 January 2006 to 31 December 2006, while the performance of data assimilation was tested in wheat growing season (from 1 March to 31 May) and maize growing season (from 1 July to 30 September).

### 2.3.5. Statistical Metrics

Model assessment consisted of the standard statistical evaluations, including *bias*, mean absolute error (MAE), root mean square error (RMSE), determination coefficient ( $R^2$ ), and effectiveness criterion (*Eff*). The statistical metrics are calculated as follows:

$$bias = \frac{1}{n} \sum_{i=1}^n (S_i - O_i) \quad (11)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |S_i - O_i| \quad (12)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (S_i - O_i)^2} \quad (13)$$

$$R^2 = \frac{\sum_{i=1}^n (S_i - \bar{S})^2}{\sum_{i=1}^n (O_i - \bar{S})^2} \quad (14)$$

$$Eff(\%) = 100 \cdot \left( 1 - \frac{\sum_{i=1}^n (S_i^u - O_i)^2}{\sum_{i=1}^n (S_i^b - O_i)^2} \right) \quad (15)$$

where superscript *u* and *b* represent the update after data assimilation and benchmark run, respectively; *O* and *S* represent the observation and simulation, respectively;  $\bar{S}$  is the average of simulated values. The *Eff* ranges between negative infinity and 100%, and a value larger than 0 corresponds to a positive impact in skill resulting from the data assimilation. Conversely, a value below 0 corresponds to a negative impact. The greater



value of  $Eff$  indicates better performance following data assimilation compared to the benchmark run.

### 3. Results

The data assimilation approach used in this study was EnKF. It was used to reduce the simulation error of SWC and surface flux. The data assimilation results were compared with the results of the benchmark run, and the observation data of the Weishan flux station were used for verification. In terms of SWC assimilation, we also evaluated the effect of root and deep SWC assimilation. The results showed that the model could simulate SWC effectively, and the data assimilation technology can effectively reduce the error of SWC simulation. At the same time, surface flux assimilation also achieved encouraging results.

#### 3.1. Benchmark Run

In order to evaluate the performance of the data assimilation approach, the results from the benchmark run are needed, i.e., the open-loop simulation with the same set of forcing data, parameters, and initial states yet without updating model states. Figure 3 shows the simulation results of SWC in three layers by benchmark run. The *bias* of SWCs, SWCr, and SWCd is 0.033, 0.047, and  $-0.009 \text{ m}^3 \cdot \text{m}^{-3}$ , respectively. Although the entire error is relatively small, the simulation does not fit well with the observation value, the MAE of the three layers from top to bottom is 0.171, 0.163, and  $0.147 \text{ m}^3 \cdot \text{m}^{-3}$ , respectively, and the RMSE is 0.055, 0.053 and  $0.053 \text{ m}^3 \cdot \text{m}^{-3}$ . Figure 2 showed the simulation fluctuated violently a few days before and after the rainfall or irrigation, which indicated that SWC is a very sensitive state to the rainfall and irrigation in the HELP model. The SWCs had a severe fluctuation; moreover, SWCr and SWCd also had an obvious increase after rainfall or irrigation.

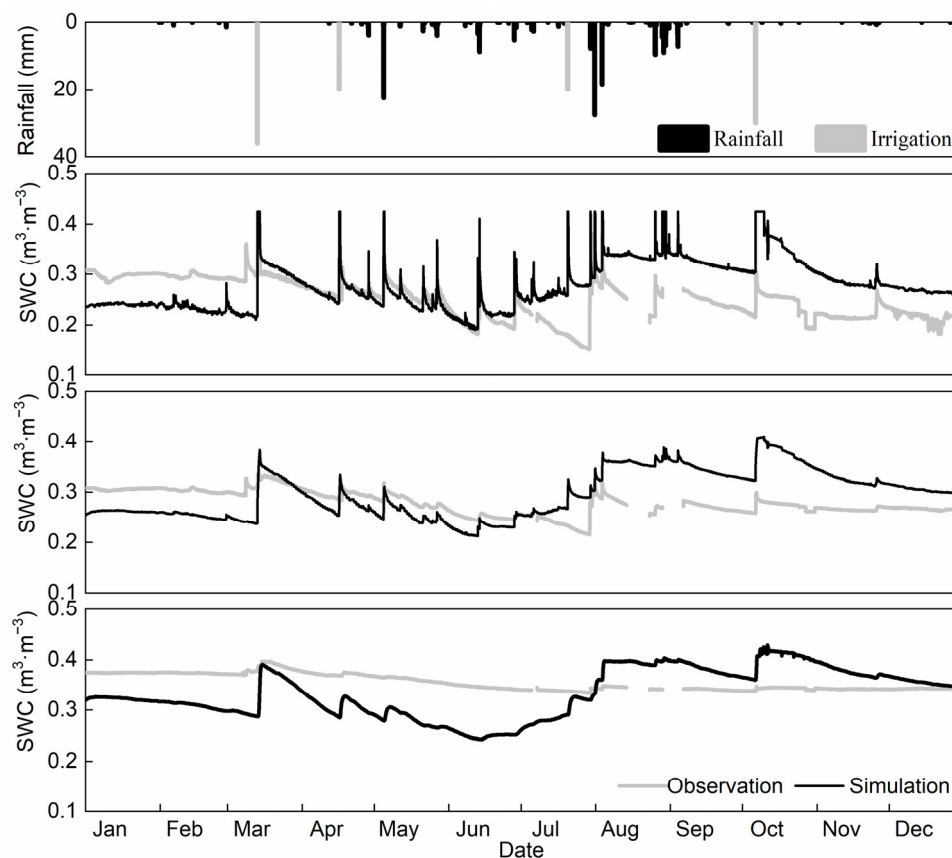
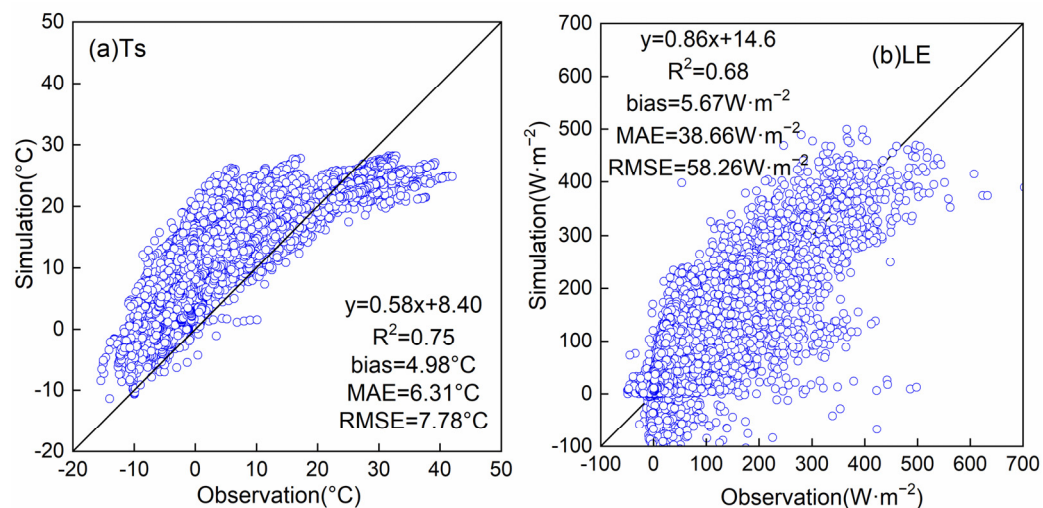


Figure 3. Comparison of SWC simulation in three layers to observation over benchmark run.

Figure 4 shows the simulation and observation of the Ts and LE. The simulation of Ts had a systematic error with a *bias* of 4.98 °C, the MAE was 6.31 °C, and the RMSE was 7.78 °C, but the simulation fitted well with the observation with the  $R^2$  of 0.75. The simulation of heat fluxes was validated using the observation from the eddy covariance system. The H and LE observed by the eddy covariance system were corrected by the energy balance closure scheme [49]. The relative errors of  $R_n$ ,  $G$ ,  $H$ , and  $LE$  are  $-17.8\%$ ,  $103.9\%$ ,  $-0.5\%$ , and  $9.0\%$ , respectively. There was a large error in the simulation of  $G$ , but the absolute value of  $G$  is small relatively and the ratio of  $G$  and  $R_n$  is small, thus the simulation of  $LE$  does not have an obvious systematic error, but the random error is large, the RMSE is  $58.26 \text{ W}\cdot\text{m}^{-2}$  with an  $R^2$  of 0.68.



**Figure 4.** Scatter plots of Ts and LE over benchmark run.

The simulation error of the HELP model is the combination of error in input data, error in calibrated parameters, and error from the model structure. The input error is the reflectance of instrument measurement uncertainty and/or tempo-spatial heterogeneous land water cycle; the calibrated parameters error is the compromise of empirical relationship and the complex and inenarrable physical mechanism; model structure error is the simplification of equations describing the hydrological processes. These errors are inevitable and can hardly be divided into different sources, which have different influences on the simulation results [50]. The SWC is sensitive to rainfall and irrigation; thus, the interpolation error of rainfall or non-uniform irrigation may cause large errors in the simulation of SWC. The systematic error of Ts may be caused by over-calibrated to the insensitive parameter, or model structure error, while the random error in the simulation of LE may be due to the difficulty in generalizing the evaporation process.

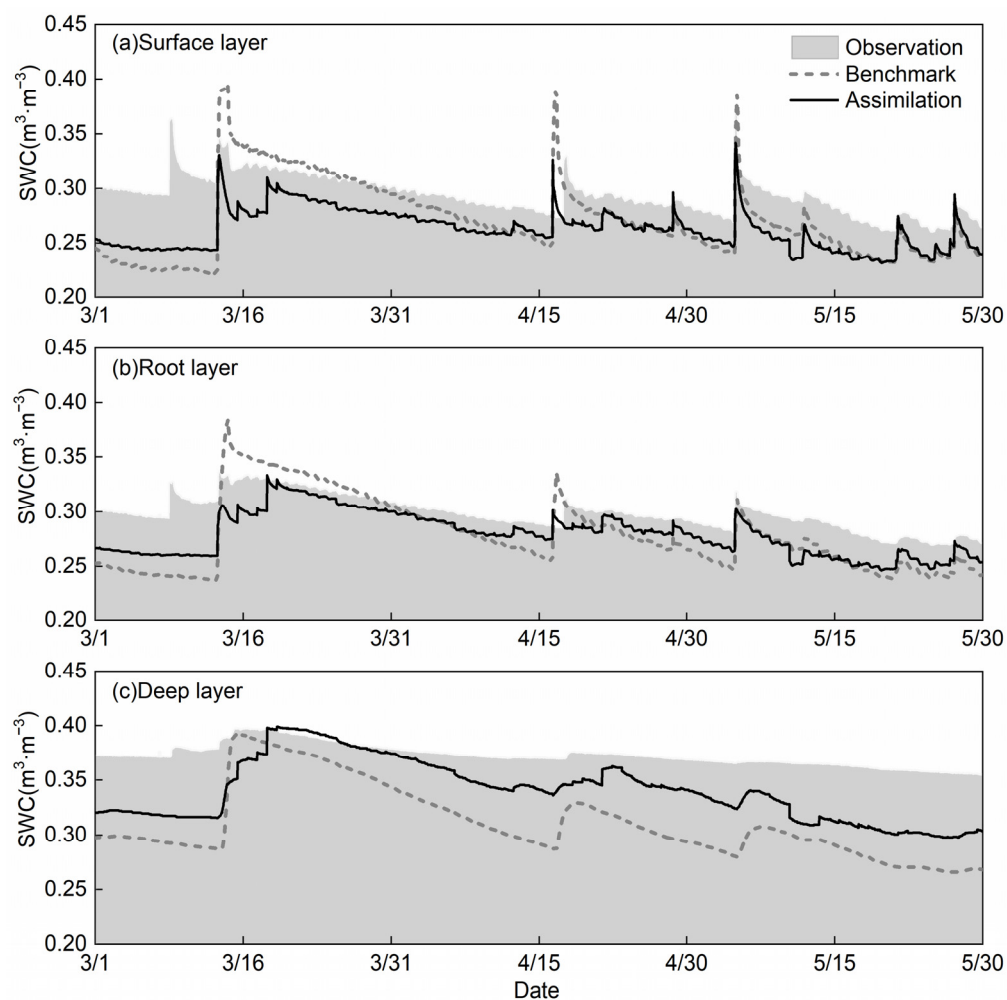
### 3.2. Data Assimilation in SWC Estimation

Figures 5 and 6 are the comparison of SWC under data assimilation run, benchmark run, and observation from the wheat season (1 March to 31 May) and maize season (1 July to 30 September), respectively. Table 3 shows the error statistics of the data assimilation run and benchmark run. The results show that compared with the benchmark run, the updated SWC simulation in both wheat season and maize season have different degrees of improvement; the *Eff* ranges from 9.31% to 74.17%.

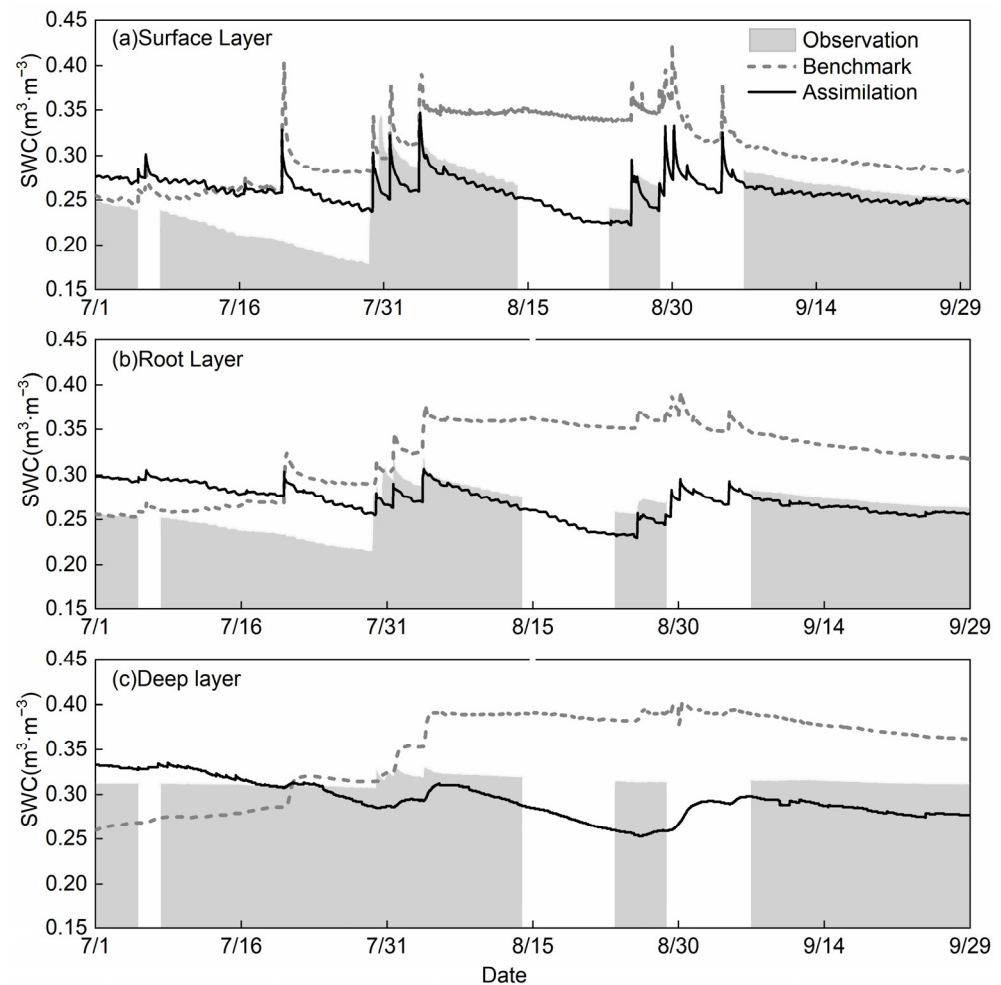
From the statistical results, we can see that in the wheat growing season, the *bias* and MAE of updated SWCs are slightly larger than the benchmark run. The reason is the simulation of SWCs from benchmark run changes severely before, and after the precipitation/irrigation, i.e., the SWCs rise sharply after the precipitation/irrigation while reducing significantly in a short period of time. Therefore, from the perspective of error statistics, during the whole growing season, the error is averaged out, and the *bias* and

MAE remain at a lower level. However, considering the simulation process during the growing season, the time series of updated results from data assimilation fits well with the observation curve, with a slightly systematic underestimation. Compared with the benchmark run, the RMSE from the data assimilation run reduces from 0.04 to 0.03  $\text{m}^3 \cdot \text{m}^{-3}$ . The *Eff* of SWCs is 9.31%, which means after the data assimilation, the simulation results of SWCs have improved to some extent.

The SWCr and SWCd of the wheat growing season have great improvement after data assimilation compared with the benchmark run, with an *Eff* of 43.68% and 66.67%, respectively. From the perspective of error statistics from the whole growing season, the *bias* of SWCr and SWCd reduced by 0.0032 and 0.0782  $\text{m}^3 \cdot \text{m}^{-3}$ , respectively, and the MAE reduced by 0.0103 and 0.0754  $\text{m}^3 \cdot \text{m}^{-3}$ , respectively. The SWCr was also influenced by precipitation/irrigation. The time series after data assimilation has a similar pattern as the surface layer; the simulation from the benchmark run shows a severe changing pattern, which might attribute to the recession process of the HELP model. After the data assimilation, the time series of SWC fits much better with the observation, with an RMSE reduced from 0.03 to 0.02  $\text{m}^3 \cdot \text{m}^{-3}$ . The SWCd did not receive much influence by precipitation/irrigation compared with surface and root layers, but the simulation from the benchmark run shows an obvious underestimation compared with the observation. The data assimilation reduces the systematic error, with an RMSE reduced from 0.07 to 0.04  $\text{m}^3 \cdot \text{m}^{-3}$ , which shows a large improvement compared with the benchmark run.



**Figure 5.** Simulation and observation of SWC in three layers in wheat season.



**Figure 6.** Simulation and observation of SWC in three layers in maize season.

Compared with the wheat growing season, the data assimilation results in the maize growing season performed better. The MAE of SWCs, SWCr, and SWCd are reduced by 0.082, 0.1082, and 0.0744  $\text{m}^3 \cdot \text{m}^{-3}$  compared with the benchmark run, respectively (Table 5). According to the simulation process, the time series simulation of SWC in three layers from the benchmark run show an overestimation, especially for SWCs and SWCr, with a *bias* of 0.1873 and 0.1997  $\text{m}^3 \cdot \text{m}^{-3}$ , respectively. After data assimilation, the systematic error reduces to a large extent, and the *bias* of the top two layers of SWC reduced to 0.037 and 0.0362  $\text{m}^3 \cdot \text{m}^{-3}$ , respectively. However, after data assimilation, the time series of simulated SWCd showed an underestimation compared with the benchmark run; the *bias* changed from 0.0927 to  $-0.0545 \text{ m}^3 \cdot \text{m}^{-3}$ . The reason is according to the HELP model, the SWCd tended to have a similar pattern with the surface and root layer since the infiltration rate in the Richards equation is constant, and this led to the results that the simulated SWCd is influenced by precipitation and/or irrigation to a large extent. Compared with the wheat growing season, the observed SWCd in the maize season roughly remained constant regardless of precipitation or irrigation because it has a longer infiltration time in a wheat-maize rotation system [51]. The RMSEs of the three layers were also reduced by 0.03  $\text{m}^3 \cdot \text{m}^{-3}$ , 0.03  $\text{m}^3 \cdot \text{m}^{-3}$ , and 0.02  $\text{m}^3 \cdot \text{m}^{-3}$ , respectively. Overall, the data assimilation *Eff* of the three layers is above 60%, which means the simulation results from the data assimilation run fit better with the observation.

**Table 5.** The characteristic parameters of SWC at Weishan flux station.

Evaluation Criteria		Wheat Season			Maize Season		
		SWCs	SWCr	SWCd	SWCs	SWCr	SWCd
<i>bias</i> ( $\text{m}^3 \cdot \text{m}^{-3}$ )	benchmark	−0.0757	−0.0664	−0.1647	0.1873	0.1997	0.0927
	assimilation	−0.1045	−0.0632	−0.0865	0.037	0.0362	−0.0545
MAE ( $\text{m}^3 \cdot \text{m}^{-3}$ )	benchmark	0.0958	0.0842	0.1647	0.19	0.1998	0.1486
	assimilation	0.1057	0.0639	0.0883	0.108	0.0916	0.0742
RMSE ( $\text{m}^3 \cdot \text{m}^{-3}$ )	benchmark	0.04	0.03	0.07	0.06	0.06	0.05
	assimilation	0.03	0.02	0.04	0.03	0.03	0.03
<i>Eff</i> (%)		9.31	43.68	66.67	63.17	74.17	72.34

Results showed that, in the wheat growing season, the improvement of data assimilation was mainly demonstrated in the random error, while in the maize growing season, the improvement was mainly demonstrated in the systematic error. Overall, the degree of improvement increased from the top layer to the bottom layer.

### 3.3. Data Assimilation in Surface Flux Estimation

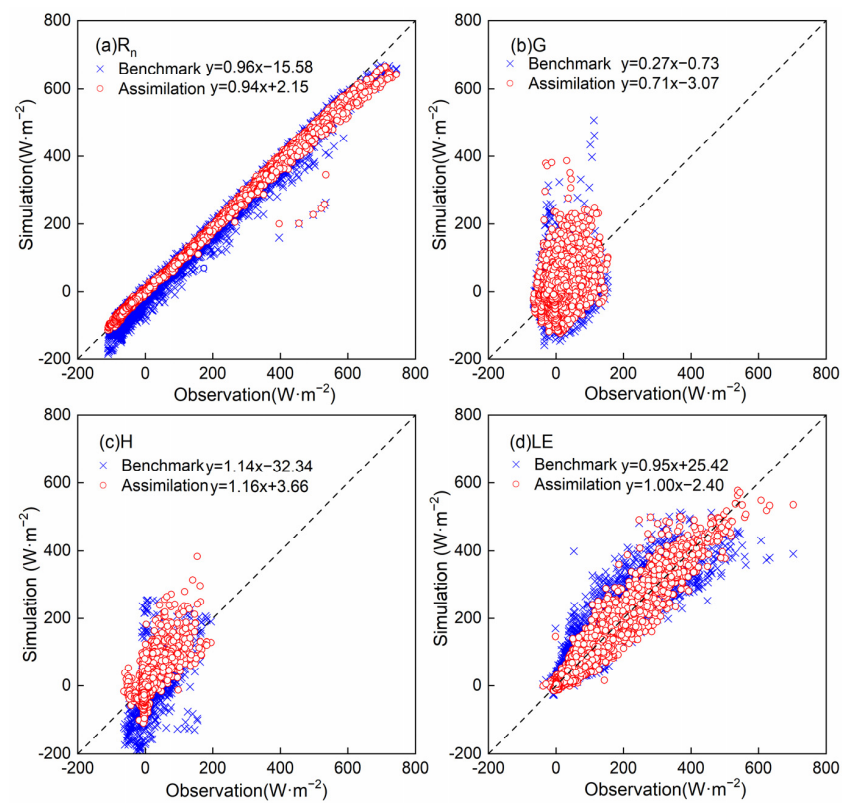
In this study, the estimated LE based on remote sensing using the SEBS model was used as the observation operator in the data assimilation system. However, further verification of data assimilation performance in the surface flux simulation is also needed. Since the background error covariance estimation of the assimilation system is calculated by the sampling error matrix consisting of model states. Therefore, using LE as an observation operator in the assimilation system can only update the model states at observation time, and the surface flux as the model output will be simulated at the next time step using updated model states as the initial condition. Moreover, the assimilation procedure only happens at remote sensing LE estimation is available, i.e., satellite over passing time on clear days. In other words, using LE as an observation operator does not guarantee the improvement of LE simulation results.

Figures 7 and 8 are the scatter plots of surface flux simulation compared with an observation by benchmark run and data assimilation run during wheat and maize growing season, respectively (error statistics see Table 6). In general, compared with the benchmark run, four components of surface heat flux simulation in both wheat and maize season performed better in various degrees. For the wheat growing season, after data assimilation, the relative error of Rn, G, H, and LE simulation was substantially reduced to under  $10 \text{ W} \cdot \text{m}^{-2}$ , indicating the reduction of systematic error. Moreover, the discrete random error was also reduced to a certain extent according to the RMSE and  $R^2$  statistics; the RMSE of LE reduced from  $54.03$  to  $32.80 \text{ W} \cdot \text{m}^{-2}$  after data assimilation. The simulated results of surface heat flux in maize season over benchmark run were not as good as in wheat season due to the higher canopy height [46], but after the data assimilation procedure, the simulation results improved except for G, for which the *Eff* showed a negative value. The RMSE of LE reduced from  $119.78$  to  $67.44 \text{ W} \cdot \text{m}^{-2}$ , and the  $R^2$  increased from  $0.56$  to  $0.72$ .

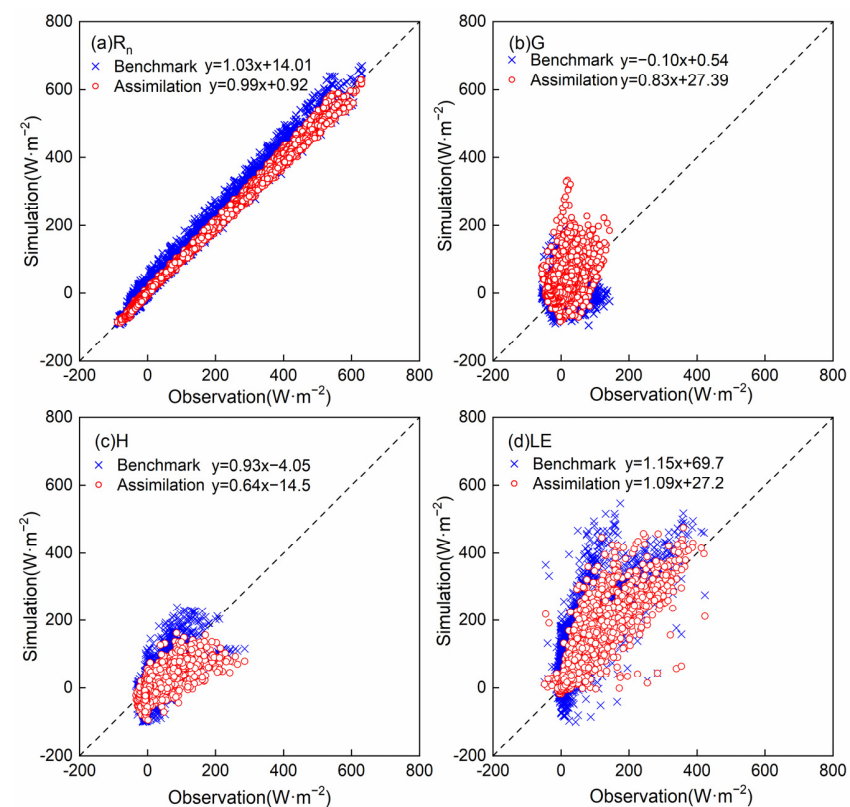
**Table 6.** Error statistics of surface flux simulation.

Evaluation Criteria		Wheat Season				Maize Season			
		Rn	G	H	LE	Rn	G	H	LE
<i>bias</i> ( $\text{W} \cdot \text{m}^{-2}$ )	benchmark	−19.40	−5.46	−29.73	19.99	17.72	−5.44	−6.79	80.96
	assimilation	−3.81	−4.96	6.63	−2.04	−0.21	26.46	−28.43	33.86
RMSE ( $\text{W} \cdot \text{m}^{-2}$ )	benchmark	30.05	57.81	71.07	54.03	26.03	40.53	51.62	119.78
	assimilation	17.76	51.63	47.72	32.80	9.81	48.44	45.50	67.44
$R^2$	benchmark	0.99	0.03	0.37	0.85	0.99	0.01	0.50	0.56
	assimilation	1.00	0.20	0.53	0.94	1.00	0.27	0.59	0.72
<i>Eff</i> (%)		65.08	20.21	54.91	63.15	85.80	−42.79	22.32	68.30



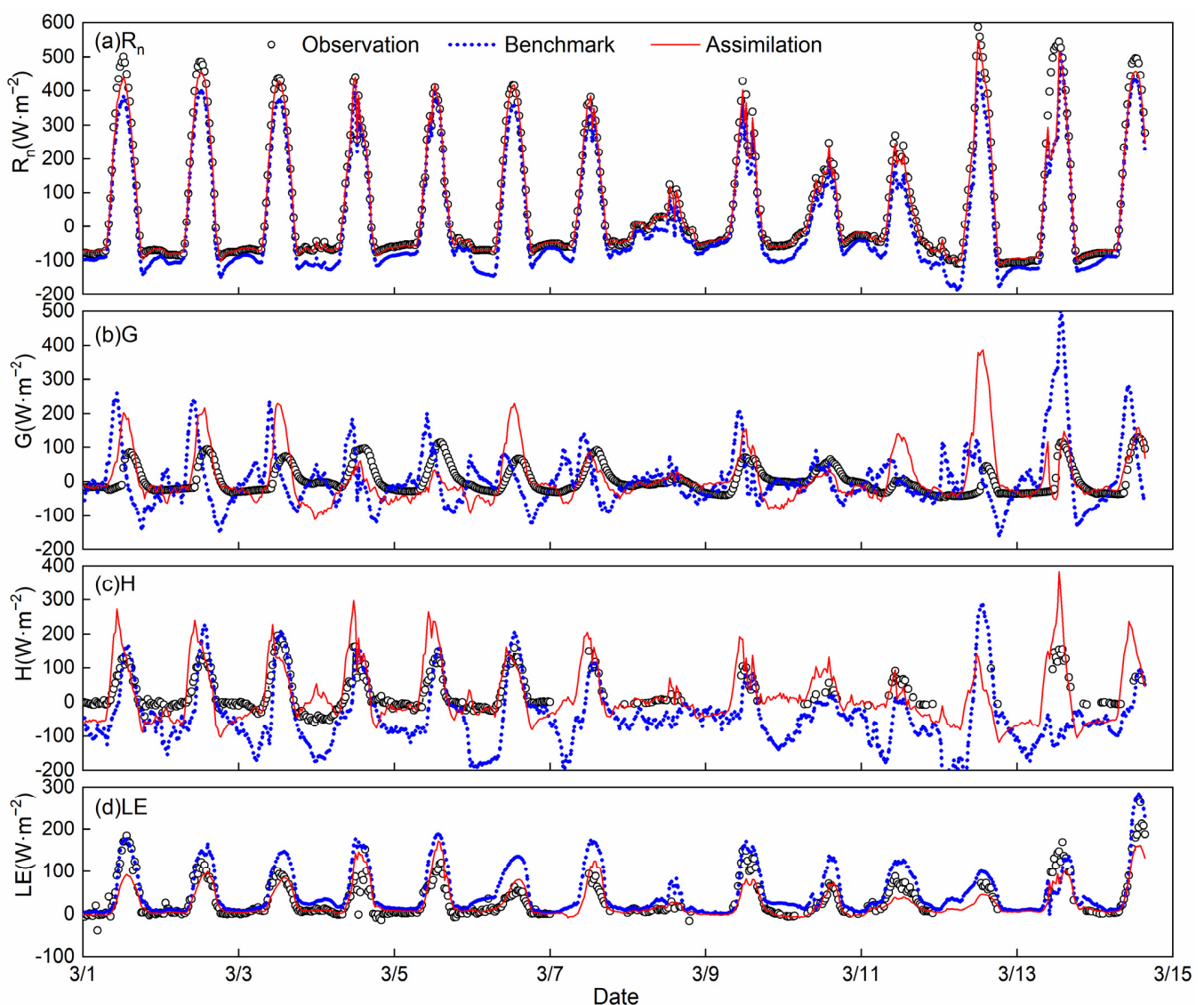


**Figure 7.** Scatter plots of surface flux simulation compared with observation by benchmark run and data assimilation run during wheat growing season.

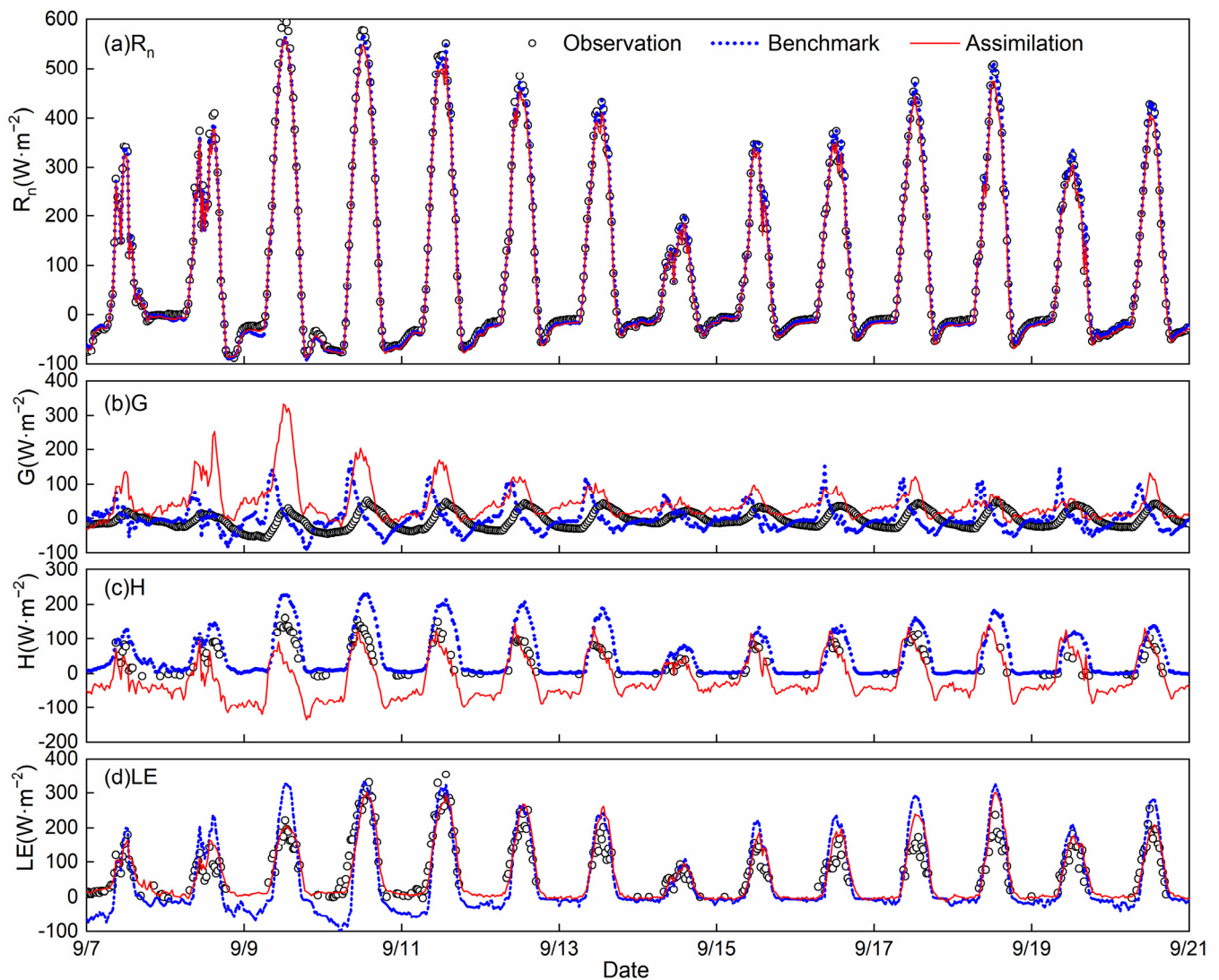


**Figure 8.** Scatter plots of surface flux simulation compared with observation by benchmark run and data assimilation run during maize growing season.

To further explore the performance of data assimilation results of surface heat flux, Figures 9 and 10 showed the two-week continuous simulation results for  $R_n$ ,  $G$ ,  $H$ , and  $LE$  from 1 March to 14 March in wheat season and 7 September to 20 September in maize season, respectively. The assimilation curves of the four variables had different correction effects compared with the benchmark run. For  $R_n$ ,  $H$ , and  $LE$  simulation, the assimilation curves were closer to the observed values than the simulation results, especially for the peaks of the curve, because in the HELP model, the error accumulation of these three variables is time-related, for example, the  $R_n$  at time step  $t$  is forced by state variables ( $T_a$ ,  $T_s$ , etc.) calculated at time  $t - 1$ , when the data assimilation occurred at  $t - 1$ , variables  $T_a$  and  $T_s$  was updated by introducing the new observation which may lead to improvement or deterioration of the results. On the contrary, the calculation of  $G$  is derived by an empirical equation related only to  $R_n$ , which means it has weaker relations with the 13 updated state variables. This may explain the bad performance of data assimilation of  $G$ .



**Figure 9.** Comparison of surface flux assimilation results and benchmark run with ground measurements during wheat growing season from 1 March 2006 to 14 March 2006.



**Figure 10.** Comparison of surface flux assimilation results and benchmark run with ground measurements during maize growing season from 7 September 2006 to 20 September 2006.

Overall, the verification results of data assimilation compared with observation values showed that updating the state variables of the HELP model could improve model output during the next time steps by making a prediction using the updated states. Noted that in our research, the update of state variables only happened when remote sensing data are available to provide observation value at certain times during clear days, which means by the time the model ran into the data assimilation procedure, the updated state variables were treated as the initial states of the model to support the model running in the following time steps until next available observation value. The results showed that after updating, the improvement of model simulation would last for a certain time, which proves the feasibility of this data assimilation scheme in improving the model performance in practical application.

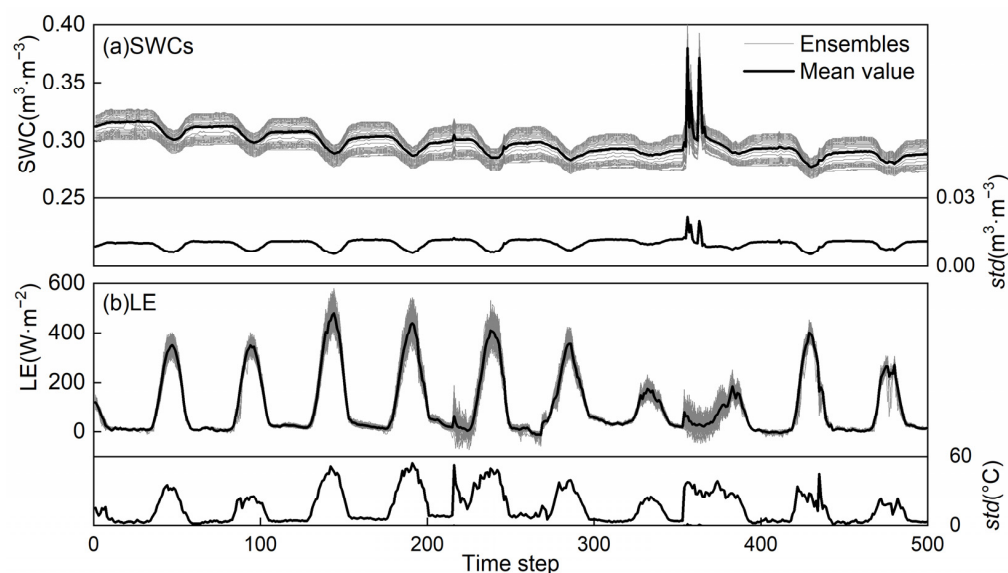
## 4. Discussion

### 4.1. Analysis of Ensemble Generation

The generation of the ensembles in the data assimilation scheme is a critical issue in ensuring the effective update of state variables. In this study, instead of setting the background covariance error matrix, we generated ensembles by adding random perturba-

tions to the forcing data as described in Section 2.3.4. We assumed that model simulation uncertainties will be embodied in the ensembles and will be correctly reflected in the prediction of state variables and outputs. In this case, numerical modeling was tested to make sure the random error setting to the forcing data was transmitted to the key variables by model simulation.

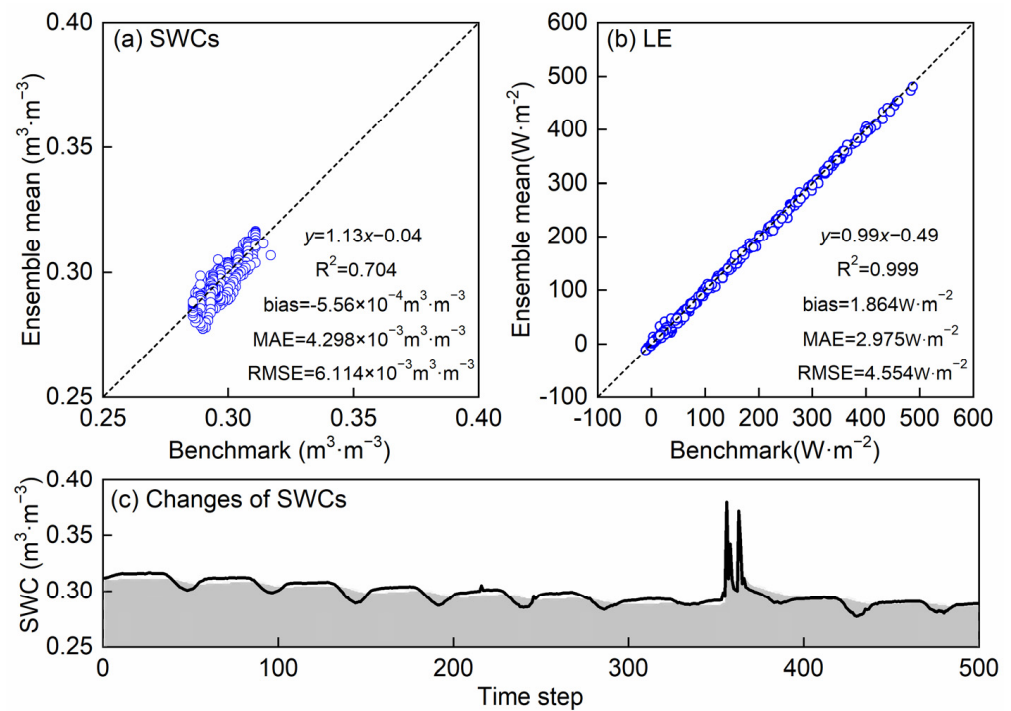
Figure 11 shows the ensemble perturbation and its standard deviation (*std*) of SWCs and LE from 100 ensembles (the time step set to 500). For each ensemble, the initial state variables originated from the ground measurements. The *std* of SWC showed a wide variety, ranging from 0.0053 to 0.0216  $\text{m}^3 \cdot \text{m}^{-3}$ . The ensemble members were perturbed severely when the precipitation event happened because extra random error from perturbation of precipitation was introduced to the model. After precipitation, the ensembles of SWC quickly converged to the previous level before precipitation. The ensemble of LE showed a similar pattern as SWC, and the *std* ranged from 1.07 to 54.09  $\text{W} \cdot \text{m}^{-2}$ . Different from the SWC ensemble, the influence of precipitation on the LE ensemble lasted a longer time, which reflected that SWC and LE had different response speeds to precipitation. Results showed random noise added to the forcing data could be corrected and conducted to the state variable and thus output value. The dispersion of the ensemble also showed different sensitivity degrees to forcing data.



**Figure 11.** Ensemble perturbation and its *std*.

Figure 12a,b are the scatter plots of benchmark run and ensemble mean for SWCs and LE, respectively. By the benchmark run, the HELP model ran freely without putting perturbation to the forcing data. In this case, the ensemble mean values of SWC and LE are supposed to approach the benchmark run, indicating the ensemble generation does not bring systematic error to the model simulation. The *bias* of SWCs of the ensemble mean compared with the benchmark run is 0.0019  $\text{m}^3 \cdot \text{m}^{-3}$ , which can be treated as unbiased as a whole, but the time process curve (Figure 12c) showed that the ensemble mean value changed more intense than the benchmark run, especially when precipitation occurred (highlighted with a red square). This indicated that the model error of SWC might be magnified by adding random noise to the forcing data. However, for the LE variable, the ensemble mean value was well fitted with the benchmark run with an  $R^2$  of 0.999. However, a slight systematic error is reflected in the LE ensemble, with a *bias* of  $-0.0178 \text{ W} \cdot \text{m}^{-2}$ , which is acceptable, though. Basically, the ensemble generation in this study is reasonable and correctly reflects the model error into the state variables and output.





**Figure 12.** The comparison of benchmark run and ensemble mean.

#### 4.2. Dependency Analysis of State Variables

In the EnKF data assimilation scheme, forcing data, state variables, and output are related by the Kalman gain matrix and observation operator. Theoretically, if the three components have a relatively low correlation, the effect of data assimilation may not be significant. Tables 7 and 8 showed the correlation coefficient of HELP model state variables and surface flux in wheat and maize season, respectively. Benchmark run results were used in this analysis to make sure the correlation of state variables was correctly reflected through the model. The SWC of three layers had a high correlation coefficient, and LE had a relatively high correlation coefficient with SWC, which approved the positive effect of data assimilation of SWC using LE as an observation operator. Compared with the wheat season, the correlation coefficient of LE and SWC in the maize season was higher, which indicated the results in Section 3.2 that for SWC estimation, the maize season had better performance than the wheat season.

For both wheat season and maize season, LE had the highest correlation with all the state variables and surface flux, which proved that using remote sensing LE estimation as an observation operator was reasonable and feasible. The results also showed that other than LE,  $T_s$  also had a relatively higher correlation coefficient with surface flux. The correlation coefficient of  $T_s$  and LE in wheat and maize season are 0.63 and 0.66, respectively. Figure 13 is the scatter plot of  $T_s$  simulation compared with an observation by benchmark run, and data assimilation ran during wheat and maize season, respectively. Results showed data assimilation procedure had a larger effect on  $T_s$  simulation compared with SWC. In the wheat season, the benchmark run had obvious systematic error compared with observation; after data assimilation, the *bias* reduced to  $-1.074\text{ }^\circ\text{C}$ , RMSE reduced to  $1.786\text{ }^\circ\text{C}$ , and the *Eff* of data assimilation was 94.8%, and  $R^2$  increased from 0.661 to 0.955. In maize season, the benchmark run had better results, but after data assimilation, the random error also reduced to a large extent, with an *Eff* of 73.0%,  $R^2$  increased from 0.943 to 0.968, and RMSE reduced from 2.473 to  $1.286\text{ }^\circ\text{C}$ . Results showed the assimilation using LE as an observation operator had an obvious effect in improving the simulation  $T_s$  as expected.

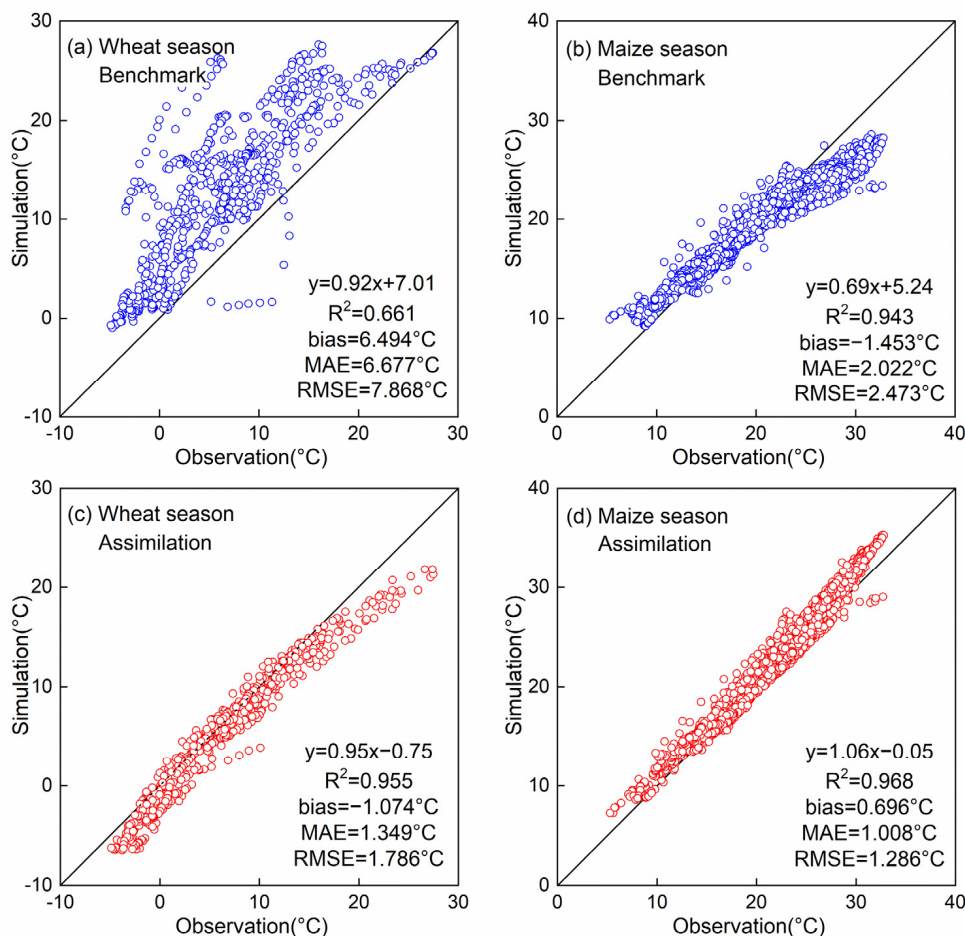


**Table 7.** Correlation coefficient of HELP model state variables and surface flux in wheat season.

	SWCs	SWCr	SWCd	Ts	Rn	G	H	LE
SWCs	1	0.94	0.81	0.03	−0.10	−0.05	0.16	0.23
SWCr	0.94	1	0.95	0.08	−0.08	−0.03	0.14	0.19
SWCd	0.81	0.95	1	0.10	−0.09	−0.02	0.12	0.19
Ts	0.03	0.08	0.10	1	0.66	0.88	0.69	0.63
Rn	−0.10	−0.08	−0.09	0.66	1	0.71	0.73	0.89
G	−0.05	−0.03	−0.02	0.88	0.71	1	0.59	0.64
H	0.16	0.14	0.12	0.69	0.73	0.59	1	0.41
LE	0.23	0.19	0.19	0.63	0.89	0.64	0.41	1

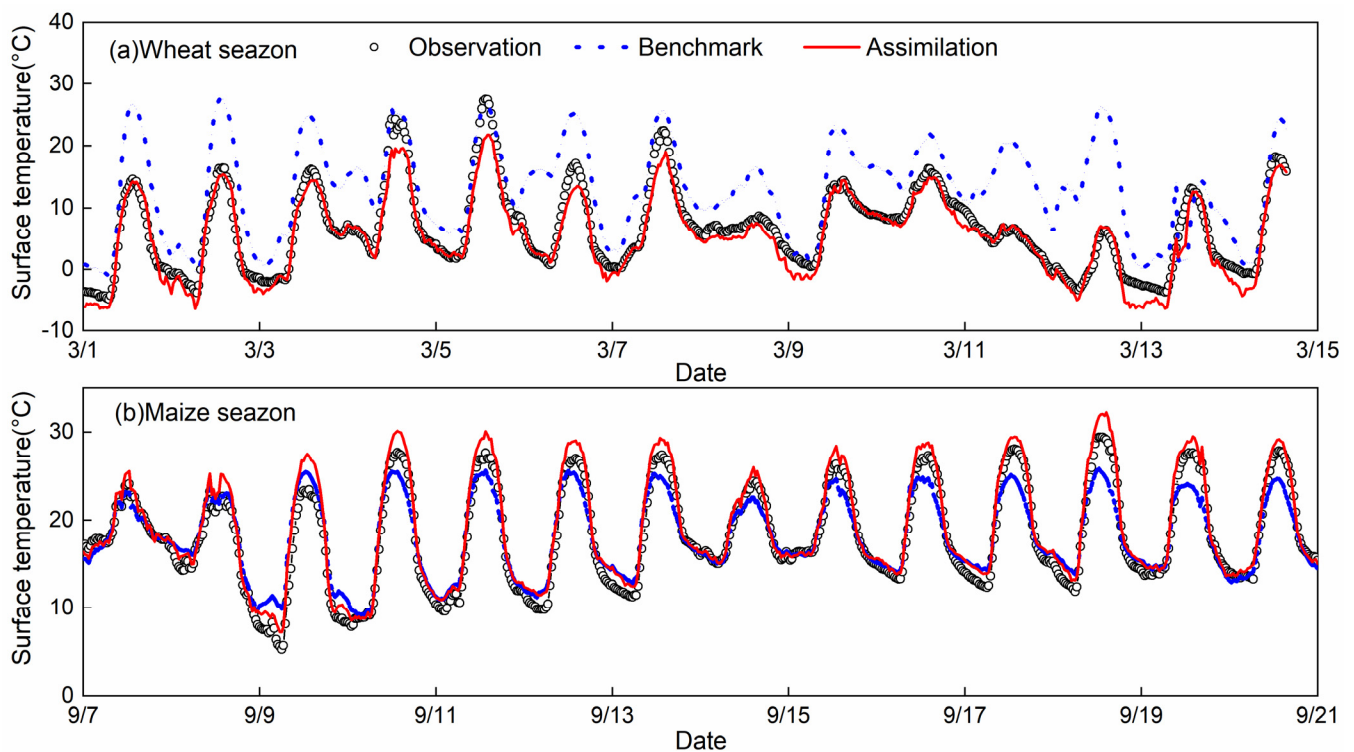
**Table 8.** Correlation coefficient of HELP model state variables and surface flux in maize season.

	SWCs	SWCr	SWCd	Ts	Rn	G	H	LE
SWCs	1	0.99	0.83	0.19	0.02	−0.14	0.02	0.41
SWCr	0.99	1	0.86	0.21	0.02	−0.14	0.01	0.41
SWCd	0.83	0.86	1	0.37	0.03	−0.09	−0.07	0.37
Ts	0.19	0.21	0.37	1	0.66	0.85	0.39	0.66
Rn	0.02	0.02	0.03	0.66	1	0.77	0.87	0.84
G	−0.14	−0.14	−0.09	0.85	0.77	1	0.81	0.60
H	0.02	0.01	−0.07	0.39	0.87	0.81	1	0.54
LE	0.41	0.41	0.37	0.66	0.84	0.60	0.54	1



**Figure 13.** The scatter plot of Ts simulation compared with observation by benchmark run and data assimilation run.

Figure 14a,b showed the two-week continuous simulation results for Ts from 1 March to 14 March in wheat season and 7 September to 20 September in maize season, respectively. Results showed the data assimilation improved the Ts simulation, especially for errors in the peaks of the curve. However, despite the data assimilation having overall great performance, it did not improve the simulation for every single day. For example, on 4 March and 10 September, after data assimilation, the error of simulation increased compared with the benchmark run, which is due to the complexity of the mechanism in the energy transfer process and simplification in the model structure.



**Figure 14.** The two-week continuous simulation results for Ts.

#### 4.3. Limitations and Future Works

In this study, the data assimilation method was used to simulate SWC and surface fluxes continuously, which was verified in the Weishan Irrigation District. However, there is some experience in this method, such as the occurrence time of the maximum  $T_a$  was determined by the method of fitting ground observations. In addition, this method was only validated under wheat and maize crop types in the Weishan Irrigation District. The extension of the model also requires ground verification in different climate zones and underlying surface types. At the same time, the applicable conditions and limitations of the model should be analyzed, and the parameter values of the model under different underlying surfaces should be modified and improved.

In addition, this study only conducted a simulation for one year. As the underlying surface is relatively uniform, parameterization schemes for different underlying surfaces are also relatively simple. In the process of SWC assimilation simulation, integration application with existing SWC products such as SMAP, CLDAS, and SMOS was not considered. At present, SWC fusion is a hot topic, and the existing SWC products have the advantage of high temporal resolution. In the future, it is necessary to further refine the parameterization scheme of the model for different underlying surface types. In order to provide a reference for regional water resources management, a long-term series of SWC and surface fluxes simulations was carried out by integrating existing SWC products. This study proved that using remote sensing observations as data assimilation can effectively improve the continuous simulation of SWC and surface flux in land surface process models. In this

study, only one remote sensing data source was used to analyze data assimilation results. In the following research, multi-objective data assimilation of multi-source remote sensing data should also be considered. Remote sensing rainfall and SWC should be introduced into the data assimilation system for multi-objective data assimilation.

## 5. Conclusions

Land surface flux and SWC are the important parameters of the hydrological cycle and play an important role in the exchange of matter and energy between different components in the earth system. At present, there are many methods or studies on theoretical calculation, remote sensing estimation, and model simulation of the above parameters, but the observation error and model error exist inevitably, and sometimes the error is even larger. How to reduce these errors and improve the accuracy of the model or calculation results is the main issue to be answered in this research. It is of great significance to promote hydrology research and development. In this study, an EnKF technique was coupled to a HELP model to update model states, including SWC and Ts. The remotely sensed LE estimated by SEBS was used as the observation value in the data assimilation system to update the model states such as SWC and Ts, etc. The model was validated by the observation data in 2006 at the Weishan flux station, where the open-loop estimation without state updating was treated as the benchmark run. The main conclusions are given as follows:

1. A land surface data assimilation system was built to the HELP model by the EnKF technique. The remotely sensed LE estimated by the remotely sensed SEBS model was used as the observation value in the data assimilation system.
2. The model was validated by the observation data in 2006 at the Weishan flux station, where the open-loop estimation without state updating was treated as the benchmark run.
3. The RMSE of SWC reduced by 30~50% compared to the benchmark run, while the surface fluxes also had significant improvement to different extents, among which the RMSE of LE estimation from the wheat season and maize season reduced by 33% and 44%, respectively.
4. The data assimilation system has great potential to be used in land surface models in agriculture and water management.

**Author Contributions:** Methodology, H.C.; Data processing, R.L.; Data analysis and validation, H.C., R.L. and Z.W.; Funding acquisition, B.Z.; Project administration, B.Z. and Z.W.; Proofread, H.C. and B.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Key Research and Development Program of China (No.2021YFB3900602), the Chinese National Science Fund (No. 52130906), the Independent Research Project of State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin (No. SKL2022TS13), and the Fund of China Institute of Water Resources and Hydropower Research (No. ID0145B022021, and ID0145B052021).

**Data Availability Statement:** The data presented in this study are available on request from the first author (H.C.).

**Acknowledgments:** We would like to thank the anonymous reviewers for their long-term guidance and constructive comments. The author would like to thank the support of International Talented Program by IWHR.

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

## References

1. McVicar, T.R.; Roderick, M.L.; Donohue, R.J.; Li, L.T.; Van Niel, T.G.; Thomas, A.; Grieser, J.; Jhajharia, D.; Himri, Y.; Mahowald, N.M.; et al. Global review and synthesis of trends in observed terrestrial near-surface wind speeds: Implications for evaporation. *J. Hydrol.* **2012**, *416–417*, 182–205. [[CrossRef](#)]
2. Bousbih, S.; Zribi, M.; El Hajj, M.; Baghdadi, N.; Lili-Chabaane, Z.; Gao, Q.; Fanise, P. Soil Moisture and Irrigation Mapping in A Semi-Arid Region, Based on the Synergetic Use of Sentinel-1 and Sentinel-2 Data. *Remote Sens.* **2018**, *10*, 1953. [[CrossRef](#)]

3. El Hajj, M.; Baghdadi, N.; Zribi, M.; Bazzi, H. Synergic Use of Sentinel-1 and Sentinel-2 Images for Operational Soil Moisture Mapping at High Spatial Resolution over Agricultural Areas. *Remote Sens.* **2017**, *9*, 1292. [[CrossRef](#)]
4. Petropoulos, G.; Srivastava, P.; Piles, M.; Pearson, S. Earth Observation-Based Operational Estimation of Soil Moisture and Evapotranspiration for Agricultural Crops in Support of Sustainable Water Management. *Sustainability* **2018**, *10*, 181. [[CrossRef](#)]
5. Long, D.; Bai, L.; Yan, L.; Zhang, C.; Yang, W.; Lei, H.; Quan, J.; Meng, X.; Shi, C. Generation of spatially complete and daily continuous surface soil moisture of high spatial resolution. *Remote Sens. Environ.* **2019**, *233*, 111364. [[CrossRef](#)]
6. Lin, R.; Chen, H.; Wei, Z.; Li, Y.; Zhang, B.; Sun, H.; Cheng, M. Improved Surface Soil Moisture Estimation Model in Semi-Arid Regions Using the Vegetation Red-Edge Band Sensitive to Plant Growth. *Atmosphere* **2022**, *13*, 930. [[CrossRef](#)]
7. Huang, Q.; Wang, C.; Meng, Y.; Chen, J.; Yue, A. Landslide Monitoring Using Change Detection in Multitemporal Optical Imagery. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 312–316. [[CrossRef](#)]
8. Ranney, K.J.; Niemann, J.D.; Lehman, B.M.; Green, T.R.; Jones, A.S. A method to downscale soil moisture to fine resolutions using topographic, vegetation, and soil data. *Adv. Water Resour.* **2015**, *76*, 81–96. [[CrossRef](#)]
9. Turk, F.; Mostovoy, G.; Anantharaj, V. Soil Moisture Sensitivity to NRL-Blend High-Resolution Precipitation Products: Analysis of Simulations with Two Land Surface Models. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2010**, *3*, 32–48. [[CrossRef](#)]
10. Srivastava, A.; Saco, P.M.; Rodriguez, J.F.; Kumari, N.; Chun, K.P.; Yetemen, O. The role of landscape morphology on soil moisture variability in semi-arid ecosystems. *Hydrol. Process.* **2020**, *35*, e13990. [[CrossRef](#)]
11. Wang, Y.; Cao, J.; Liu, Y.; Zhu, Y.; Fang, X.; Huang, Q.; Chen, J. Spatiotemporal Analysis of Soil Moisture Variation in the Jiangsu Water Supply Area of the South-to-North Water Diversion Using ESA CCI Data. *Remote Sens.* **2022**, *14*, 256. [[CrossRef](#)]
12. Famiglietti, J.S.; Ryu, D.; Berg, A.A.; Rodell, M.; Jackson, T.J. Field observations of soil moisture variability across scales. *Water Resour. Res.* **2008**, *44*, W01423. [[CrossRef](#)]
13. Baugh, C.; de Rosnay, P.; Lawrence, H.; Jurlina, T.; Drusch, M.; Zsoter, E.; Prudhomme, C. The Impact of SMOS Soil Moisture Data Assimilation within the Operational Global Flood Awareness System (GloFAS). *Remote Sens.* **2020**, *12*, 1490. [[CrossRef](#)]
14. Zhao, W.; Sánchez, N.; Lu, H.; Li, A. A spatial downscaling approach for the SMAP passive surface soil moisture product using random forest regression. *J. Hydrol.* **2018**, *563*, 1009–1024. [[CrossRef](#)]
15. Abowarda, A.S.; Bai, L.; Zhang, C.; Long, D.; Li, X.; Huang, Q.; Sun, Z. Generating surface soil moisture at 30 m spatial resolution using both data fusion and machine learning toward better water resources management at the field scale. *Remote Sens. Environ.* **2021**, *255*, 112301. [[CrossRef](#)]
16. Zribi, M.; Chahbi, A.; Shabou, M.; Lili-Chabaane, Z.; Duchemin, B.; Baghdadi, N.; Amri, R.; Chehbouni, A. Soil surface moisture estimation over a semi-arid region using ENVISAT ASAR radar data for soil evaporation evaluation. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 345–358. [[CrossRef](#)]
17. Gao, Q.; Zribi, M.; Escorihuela, M.J.; Baghdadi, N. Synergetic Use of Sentinel-1 and Sentinel-2 Data for Soil Moisture Mapping at 100 m Resolution. *Sensors* **2017**, *17*, 1966. [[CrossRef](#)]
18. Robinson, D.A.; Campbell, C.S.; Hopmans, J.W.; Hornbuckle, B.K.; Jones, S.B.; Knight, R.; Ogden, F.; Selker, J.; Wendroth, O. Soil Moisture Measurement for Ecological and Hydrological Watershed-Scale Observatories: A Review. *Vadose Zone J.* **2008**, *7*, 358–389. [[CrossRef](#)]
19. Agyeman, B.T.; Bo, S.; Sahoo, S.R.; Yin, X.; Liu, J.; Shah, S.L. Soil moisture map construction by sequential data assimilation using an extended Kalman filter. *J. Hydrol.* **2021**, *598*, 126425. [[CrossRef](#)]
20. Sabater, J.M.; Jarlan, L.; Calvet, J.-C.; Bouyssel, F.; Rosnay, P.D. From Near-Surface to Root-Zone Soil Moisture Using Different Assimilation Techniques. *J. Hydrometeorol.* **2007**, *8*, 194–206. [[CrossRef](#)]
21. Baldwin, D.; Manfreda, S.; Keller, K.; Smithwick, E.A.H. Predicting root zone soil moisture with soil properties and satellite near-surface moisture data across the conterminous United States. *J. Hydrol.* **2017**, *546*, 393–404. [[CrossRef](#)]
22. Khaki, M.; Awange, J. Altimetry-derived surface water data assimilation over the Nile Basin. *Sci. Total Environ.* **2020**, *735*, 139008. [[CrossRef](#)] [[PubMed](#)]
23. Margulis, S.A.; McLaughlin, D.; Entekhabi, D.; Dunne, S. Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 Field Experiment. *Water Resour. Res.* **2002**, *38*, 35-1–35-18. [[CrossRef](#)]
24. Bosilovich, M.G.; Radakovich, J.D.; da Silva, A.; Todling, R.; Verter, F. Skin Temperature Analysis and Bias Correction in a Coupled Land-Atmosphere Data Assimilation System. *J. Meteorol. Soc. Jpn. Ser. II* **2007**, *85A*, 205–228. [[CrossRef](#)]
25. Kalman, R.E. A New Approach to Linear Filtering and Prediction Problems. *J. Basic Eng.* **1960**, *82*, 35–45. [[CrossRef](#)]
26. Komma, J.; Blöschl, G.; Reszler, C. Soil moisture updating by Ensemble Kalman Filtering in real-time flood forecasting. *J. Hydrol.* **2008**, *357*, 228–242. [[CrossRef](#)]
27. Thirel, G.; Martin, E.; Mahfouf, J.F.; Massart, S.; Ricci, S.; Habets, F. A past discharges assimilation system for ensemble streamflow forecasts over France—Part 1: Description and validation of the assimilation system. *Hydrol. Earth Syst. Sci.* **2010**, *14*, 1623–1637. [[CrossRef](#)]
28. Walker, J.P.; Houser, P.R. A methodology for initializing soil moisture in a global climate model: Assimilation of near-surface soil moisture observations. *J. Geophys. Res. Atmos.* **2001**, *106*, 11761–11774. [[CrossRef](#)]
29. Crow, W.T.; Koster, R.D.; Reichle, R.H.; Sharif, H.O. Relevance of time-varying and time-invariant retrieval error sources on the utility of spaceborne soil moisture products. *Geophys. Res. Lett.* **2005**, *32*, 1–5. [[CrossRef](#)]
30. Chen, F.; Crow, W.T.; Starks, P.J.; Moriasi, D.N. Improving hydrologic predictions of a catchment model via assimilation of surface soil moisture. *Adv. Water Resour.* **2011**, *34*, 526–536. [[CrossRef](#)]



31. Qin, C.; Jia, Y.; Su, Z.; Zhou, Z.; Qiu, Y.; Suhui, S. Integrating Remote Sensing Information into A Distributed Hydrological Model for Improving Water Budget Predictions in Large-scale Basins through Data Assimilation. *Sensors* **2008**, *8*, 4441–4465. [[CrossRef](#)] [[PubMed](#)]
32. Pan, M.; Wood, E.; Wojcik, R.; McCabe, M. Estimation of regional terrestrial water cycle using multi-sensor remote sensing observations and data assimilation. *Remote Sens. Environ.* **2008**, *112*, 1282–1294. [[CrossRef](#)]
33. Sun, H.; Wang, L.; Lin, R.; Zhang, Z.; Zhang, B. Mapping Plastic Greenhouses with Two-Temporal Sentinel-2 Images and 1D-CNN Deep Learning. *Remote Sens.* **2021**, *13*, 2820. [[CrossRef](#)]
34. Yi, Y.; Yang, D.; Huang, J.; Chen, D. Evaluation of MODIS surface reflectance products for wheat leaf area index (LAI) retrieval. *ISPRS J. Photogramm. Remote Sens.* **2008**, *63*, 661–677. [[CrossRef](#)]
35. Liang, S. Narrowband to broadband conversions of land surface albedo I: Algorithms. *Remote Sens. Environ.* **2001**, *76*, 213–238. [[CrossRef](#)]
36. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.; Gao, X.; Ferreira, L. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [[CrossRef](#)]
37. Sellers, P.J.; Randall, D.A.; Collatz, G.J.; Berry, J.A.; Field, C.B.; Dazlich, D.A.; Zhang, C.; Collelo, G.D.; Bounoua, L. A Revised Land Surface Parameterization (SiB2) for Atmospheric GCMs. Part I: Model Formulation. *J. Clim.* **1996**, *9*, 676–705. [[CrossRef](#)]
38. Evensen, G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res. Oceans* **1994**, *99*, 10143–10162. [[CrossRef](#)]
39. Kalman, R.E.; Bucy, R.S. New Results in Linear Filtering and Prediction Theory. *J. Basic Eng.* **1961**, *83*, 95–108. [[CrossRef](#)]
40. van der Kwast, J.; Timmermans, W.; Gieske, A.; Su, Z.; Olioso, A.; Jia, L.; Elbers, J.; Karssenber, D.; de Jong, S. Evaluation of the Surface Energy Balance System (SEBS) applied to ASTER imagery with flux-measurements at the SPARC 2004 site (Barrax, Spain). *Hydrol. Earth Syst. Sci.* **2009**, *13*, 1337–1347. [[CrossRef](#)]
41. Su, Z.; Schmugge, T.; Kustas, W.P.; Massman, W.J. An evaluation of two models for estimation of the roughness height for heat transfer between the land surface and the atmosphere. *J. Appl. Meteorol. Climatol.* **2001**, *40*, 1933–1951. [[CrossRef](#)]
42. Su, Z. The surface energy balance system (SEBS) for estimation of turbulent heat fluxes. *Hydrol. Earth Syst. Sci.* **2002**, *6*, 85–100. [[CrossRef](#)]
43. Shaw, R.; Pereira, A. Aerodynamic roughness of a plant canopy: A numerical experiment. *Agric. Meteorol.* **1982**, *26*, 51–65. [[CrossRef](#)]
44. Choudhury, B.J.; Monteith, J.L. A four-layer model for the heat budget of homogeneous land surfaces. *Q. J. R. Meteorol. Soc.* **1988**, *114*, 373–398. [[CrossRef](#)]
45. Shuttleworth, W.J.; Wallace, J.S. Evaporation from sparse crops—an energy combination theory. *Q. J. R. Meteorol. Soc.* **1985**, *111*, 839–855. [[CrossRef](#)]
46. Yang, D.; Chen, H.; Lei, H. Estimation of evapotranspiration using a remote sensing model over agricultural land in the North China Plain. *Int. J. Remote Sens.* **2010**, *31*, 3783–3798. [[CrossRef](#)]
47. Clark, M.P.; Rupp, D.E.; Woods, R.A.; Zheng, X.; Ibbitt, R.P.; Slater, A.G.; Schmidt, J.; Uddstrom, M.J. Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model. *Adv. Water Resour.* **2008**, *31*, 1309–1324. [[CrossRef](#)]
48. Chen, H.; Yang, D.; Hong, Y.; Gourley, J.J.; Zhang, Y. Hydrological data assimilation with the Ensemble Square-Root-Filter: Use of streamflow observations to update model states for real-time flash flood forecasting. *Adv. Water Resour.* **2013**, *59*, 209–220. [[CrossRef](#)]
49. Lei, H.; Yang, D. Interannual and seasonal variability in evapotranspiration and energy partitioning over an irrigated cropland in the North China Plain. *Agric. For. Meteorol.* **2010**, *150*, 581–589. [[CrossRef](#)]
50. Lei, H.; Yang, D.; Shen, Y.; Liu, Y.; Zhang, Y. Simulation of evapotranspiration and carbon dioxide flux in the wheat-maize rotation croplands of the North China Plain using the Simple Biosphere Model. *Hydrol. Process.* **2011**, *25*, 3107–3120. [[CrossRef](#)]
51. Govaerts, B.; Fuentes, M.; Mezzalama, M.; Nicol, J.M.; Deckers, J.; Etchevers, J.D.; Figueroa-Sandoval, B.; Sayre, K.D. Infiltration, soil moisture, root rot and nematode populations after 12 years of different tillage, residue and crop rotation managements. *Soil Tillage Res.* **2007**, *94*, 209–219. [[CrossRef](#)]