An Exponential Filter Model-Based Root-Zone Soil Moisture Estimation Methodology from Multiple Datasets

Yanqing Yang 1,2,3, Zhenxin Bao 1,3,*, Houfa Wu 1,2,3, Guoqing Wang 1,3, Cuishan Liu 1,3, Jie Wang 1,3,4 and Jianyun Zhang 1,3

1 State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute, Nanjing 210029, China; 2017323060018@stu.scu.edu.cn (Y.Y.); 2019323060018@stu.scu.edu.cn (H.W.);
gqwang@nhri.cn (G.W.); csliu@nhri.cn (C.L.); 05141805@cumt.edu.cn (J.W.); jyzhang@nhri.cn (J.Z.)
2 State Key Laboratory of Hydraulics and Mountain River Engineering, College of Water Resource and Hydropower, Sichuan University, Chengdu 610065, China
3 Research Center for Climate Change, Ministry of Water Resources, Nanjing 210029, China
4 School of Civil Engineering, Tianjin University, Tianjin 300350, China
* Correspondence: zxbaonhri.cn; Tel.: +86-25-85828830

Abstract: Modern smart agriculture initiative presents more requests for soil moisture (SM) monitoring over large agricultural areas. Remote sensing techniques facilitate high-resolution surface SM (SSM) estimation at a large scale but lack root zone SM (RZSM) information. Establishing the deduction method of RZSM from the SSM has long been the focus of most attention. Data assimilation methods are promising techniques for RZSM estimation, developing numerous assimilated reanalysis datasets, e.g., ERA5 and the latest Soil Moisture Active and Passive (SMAP) L4 SM product. However, data latency and large computation during data collecting and processing often inhibits further applications. This work proposes a rapid estimation scheme for estimating RZSM with short latency and small computations, based on the Exponential Filter (EF) method. The EF model with single parameter T was firstly calibrated and validated using the SSM and RZSM of ERA5 reanalysis dataset, obtaining the optimum parameter T map for each grid. Then, the fast-updating SMAP L3 SSM product together with the scale-matched optimum T were adopted as inputs into the EF model to retrieve RZSM estimation of each grid. Specifically, such estimation scheme was tested over the central and eastern agricultural areas of China, using a dense monitoring network of 796 SM observation sites, which contains various land uses, as well as meteorological and hydrological conditions. The calibrated optimum parameter T presented an increasing trend with good physical explanations. Furthermore, all the estimated RZSMs were found to have good performances on capturing the temporal-spatial variations of RZSM and well reflecting seasonal RZSM changes. Overall, such an estimation scheme was proven to be a desirable alternative for estimating RZSM over large agricultural areas.

Keywords: SMAP; root zone; soil moisture; Exponential Filter; agricultural water management

1. Introduction

Soil moisture (SM) plays a key role in the hydrologic cycle processes by controlling the rainfall distributions between surface runoff, infiltration, and ground water recharge, and the evapotranspiration rate from bare and vegetated areas [1,2]. Hence, a good understanding of SM information is essential to improve weather forecasting [3], drought and flood predictions [4–7], agricultural water resources management [8], and climate change investigations [9]. Traditional methods to obtain SM information are based on those station-observed instruments, including the gravimetric method [10], time domain reflectometry (TDR), gamma ray scanners, and neutron probes [11], data quality of which depend on the sampling frequency and distribution density of measuring stations at regional to global scales. Newly arising remote sensing (RS) techniques can continually estimate SM at a large
scale with increasing spatial-temporal resolution. However, remote sensing methods only provide surface SM at 0–10 cm depth, lacking SM information at deeper depths (especially rootzone areas), which has been proven much more important for crop growth [12,13]. Such limitations of RS techniques inhibited their wide applications and propagations in modern agriculture [14]; thus, investigating the relationships between rootzone and surface SM, and constructing a reliable deduction method for SM at rootzone depths using the surface satellite SM information, has been the focus in current studies [15–18].

Many scholars have attempted to estimate root zone SM (RZSM) from the surface SM (SSM) through numerous methods, including physical models [19], data assimilation methods [15,20–22], neural networks [23,24], deep learning algorithms [25], etc. Thereinto, data assimilation techniques are physically based and widely used method for estimating rootzone SM at larger scale [26]. The primary principle of data assimilation method is to integrate the surface SM (or relating to) observations into the land surface model [27], to update the simulation results of deep-layer soil moisture of the model. The Ensemble Kalman Filter (EnKF) is generally regarded as the most promising assimilation technique to estimate profile SM [28]. This approach has been strengthened due to the increasingly mature EnKF methods and the enhanced availability of multi-source SM datasets [19]; nonetheless, several drawbacks of EnKF are highlighted in the nonlinear systems and high computational demands [29]. The most well-known and representative assimilated RZSM dataset is provided by the Soil Moisture Active and Passive (SMAP) L4 product [30–32]. Such product assimilated the surface SMAP-L1 brightness temperature observations using EnKF into the watershed land surface model developed by NASA and provided a global 9 km × 9 km grid, 3-h resolution of 0–100 cm SM information [22]. However, SMAP L4 product defines soil moisture in the top 1 m of the soil column as root zone soil moisture, such depths cannot directly provide the RZSM dynamic information at the crop root layer (usually at 0–30 cm depth) for those agricultural areas.

There are also some other reanalysis datasets developed based on global land surface models and data assimilation systems, providing multi-layer RZSM series with long covering periods [33,34]. For example, ERA5 is the most recent reanalysis dataset developed by the European Centre for Medium-Range Weather Forecasts (ECMWF), with four layers of soil moisture simulated in the ECMWF surface model: Layer 1: 0–7 cm, Layer 2: 7–28 cm, Layer 3: 28–100 cm, and Layer 4: 100–289 cm [33]. Despite a high spatial-temporal resolution (i.e., 1 h and 9 km) and four-layer SM of ERA5, the latest quality-assured ERA5 data cannot be released until 2–3 months before the present. Such data latency of ERA5 inhibits its further application to crop guidance (e.g., drought forecasting) in modern agriculture.

An alternative and increasingly more useful approach is to deduce RZSM from the near real-time satellite SSM using the Exponential Filter (EF) method [35]. This method was firstly proposed by Wagner et al. [36], which transformed the observed SSM series with time into dynamic signals of soil moisture at deeper depths, and on this basis, derived the concept of deducing RZSM using the Soil Wetness Index (SWI). EF method is simple and practical, with only one parameter, i.e., characteristic time length (T), and has been widely used in rainfall runoff simulations to improve the prediction accuracy of soil moisture in the root-zone layer [16,37,38]. Increasingly, numerous satellite sensors (including optical, and active and passive microwave sensors) can provide multi-source SSM datasets, facilitating RZSM simulations for EF method at large scales, especially for those with almost real time latency (e.g., SMAP L3). Stefan et al. [39] adopted the EF model to retrieve RZSM estimates from SMAP SSM, with parameter T calibrated per land cover type using a long-term simulated SM dataset by land surface model. Such a calibration scheme was proved an adequate approach in applying EF to derive RZSM at large scale. However, parameter T is prone to be affected by all physical processes affecting the soil moisture, such as evapotranspiration, the hydraulic characteristics of the soil, the soil depth, and the number of soil layers [40]; thus, merely considering the land cover type is still defective, and a wide range of calibration sites containing various meteorologic and hydrological conditions should be involved in and further verified.
This work aimed to test the applicability of EF methods in estimating RZSM from SMAP L3 SSM dataset, and to establish a scheme suited for RZSM estimation with short latency and high computational efficiency over large agricultural areas. To this purpose, this work calibrated and validated the EF model using ERA5 reanalysis SM dataset over central and eastern agricultural areas of China, obtaining the optimum parameter $T$ of each grid according to the largest Nash–Sutcliffe Efficient (NSE) between the ERA5-derived and ERA5-provided RZSM. The ERA5-derived RZSM was then evaluated against the observed RZSM series in different agricultural zonings, including the spatial-temporal and seasonal performance from site to regional scale. The calibrated optimum $T$ values were interpolated to the grid matched with the spatial resolution of SMAP L3, and on this basis, optimum $T$ value together with SMAP L3 SSM was used as inputs of EF model to retrieve RZSM estimation for each grid. The performances of SMAP L3-derived RZSM series were also evaluated compared to the observations. Further, the applicability of the proposed calculating scheme was discussed, as well as the implications for modern agricultural water management.

2. Data and Methodology

2.1. Study Area

The central and eastern agricultural areas of China ($105.14^\circ$E–$131.31^\circ$E, $28.8^\circ$N–$46.2^\circ$N), including 13 major food producing provinces with a covering area of 1.77 million km$^2$, were chosen as the study area (Figure 1). The cultivated area of the study area reaches ~45% of the total cultivated area in China, and the grain output accounts for ~55% of the total national output. The study area is divided into four agricultural areas as: I. Middle-lower Yangtze Plain; II. Huang-Huai-Hai Plain; III. Northeast China Plain; and IV. Loess Plateau, according to China’s nine major agricultural regions. Some key soil properties (e.g., texture, sand/silt/clay fraction, and bulk density) for topsoil (0–30 cm) within the four agricultural areas are listed in Table 1. For those relatively humid areas as Middle-lower Yangtze Plain and Northeast China plain, soil texture is Medium/Fine, with higher clay fraction, lower sand fraction, and lower bulk density. Whilst, those relatively arid areas, particularly for the Loess Plateau, showed the highest sand fraction, lowest clay faction, and largest bulk density.

Table 1. Key physical soil properties for topsoil (0–30 cm) in four agricultural zoning areas. (Note that only three simplified textural classes were used: Coarse-, Medium-, and Fine-textured, due to the scale (1:5 million) of the Soil Map of the World, and that numbers in the brackets of the last column indicate the average value of bulk density).

| Soil Properties          | Texture          | Sand Fraction (%) | Silt Fraction (%) | Clay Fraction (%) | Bulk Density      |
|--------------------------|------------------|-------------------|-------------------|-------------------|-------------------|
| Middle-lower Yangtze Plain | Medium/Fine      | 38.1              | 37.7              | 24.2              | 1.18–1.74 (1.39)  |
| Huang-Huai-Hai Plain     | Medium           | 45.9              | 34.1              | 20.0              | 1.21–1.79 (1.43)  |
| Northeast China Plain    | Medium           | 40.1              | 37.5              | 22.4              | 1.21–1.71 (1.40)  |
| Loess Plateau            | Medium/Coarse    | 48.6              | 34.3              | 17.1              | 1.21–1.74 (1.44)  |
Land uses of those areas mainly include Cropland (CRO), Forestland (FOR), Grassland (GRA), Mixed land (MIX), and Barren land (BAR), based on the Remote Sensing Monitoring Data on Land Use of China that are generated in 2015. The land use data was downloaded from the Resources and Environment Data Cloud Platform (http://www.resdc.cn (accessed on 6 July 2021)) of the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences. Soil moisture, particularly that in rootzone areas, is highly related with crop growth in those agricultural areas, especially for the North China Plain areas, which has long been perturbed by a serious water shortage issue due to extensive irrigation [41]. Therefore, obtaining rootzone SM information is critically needed for drought forecasting, agricultural irrigation guidance, as well as water resources management within those areas [7,42].

2.2. In Situ SM Data

The in situ, ground station-based SM observations are fundamental datasets for understanding the true SM dynamics. A total of 796 SM monitoring stations are distributed in the study area, with the most densely distributed stations in Huang-Huai-Hai Plain (region II) (Figure 1). There are 419, 48, 39, 288, and 2 stations located in the Cropland, Forestland, Grassland, Mix land, and Barren land areas, respectively. Each station adopted the gravimetric method for SM measurement [10], which is generally recognized as the standard method. At all sites, four soil samples under 0–10 cm, 10–30 cm, and 30–50 cm surface layer were collected by a soil auger, and 3 parallel soil samples were analyzed for each soil sample. The soil samples were dried at 105 °C for 8 h in the oven and weighed after cooling. In terms of error control, the allowable error did not exceed 1% for parallel determination when SM is less than 5%, and that did not exceed 2% when SM is higher than 40%. The in-situ station observed SM data started from 1 January 2006 to 31 November 2017, sampling on the 1st, 11st, 21st day per month at 00:00 UTC.
According to Yang et al. [43], SM observation at three times per month could well represent the SM distribution within a month, which showed good reflection of the ground SM changes with local precipitation and evapotranspiration. Additionally, the in situ SM observations provide gravimetric units (kg/kg), while most satellite/reanalysis SM datasets provide volumetric units (m$^3$/m$^3$); thus, the first step is to uniform the units of different datasets. The soil bulk density data used for unit conversion is from the Harmonized World Soil Database (HWSD), of which the data source of China is the 1:1,000,000 soil data provided by Institute of Soil Science, Chinese Academy of Sciences, for the second national land survey. Specific descriptions on the calculating formulas can be seen in our previous work [43].

2.3. Satellite/Reanalysis SM Datasets

Based on the evaluation work of multi-source surface SM datasets over the central and eastern agricultural areas of China [43], several satellite products and reanalysis datasets were demonstrated with better spatial-temporal performances and higher estimation accuracy. Specially, ERA5 reanalysis dataset produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) had been proven the most promising alternatives for SM estimation [33]. ERA5 SM dataset covers the period from 1981 to the present, providing four-layer soil moisture data, with high spatial-temporal resolution (i.e., 1-hourly and 0.1°). A high-quality ERA5 dataset is available at 2–3 months before the present, and the latency of the preliminary dataset can be controlled within ~5 days.

The Soil Moisture Active Passive (SMAP) mission is an orbiting observatory to measure the amount of water in the surface soil, which was launched in January 2015 and started operation in April 2015. The enhanced level 3 radiometer global daily 9 km EASE-Grid surface soil moisture products generated from SMAP were chosen as the input to EF model in this work, covering the period from 1 April 2015 to 31 December 2017 [44]. Meanwhile, the level 4 global 3-hourly 9 km EASE-Grid surface and root zone soil moisture analysis update products (at depths of 0–100cm), that were assimilated and derived from Catchment Land Surface Model of SMAP [45], were added herein to compare with the EF-simulated root zone soil moisture. All SMAP L3 and L4 soil moisture products were downloaded from NASA Distributed Active Archive Center (DAAC) at National Snow and Ice Data Center (NSIDC). It should be noted that only SM at 00:00 UTC of SMAP L4 was chosen to match the observed SM time series (Figure S2). A summary of ERA5 and SMAP soil moisture datasets can be seen in Table 2.

### Table 2. Summary of ERA5 and SMAP soil moisture datasets.

| Dataset       | Retrieval/Assimilation Method            | Period          | Spatial Coverage | Temporal Resolution | Spatial Resolution | Depth               | Latency          | References/Links                                                                 |
|---------------|------------------------------------------|-----------------|------------------|---------------------|--------------------|---------------------|-------------------|-----------------------------------------------------------------------------------|
| ERA 5         | ECMWF-Integrated Forecast System         | 1981-present    | Global           | 1-hourly            | 0.1°               | 0–7 cm; 7–28 cm; 28–100 cm; 100–289 cm | 5 days            | Preliminary data; 3 months (Accurate data)                                       | [https://www.ecmwf.int/en/forecasts/datasets](https://www.ecmwf.int/en/forecasts/datasets) (accessed on 20 August 2021) |
| SMAP L3       | Backus-Gilbert Optimal Interpolation     | 2015-present    | Global           | Diurnal             | 9 km               | 0–5 cm              | Within 50 h       | [https://nsidc.org/data/smap/smap-data.html](https://nsidc.org/data/smap/smap-data.html) (accessed on 2 August 2021) |
| SMAP L4       | Ensemble Kalman Filter                   | 2015-present    | Global           | 3-hourly            | 9 km               | 0–100 cm            | ~7 days           |                                                                                 |

2.4. Methodology

**Exponential Filter Method.** According to Wagner et al. [36], water fluxes across the surface and the root zone layer are assumed proportional to the differences of volumetric SM between the two layers. That way, the water balance equations for the root zone SM can be simplified as

\[
L \frac{d \varnothing_R}{d_t} = C(\varnothing_S - \varnothing_R),
\]  

(1)
where $\varnothing_S$ is the surface SM (cm$^3$/cm$^3$); $\varnothing_R$ is the root zone SM (cm$^3$/cm$^3$); $L$ is the depth of root zone layer (cm); and $C$ is the pseudo diffusivity coefficient (cm/d), which is not only highly related with soil properties, but also influenced by plant species and meteorological conditions. Defining $T = L/C$ gives the following analytical solution of the Equation (1):

$$\varnothing_R(t) = \frac{1}{T} \int_{\infty}^{t} \varnothing_S(\tau) \exp\left(-\frac{t-\tau}{T}\right) d\tau,$$

where $T$ is the characteristic time length (d), which increases with the layer depth $L$, and decreases with the pseudo diffusivity coefficient $C$. The parameter $T$ is usually used as the time scale of SM variation, and is a multi-factor comprehensive parameter that affects the dynamic change of soil moisture [40]. Note that although transpiration processes are not involved in this simplified water balance model, and that soil hydraulic conductivity is also assumed constant in the calculating processes, such method can provide a useful tool for estimating profile SM content due to the weakening influences of measurements with increasing time lag [36].

Further, the discrete forms of Equation (2) are defined as the following formula (using the standardized SM data):

$$\varnothing_R(t_n) = \frac{\sum \varnothing_s(t_i) \exp\left(-\frac{t_n-t_i}{T}\right)}{\sum \exp\left(-\frac{t_n-t_i}{T}\right)},$$

where $\varnothing_R$ is the standardized rootzone SM (cm$^3$/cm$^3$); $\varnothing_s$ is the standardized surface SM (cm$^3$/cm$^3$). Herein, the standardization formula uses min-max normalization.

Albergel et al. [37] also proposed an iterative scheme of Equation (3) for further convenient calculation:

$$\varnothing_R(t_n) = \varnothing_R(t_{n-1}) + K_n[\varnothing_s(t_n) - \varnothing_R(t_{n-1})],$$

where the gain $K$ at time $t_n(K_n)$ is defined as the followings:

$$K_n = \frac{K_{n-1}}{K_{n-1} + \exp\left(-\frac{t_n-t_{n-1}}{T}\right)}.$$

The initial value is given as $K_1 = 1$, and $\varnothing_R(t_1) = \varnothing_s(t_1)$. It is noteworthy that although Equation (5) is an explicit recursive formula, and prone to accumulate estimation errors in theory; however, the root zone SM at time $t_n$ (i.e., $\varnothing_R(t_n)$) is estimated based on the surface SM at the corresponding time, not merely using the estimated value at $t_{n-1}$ (i.e., $\varnothing_R(t_{n-1})$), which can largely reduce the accumulated estimation errors.

**Estimation Scheme.** Herein, an estimation scheme was proposed to briefly outline the calculating steps for estimating RZSM using satellite surface SM based on the EF model (Figure 2). Two key steps were mainly involved in this scheme. The first step (i) was to calibrate and validate the EF model using the surface and root zone SM data of ERA5. When establishing the EF model, 2/3 of the total ERA5 SM data series duration was divided as calibration period (2006–2013) and the rest 1/3 was as the validation period (2014–2017). Specifically, the ERA5 SSM dataset (first layer, 0–7cm) was firstly adopted in EF model to simulate the RZSM in the range of $T$ ([1–50]) for each 10 km grid. Then, the Nash–Sutcliffe Efficient (NSE) was computed between the second layer of ERA5 SM (7–28cm) and all the simulated RZSMs, to determine the optimum $T$ of each grid. On this basis, an optimum $T$ map at 10 km grids was constructed, together with the simulated RZSM series corresponding to the optimal parameter $T$ for each grid (hereinafter referred to as ERA5-derived RZSM).
Figure 2. Estimation scheme for calibrating and validating EF models using ERA5 surface and root zone SM datasets (i), and estimating the root zone SM from SMAP L3 surface SM using the established EF model (ii).

The second step (ii) was to derive the RZSM from SMAP L3 surface SM dataset using the established EF model in step (i). Firstly, interpolate the optimum T map (~10 km) obtained in step (i) to match the scale of SMAP L3 SSM dataset (~9 km), using the inverse distance weighted (IDW) interpolation, to obtain a new optimum T map at 9 km grid. Then, the EF model was driven by the matched optimum parameter T and the SMAP L3 surface SM data, to estimate the RZSM at each 9 km grid (Hereinafter referred to as SMAP L3-derived RZSM). Lastly, all estimated RZSMs by EF model, and ERA5- and SMAP-provided RZSMs were compared and evaluated against the ground observed SM series, including the spatial-temporal and seasonal performances.

**Performance Metrics.** Herein, three statistical metrics as Nash-Sutcliffe Efficient (NSE), Relative Error (RE), and Root Mean Squared Error (RMSE) were adopted to evaluate the performance of EF model. The higher NSE, lower RE and RMSE, indicating the better
performance of EF model in simulating RZSM. Herein, NSE, RE, and RMSE was calculated according to the following formulas:

\[
\text{NSE} = 1 - \frac{\sum_{t=t_1}^{t_n} (\varnothing_{\text{ERA5}}^t - \varnothing_{\text{sim}}^t)^2}{\sum_{t=t_1}^{t_n} (\varnothing_{\text{ERA5}}^t - \overline{\varnothing_{\text{ERA5}}})^2}, \tag{6}
\]

where \(t_1\) and \(t_n\) is the start and end time of RZSM dataset series, \(\varnothing_{\text{ERA5}}^t\) and \(\varnothing_{\text{sim}}^t\) is the ERA5-provided and EF-simulated RZSM at time \(t\), and \(\overline{\varnothing_{\text{ERA5}}}\) is the average of ERA5-provided RZSM series.

\[
\text{RE} = \frac{(\varnothing_{\text{sim}} - \overline{\varnothing_{\text{ERA5}}})}{\overline{\varnothing_{\text{ERA5}}}}, \tag{7}
\]

\[
\text{RMSE} = \sqrt{\frac{\sum_{t=t_1}^{t_n} (\varnothing_{\text{ERA5}}^t - \varnothing_{\text{sim}}^t)^2}{n}}. \tag{8}
\]

Similarly, \(\varnothing_{\text{ERA5}}^t\) and \(\varnothing_{\text{sim}}^t\) indicates the ERA5-provided and model-simulated RZSM at time \(t\).

Moreover, statistical metrics as Correlation Coefficient (CC), \(p\) value, bias, and RMSE were chosen, to quantitatively evaluate the differentiation between model-simulated RZSM and the ground observations. The larger CC, lower bias and RMSE compared with the observations, the higher estimation accuracy of model-simulated RZSM. Of them, RMSE was calculated using the same formula with Equation (8), and Pearson correlation coefficient was computed using the following formula:

\[
\text{CC} = \frac{\sum_{i=1}^{m} (\varnothing_{\text{sim}} - \overline{\varnothing_{\text{sim}}}) (\varnothing_{\text{obs}} - \overline{\varnothing_{\text{obs}}})}{\sqrt{\sum_{i=1}^{m} (\varnothing_{\text{sim}} - \overline{\varnothing_{\text{sim}}})^2 \sum_{i=1}^{m} (\varnothing_{\text{obs}} - \overline{\varnothing_{\text{obs}}})^2}}, \tag{9}
\]

where \(\varnothing_{\text{sim}}\) is model-simulated RZSM value (m\(^3\)/m\(^3\)), \(\varnothing_{\text{obs}}\) is in-situ ground observed RZSM value (m\(^3\)/m\(^3\)), \(\overline{\varnothing_{\text{sim}}}\) and \(\overline{\varnothing_{\text{obs}}}\) is the average of model-simulated and observed RZSM series, and the difference between the former minus the latter was taken as bias:

\[
\text{bias} = \overline{\varnothing_{\text{sim}}} - \overline{\varnothing_{\text{obs}}}.
\]

The Hausdorff distance method was used to evaluate the distributed trajectory similarity of annual RZSM, which was computed as the following formula [46,47]:

\[
D_H(L_i, L_j) = \max \{h(L_i, L_j), h(L_j, L_i)\}, \tag{11}
\]

where

\[
h(L_i, L_j) = \max (a \in L_i) \min (b \in L_j) \parallel a - b \parallel;
\]

\[
h(L_j, L_i) = \max (a \in L_j) \min (b \in L_i) \parallel a - b \parallel;
\]

\(\parallel \cdot \parallel\) denotes the distance norm between line \(L_i\) and \(L_j\).

Usually, the Hausdorff distance reflects the maximum mismatch of two distributed trajectories: the larger \(D_H\), the higher mismatch between the two trajectories; and vice versa. Specifically, herein the larger Hausdorff distance between the observed and one provided/model-simulated RZSM dataset, the worse performance of this dataset in reflecting the annual RZSM changes.
3. Results

3.1. Exponential Filter Model Calibration and Validation

3.1.1. Distribution of Optimum T Parameter

The EF model was calibrated and validated using ERA5 SSM and RZSM datasets for each 10 km grid, obtaining the distribution map of the optimum T parameter (T\text{opt}) (Figure 3). In the upper left box diagram, T\text{opt} for all grids were in the range of [2, 4], with an average of ~3 days. In terms of spatial T\text{opt} distribution, the optimum T parameter showed an increasing trend from the eastern humid areas to the western semi-humid areas: the minimum T\text{opt} occurred in the southern Middle-lower Yangtze Plain and northeastern Northeast China Plain areas (humid areas with annual rainfall >800 mm), with a value of 1–3 days; while larger T\text{opt} was found in those semi-humid (Huang-Huai-Hai Plain) and semi-arid (Loess Plateau) areas, with most values in 4–10 days. As mentioned above, T\text{opt} is a comprehensive parameter affected by multiple factors, which is not only related to the climatic conditions and vegetation in the region, but also affected by the soil properties. For instance, the fine texture with low sand content and high clay content, will affect the movement and stagnation of soil water, which can be reflected in the parameter T\text{opt}.

![Figure 3. Distributions of optimum parameter T calibrated at 10 km grid. (Note that in the upper left violin diagrams are the distributions of the optimum parameter T for all grids).](image)

3.1.2. Overall Performances

For all 10 km grids, the performance of EF model was evaluated using the three statistical metrics, i.e., NSE, RE, and RMSE. Overall, EF model showed good performances both in the calibration and validation period, as presented by the violin diagrams (Figure 4). The average NSE (NSE) of all grids reached 0.82 and 0.78 in calibration (2006–2013) and validation (2014–2017) period, respectively, indicating a good reliability of EF model calibrated using ERA5 datasets in those study areas. The EF model also had low RE and RMSE both in the calibration and validation period (RE: ~10%; RMSE: ~0.08 m\textsuperscript{3}/m\textsuperscript{3}), indicating low bias between the simulated RZSM and the ERA5-provided RZSM at ~10–30 cm depth.
3.2. Evaluation on ERA5-Derived RZSM against In Situ Observations

In this section, the ERA5-derived RZSM using the EF model was evaluated against the in situ observed RZSM (10–30 cm). Specifically, the spatial-temporal comparisons between ERA5-derived and observed RZSM were performed at a regional scale. Furthermore, a quantitative evaluation was carried out by interpolating ERA5-derived RZSM dataset from grid to each observation station.

3.2.1. Spatial Comparison between Observed and ERA5-Derived SM

By contrast with the in situ observed RZSM, both ERA5-provided RZSM and ERA5-derived RZSM by the EF model can capture the spatial variations of RZSM in a large part (Figure 5): spatial distributions of RZSM were effectively estimated in the study area, despite the overestimation in some marginal areas, e.g., the southern Middle-lower Yangtze Plain. Meanwhile, ERA5-derived RZSM by the EF model showed similar spatial performance as the ERA5-provided RZSM itself in estimating RZSM, which even presented higher estimation accuracy for some semi-humid and -arid areas against the in situ observed RZSM. Despite the slight difference between the depth of 7–28 cm and 10–30 cm, soil moisture will not show much variations between the two depths; thus, the in situ ground RZSM observations at 10–30 cm could act as the approximate “ground reference” to evaluate the ERA5-provided and -derived RZSMs at 7–28 cm.

Figure 4. Overall performance metrics of the EF model used for each 10 km grid in calibration and validation period, respectively.
3.2.2. Temporal Comparison among Different Agricultural Zoning Areas

The temporal changes of ERA5-derived RZSM by the EF model were compared with the in situ observed RZSM in Middle-lower Yangtze Plain, Huang-Huai-Hai Plain, Northeast China Plain, and Loess Plateau, respectively, by computing four statistical metrics (i.e., CC, bias, RMSE, and NSE) between the two RZSM series (Figure 6). For all agricultural areas, the EF-simulated RZSM could well capture the temporal changes of in situ observed RZSM, as presented by the large CC (all >0.7) between the simulated RZSM and the observations. Meanwhile, the bias and RMSE between ERA5-derived RZSM and observed RZSM varied with different agricultural zonings, which were minimum for Huang-Huai-Hai Plain (bias: $-0.003 \text{ m}^3/\text{m}^3$; RMSE: 0.03 m$^3$/m$^3$) and relatively larger for the other areas (bias: 0.047–0.079 m$^3$/m$^3$; RMSE: 0.06–0.08 m$^3$/m$^3$). Despite the low NSE before bias calibration, higher NSE (0.37–0.70) occurred after eliminating bias of ERA5-derived and -provided RZSM against the in situ observed RZSM. By contrast, the temporal changes of ERA5-derived RZSM by the EF model were close to that of ERA5-provided RZSM, which was further verified by the similar performance metrics (Figure 6), indicating that ERA5-derived and -provided RZSM series had similar performances on depicting the RZSM changes with time in different agricultural areas.

3.2.3. Quantitative Comparison against Ground Observation Sites

The ERA5-derived RZSM at ~10km grid was interpolated to each observation station, which was then quantitatively evaluated against the in situ observed RZSM series. The overall statistical results of all observed stations and regional statistics within four agricultural zoning areas were presented by the following violin diagrams (Figure 7). The average CC ($\overline{CC}$) between ERA5-derived RZSM and the ground observed RZSM series were in the range of 0.42–0.6, fluctuating around 0.5, and the corresponding $p$ values were below 0.05, as presented by the 1st quantile Q1 which was under the red short dot line ($p = 0.05$). Thereinto, the humid areas, i.e., Middle-lower Yangtze Plain, showed the largest CC with the observed RZSM series ($\overline{CC} = 0.60$) and the lowest $p$ value (far below the line of $p = 0.05$); meanwhile, the Loess Plateau (semi-arid areas) showed the lowest CC (0.29–0.53, $\overline{CC} = 0.42$). The bias and RMSE between ERA5-derived RZSM and the ground observed RZSM series were in the range of $-0.01$–$0.04 \text{ m}^3/\text{m}^3$ and $0.07$–$0.09 \text{ m}^3/\text{m}^3$, respectively, indicating that the estimation error of ERA5-derived RZSM was small compared to the in situ measured RZSM. By contrast, the simulated RZSM by EF model could attain a good estimation accuracy as that of ERA5-provided RZSM (Figure S1).
Figure 6. Temporal comparison of ERA5-derived RZSM compared to the in-situ ground observations and ERA5-provided RZSM in (a) Middle-lower Yangtze Plain, (b) Huang-Huai-Hai Plain, (c) Northeast China Plain, and (d) Loess Plateau. (Herein, CC, RMSE, and NSE indicates the performance metrics of ERA5-derived (in red) and -provided (in green) RZSM against the observed RZSM series; and ** indicates significant correlation at \( p < 0.001 \)).

Figure 7. Distribution violin diagrams of CC (a), \( p \) value (b), bias (c), and RMSE (d) between ERA5-derived RZSM and in situ ground observations. Presented are the median, the 1st quantile Q1 and 3rd quantile Q3 (as indicated by the box), and the Q1 – 1.5(Q3 – Q1) and Q3 + 1.5(Q3 – Q1) values (whiskers). (The red dash line in (b) indicates \( p = 0.05 \)).
3.3. Evaluation of Root-Zone SM Estimated from SMAP L3 Surface SM

This section attempted to derive the RZSM from SMAP L3 SSM using the abovementioned established ERA5-based EF model, and on basis of which evaluated the SMAP L3-derived RZSM against in situ observed RZSM.

3.3.1. Temporal Comparison between Observed and SMAP L3-Derived RZSM

Four statistical metrics (i.e., CC, bias, RMSE, and NSE) compared to the observed RZSM series were computed to evaluate the temporal performance of SMAP L3-derived RZSM (Table 3). Overall, SMAP L3-derived RZSM in all agricultural areas could well capture the temporal changes of in situ observed RZSM, as presented by the large CC (CC: 0.53–0.82, p < 0.001) between the simulated RZSM and the observations. Furthermore, the bias and RMSE between SMAP L3-derived RZSM and observed RZSM varied with different agricultural zonings, which were small for Middle-lower Yangtze Plain and Northeast China Plain (Humid areas) but large for Huang-Huai-Hai Plain and Loess Plateau areas (semi-humid/-arid areas). This is consistent with those conclusions that EF model provides better performances under humid conditions [14]. Meanwhile, lower NSEs (a few negative) were found for all agricultural zonings before bias calibration; however, larger NSE (0.35–0.48) could be attained once the bias of SMAP L3-derived RZSM were corrected compared to the observed RZSM. By contrast with the SMAP L4-provided RZSM product assimilated using the EnKF method, the SMAP L3-derived RZSM by the EF model had similar performance metrics for all agricultural areas.

Table 3. Performance metrics of SMAP L3-derived and L4-provided RZSMs against the in-situ ground observations in five different agricultural zoning areas. (Herein, all statistical metrics were calculated after bias correction, and ** indicates significant correlation at p < 0.001).

| Statistical Indexes          | RZSM                      | CC     | RMSE (m³/m³) | NSE  |
|------------------------------|---------------------------|--------|--------------|------|
| Middle-lower Yangtze Plain   | SMAP L3-derived by EF     | 0.58 **| 0.02         | 0.34 |
|                             | SMAP L4-provided          | 0.76 **| 0.02         | 0.4  |
| Huang-Huai-Hai Plain         | SMAP L3-derived by EF     | 0.82 **| 0.05         | 0.48 |
|                             | SMAP L4-provided          | 0.83 **| 0.02         | 0.51 |
| Northeast China Plain        | SMAP L3-derived by EF     | 0.53 **| 0.03         | 0.35 |
|                             | SMAP L4-provided          | 0.55 **| 0.03         | 0.33 |
| Loess Plateau                | SMAP L3-derived by EF     | 0.66 **| 0.05         | 0.41 |
|                             | SMAP L4-provided          | 0.77 **| 0.02         | 0.46 |

3.3.2. Accuracy Evaluation Using Ground Observation Sites

Similarly, the SMAP L3-derived RZSM at 10 km grid was interpolated to each observation station, and was then quantitatively evaluated using the in situ observed RZSM series. The following box diagrams showed the overall statistical results of all observed stations and regional statistics within four agricultural zoning areas (Figure 8). The average CC (CC) between SMAP L3-derived RZSM and the ground observed RZSM series were in the range of 0.38–0.45, fluctuating around 0.39, and the corresponding p values were below 0.05. Thereinto, those humid areas, i.e., Northeast China Plain and Middle-lower Yangtze Plain showed larger CC with the observed RZSM series (CC = 0.39, 0.45); meanwhile, the Huang-Huai-Hai Plain showed the lowest CC (CC: ~ 0.36). The bias and RMSE between SMAP L3-derived RZSM and the ground observed RZSM series were small in those humid areas, with bias = −0.04, −0.06 m³/m³, and RMSE = 0.07, 0.11 m³/m³; while, bias and RMSE were relatively large for those semi-humid/-arid areas, with bias = −0.1, −0.12 m³/m³, and RMSE = 0.11, 0.13 m³/m³. By contrast, SMAP L3-derived RZSM showed similar performance metrics with that of SMAP L4-provided RZSM product (including the interpo-
lated results) for all agricultural areas, particularly for Loess Plateau areas, indicating their similar estimation accuracy against the observed RZSM series.

Figure 8. Distribution box diagrams of CC (a), bias (b), and RMSE (c) between SMAP L3-derived and SMAP L4-provided RZSMs and in situ ground observations among different agricultural zonings. Presented are the median, the 1st quantile Q1 and 3rd quantile Q3 (as indicated by the box), and the Q1 – 1.5(Q3 – Q1) and Q3 + 1.5(Q3 – Q1) values (whiskers).

Moreover, the CC, bias and RMSE compared to the observed RZSM series between SMAP L3-derived and SMAP L4-provided (after interpolation) RZSM showed similar spatial distributions within the whole evaluated area (Figure 9). Particularly, SMAP L3-derived RZSM even outperformed the SMAP L4-provided products in the central part of Northeast China Plain (SMAP L3-derived CC: 0.49; SMAP L4-provided CC: 0.41), where the latter cannot cover at that moment.
Additionally, one typical area located in the northern part of Haihe River basin in China was zoomed in from the spatial error map (Figures 9 and 10). Larger estimation errors and uncertainties were found in those Cropland areas, where the underlying surface are strongly perturbed by complex human activities, e.g., mining groundwater, agricultural irrigation, which may largely influence the supply and transport conditions of soil water.
3.3.3. Seasonality of Estimated RZSMs

The seasonality of all estimated RZSMs were further compared with the in situ observations in Huang-Huai-Hai Plain areas (Figure 11). Herein, Huang-Huai-Hai Plain was chosen as the typical case mainly considered the most densely distributed SM observation stations in those areas that can best represent the ground reference of SM dynamics [43]. The Hausdorff distances and Person CC were computed as performance metrics of reflecting the seasonality by different RZSM versions. Of these RZSM trajectories, the ERA5-derived and -provided RZSMs showed similar performance in reflecting seasonal changes of RZSM, as presented by similar CC of 0.96 and 0.95 (p < 0.001) with the observed RZSM series. The Hausdorff distances between ERA5-derived and -provided RZSMs away from the observed RZSM trajectory also proved their similar seasonality reflection performance, reaching 0.36 (H₁) and 0.33 (H₂), respectively. Furthermore, the SMAP L3-derived RZSM showed similar performance with the SMAP L4-provided RZSM in reflecting seasonal RZSM distribution: the CC of SMAP L3-derived and SMAP L4-provided RZSM with the observed trajectory reached 0.74 and 0.77 (p < 0.001); and the corresponding Hausdorff distances were 1.5 (H₃) and 1.06 (H₄), respectively.
Figure 11. Seasonality of soil moisture distribution among different RZSMs estimated from multi-source SM datasets. (Herein \( H_1, H_2, H_3, \) and \( H_4 \) indicates the Hausdorff distance of ERA5-derived, ERA5-provided, SMAP L3-derived, and SMAP L4-provided RZSM against the observed RZSM trajectories, respectively, and note that all data series were standardized using the Z-score method).

4. Discussion

This work proposed a rapid estimation scheme for estimating RZSM with short latency and small computations, and tested over the central and eastern agricultural areas of China. Two main calculating procedures were involved in the scheme. The calibrated and validated EF model in the first step could successfully attain as good performance on estimating RZSM as the ERA5 reanalysis product itself. Firstly, the calibrated optimum parameter \( T (T_{\text{opt}}) \) showed a reasonable distribution with different wet-dry conditions, as presented by an increasing trend from the eastern humid areas (1–3 days) to the western semi-humid areas (4–10 days) (Figure 2). The single parameter \( T \) is simple to calibrate, however, giving a physically-based explanation on this parameter has long remained a key problem to be solved in previous studies [37,40,48]. From the model perspective, the larger parameter \( T \), the smaller \( K_n \), indicating that the RZSM at \( t_n (\theta_R(t_n)) \) relies less on the surface SM at \( t_n (\theta_s(t_n)) \) but more on the RZSM at the last period (\( \theta_R(t_{n-1}) \)); and vice versa (as shown in Equations (4) and (5)). Parameter \( T \) could be regarded as a “pseudodiffusivity” that controls how much of the SSM infiltrates into the root zone layers and the retention time [39]. Physically, for those humid areas, e.g., Middle-lower Yangtze Plain (annual rainfall >800 mm) with mechanism of excess storage runoff, where root-zone soil moisture is highly related with the surface soil moisture due to its role in partitioning how much rainfall to deeper soil; while, those relatively arid areas often present excess infiltration, where deeper-layer soil moisture rarely rests on the changes of surface soil moisture. Consistently, the spatial distributions of \( T_{\text{opt}} \) by EF model well reflected the varying hydraulic properties of underlying surface. Such relationships between the surface and root zone soil moisture among different wet-dry areas were also consistent with the previous report based on the China Ecosystem Research Network [49]. The good reflection of hydrological and meteorological conditions by model parameters may further strengthen the physical mechanism of EF methods to some extent [19].

Secondly, the applicability of this calibration approach was demonstrated by the overall performance indexes as well as the comparison against the in situ observed RZSM. On one hand, the EF model for all calculating girds showed high NSE (\( \text{NSE} : 0.82, 0.78 \)), and low RE (\( \text{RE:} \sim 10\% \text{m}^3/\text{m}^3 \)) and RMSE (\( \sim 0.08 \text{m}^3/\text{m}^3 \)) both in calibration and validation period (Figure 3), which indicated that the calibration and variation using ERA5-provided SSM and RZSM datasets was desirable for the EF model in those study areas. The applicability of such calibration of EF methods using reanalysis datasets was also proved by
other studies [39]. On the other hand, by contrast with the in situ observed RZSM series at 10–30 cm depth, the ERA5-derived RZSM (7–28 cm) by the EF model could capture the temporal-spatial variations of RZSM in different agricultural zonings, as presented by the large CC (all >0.7), low bias (|bias| < 0.08 m$^3$/m$^3$), and RMSE (all <0.08 m$^3$/m$^3$), as well as the high NSE (0.37–0.61) between the simulated and observed RZSM series (Figures 4 and 5). The quantitative evaluation results at each observation sites further verified the good estimation accuracy of ERA5-derived RZSM with average CC, bias and RMSE reaching the range of 0.42–0.6, 0.01–0.04 m$^3$/m$^3$, and 0.07–0.09 m$^3$/m$^3$, respectively (Figure 6). Moreover, the ERA5-derived RZSM by the EF model could well reflect the seasonal changes of in situ observed RZSM, due to the short Hausdorff distances with the observed (H = 0.36) (Figure 11). The good correspondence between the simulated RZSM and the observations at ~30 cm depth was also found in previous studies [16,39,50,51]. It should be noted that all those abovementioned evaluation results of ERA5-derived RZSM were also compared with that of ERA5-provided RZSM itself, indicating that the two RZSM series had similar performances in various aspects. Overall, the simulated RZSM by EF model could be a desirable alternative of ERA5-provided RZSM, considering its efficiency in rapidly obtaining RZSM, and thus solving the longer data latency and large computation costs of ERA5 in land surface models [25].

Necessarily, a reliable input of surface soil moisture dataset, which should be accessible with short data latency (real time is the best if possible), is the critical issue for EF methods to be solved [16,37]. This work adopted the fast-updating SMAP L3 SSM product (assured latency within 50 h) as the input to EF models. By contrast with the in situ observed RZSM, SMAP L3-derived RZSM by EF model presented good performances on capturing the temporal RZSM changes over all agricultural areas (Table 3); meanwhile, the quantitative evaluation for each observed site also proved the good estimation accuracy of SMAP L3-derived RZSM (Figure 7). More importantly, all those evaluation results of SMAP L3-derived RZSM by the EF model showed as good estimation accuracy as that of SMAP L4-provided RZSM dataset, together with the performances in capturing temporal RZSM changes and reflecting seasonality of RZSM (Figure 8). Although SMAP L3-derived RZSM was slightly inferior to SMAP L4 product in some agricultural areas (e.g., Middle-lower Yangtze Plain), this is likely due to the inherent discrepancy of surface SM (0–5 cm) provided by the two products [7,39]. The SMAP L4 RZSM dataset is known as the assimilation product merged from the lower-level SMAP data using the EnKF technique [22,32], which considers the uncertainties from both model product and satellite observations, and therefore, is found superior to model and satellite data alone [52]. Nevertheless, SMAP L4-provided products are not full area coverage (Figure S2), in some specific areas, e.g., the central part of Northeast China Plain, SMAP L3-derived RZSM even outperformed the interpolated SMAP L4-provided RZSM (Figure 9). Additionally, Cropland areas with complex human activities, e.g., crop planting and irrigating, which may cause larger estimation errors and uncertainties due to the influences on the supply, transport, and conduction conditions of soil water.

It should be noted that the RZSM depth of SMAP L4 was defined as soil moisture in the top 1 m of the soil column [32]; however, in most cases for those agricultural areas, a layer of 10–30 cm depth soil is primarily concerned due to its important role for crop growth [12,53]. According to Raza et al. [12], most temperate crop roots grow at the uppermost ~15 cm depth of the soil, and further indicated that 61–78% of the root biomass of various crops were found in the top 30 cm layer of soil. At this point, SMAP L3-derived RZSM by the EF model could provide more direct information than SMAP L4 RZSM at the 10–30 cm depth of crop root areas. Adding the data latency of SMAP L4 RZSM product (latency of various input data plus processing time, assured latency within seven days) and the high computation costs of the EnKF assimilation technique in process-based models [25,54], the estimation scheme proposed in this work might be a more convenient and rapid approach appropriate for estimating RZSM over large agricultural areas [16].
In sum, the proposed calculation scheme in this work for estimating root zone soil moisture using the EF method was initially driven by the surface and root-zone SM dataset of ERA5, which attained as good performances as ERA5-provided RZSM itself in various aspects. The input of almost real-time SMAP L3 SSM dataset, together with the good estimation accuracy of SMAP L3-derived RZSM, facilitated the estimation scheme based on the EF method to be a desirable alternative for estimating RZSM with short data latency and small computation. Under the smart agricultural initiative in modern agriculture, such estimation scheme can effectively utilize the latest released satellite SM products, and provide timely guidance for agricultural water management, e.g., drought monitoring for crop growth.

One thing should be kept in mind that although only one parameter T needed for calibration does save much computation, the physical explanation of T needs to be further considered. Parameter T was found to be related with all physical process affecting the soil moisture dynamics, e.g., evapotranspiration, the hydraulic characteristics of the soil, texture, density, thickness, and the number of soil layers [40]. Considering the influences of those processes when calibrating parameter T for each grid in future may imply more specific illustrations on the physical mechanisms of the EF model. This may contribute to improve the worse performances of EF model in relatively arid areas than those humid areas. In addition, bias correction or multi-source data fusion can enhance data quality of input to the EF model, which may further attain higher estimation accuracy of RZSM.

5. Conclusions

This work established a rapid estimation scheme suited for estimating root zone soil moisture with short latency and small computations over central and eastern agricultural areas of China. First, the EF model was calibrated and validated using the surface and root-zone SM of ERA5 reanalysis dataset, to obtain the optimum parameter T for each grid based on the largest NSE between the ERA5-derived RZSM by the EF model and ERA5-provided RZSM. Second, the ERA5-derived RZSM by the EF model was compared with the observed RZSM series in different agricultural areas, including spatial-temporal and seasonal performances from site to regional scale. Finally, a scale-matched optimum T value together with SMAP L3 SSM was used as inputs into the EF model to retrieve RZSM estimation for each grid, and the performances of SMAP L3-derived RZSM series were also quantitatively evaluated against the in situ observations. The major conclusions of this work are as follows:

1. The calibrated optimum parameter T showed an increasing trend from the eastern humid areas (1–3 days) to the western semi-humid areas (4–10 days), which is in line with the mechanism of local runoff generation, verifying the physical mechanism of the EF model to some extent;
2. The applicability of the calibration approach using ERA5 SSM and RZSM dataset was demonstrated: (1) EF model in all calculating girds showed high NSE (NSE : 0.82, 0.78), and low RE (RE: ~10% m$^3$/m$^3$) and RMSE (~0.08 m$^3$/m$^3$) both in calibration and validation period; (2) EF-simulated RZSM could capture the temporal-spatial and seasonal variations of RZSM by comparison with the in situ observed RZSM series among different agricultural zonings, as presented by the large CC (all >0.7), low bias (|bias| < 0.08 m$^3$/m$^3$) and RMSE (all <0.08 m$^3$/m$^3$), as well as the high NSE (0.37–0.61) between the simulated and observed RZSM series;
3. The SMAP L3-derived RZSM by the EF model presented good performances on capturing the temporal RZSM changes over all agricultural areas. Moreover, the quantitative evaluation at each observed site also proved the good estimation accuracy of SMAP-derived RZSM. SMAP L3-derived RZSM even outperformed the interpolated SMAP L4-provided RZSM in some specific areas;
4. The fast-updating SMAP L3 SSM product facilitated the proposed estimation scheme a desirable alternative for estimating RZSM with short data latency and high computa-
tional efficiency. Such estimation scheme presents a distinct advantage in agricultural water management under the modern smart agriculture initiative.

**Supplementary Materials:** The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/rs14081785/s1, Figure S1: Comparison of ERA5-provided and -derived RZSM series against the in situ observed RZSM for all observation sites of the study area; Figure S2: Covering scale of SMAPL3-derived (a) and SMAPL4-provided (b) RZSM datasets at the same time (at 00:00 UTC) within the study area.

**Author Contributions:** Conceptualization, Z.B. and J.Z.; methodology, Y.Y.; software, H.W.; validation, J.W.; formal analysis, J.W.; investigation, H.W.; resources, G.W.; data curation, Y.Y.; writing—original draft preparation, Y.Y.; writing—review and editing, Y.Y.; visualization, C.L.; supervision, J.Z. and G.W.; project administration, Z.B. and G.W.; funding acquisition, Z.B., C.L. and J.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is supported by the National Key Research and Development Program of China (grant numbers 2017YFA0605002), National Natural Science Foundation of China (grant numbers 41961124007, 51779145, 41830863, 51879164), “Six top talents” in Jiangsu province (grant no. RJFW-031).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Acknowledgments:** This work was supported by the National Key Research and Development Program of China (grant numbers 2017YFA0605002), National Natural Science Foundation of China (grant numbers 41961124007, 51779145, 41830863, 51879164), “Six top talents” in Jiangsu province (grant no. RJFW-031). The authors would like to thank the teams from NASA and ECMWF for making their datasets publicly available. Specifically, the authors would like to thank ECMWF for the ERA5 SM datasets generated using Copernicus Climate Change Service Information; and NASA for providing access to the SMAP L3 and L4 SM datasets.

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

**References**

1. Brocca, L.; Morbidelli, R.; Melone, F.; Moramarco, T. Soil moisture spatial variability in experimental areas of central Italy. *J. Hydrol.* 2007, 333, 356–373. [CrossRef]
2. Corradini, C. Soil moisture in the development of hydrological processes and its determination at different spatial scales. *J. Hydrol.* 2014, 516, 1–5. [CrossRef]
3. De Rosnay, P.; Balsamo, G.; Albergel, C.; Muñoz-Sabater, J.; Isaksen, L. Initialisation of Land Surface Variables for Numerical Weather Prediction. *Surf. Geophys.* 2014, 35, 607–621. [CrossRef]
4. Bindlish, R.; Crow, W.T.; Jackson, T.J. Role of Passive Microwave Remote Sensing in Improving Flood Forecasts. *IEEE Geosci. Remote Sens. Lett.* 2009, 6, 112–116. [CrossRef]
5. Koster, R.D.; Mahanama, S.P.P.; Livneh, B.; Lettenmaier, D.P.; Reichle, R.H. Skill in streamflow forecasts derived from large-scale estimates of soil moisture and snow. *Nat. Geosci.* 2010, 3, 613–616. [CrossRef]
6. Enenkel, M.; Steiner, C.; Mistelbauer, T.; Dorigo, W.; Wagner, W.; See, L.; Atzberger, C.; Schneider, S.; Rogenhofer, E. A Combined Satellite-Derived Drought Indicator to Support Humanitarian Aid Organizations. *Remote Sens.* 2016, 8, 340. [CrossRef]
7. Tavakol, A.; Rahmani, V.; Quiring, S.M.; Kumar, S.V. Evaluation analysis of NASA SMAP L3 and L4 and SPoRT-LIS soil moisture data in the United States. *Remote Sens. Environ.* 2019, 229, 234–246. [CrossRef]
8. El Hajji, M.; Baghdadi, N.; Zribi, M.; Belaud, G.; Cheviron, B.; Courault, D.; Charron, F. Soil moisture retrieval over irrigated grassland using X-band SAR data. *Remote Sens. Environ.* 2016, 176, 202–218. [CrossRef]
9. Seneviratne, S.I.; Corti, T.; Davin, E.L.; Hirschi, M.; Jaeger, E.B.; Lehner, I.; Orlowski, B.; Teuling, A.J. Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Sci. Rev.* 2010, 99, 125–161. [CrossRef]
10. Romano, N. Soil moisture at local scale: Measurements and simulations. *J. Hydrol.* 2014, 516, 6–20. [CrossRef]
11. Brocca, L.; Ciabatta, L.; Massari, C.; Camici, S.; Tarpanelli, A. Soil Moisture for Hydrological Applications: Open Questions and New Opportunities. *Water* 2017, 9, 140. [CrossRef]
12. Raza, A.; Friedel, J.K.; Moghaddam, A.; Ardakani, M.R.; Loiskandl, W.; Himmelbauer, M.; Bodner, G. Modeling growth of different Lucerne cultivars and their effect on soil water dynamics. *Agric. Water Manag.* 2013, 119, 100–110. [CrossRef]
13. Chakrabarti, S.; Bongiovanni, T.; Judge, J.; Zotarelli, L.; Bayer, C. Assimilation of smos soil moisture for quantifying drought impacts on crop yield in agricultural regions. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2017, 7, 3867–3879. [CrossRef]
14. Tobin, K.J.; Torres, R.; Crow, W.T.; Bennett, M.E. Multi-decadal analysis of root-zone soil moisture applying the exponential filter across CONUS. *Hydrol. Earth Syst. Sci.* 2017, 21, 4403–4417. [CrossRef]
15. Lia, F.; Crow, W.T.; Kustas, W.P. Towards the estimation root-zone soil moisture via the simultaneous assimilation of thermal and microwave soil moisture retrievals. *Adv. Water Resour.* **2010**, *33*, 201–214. [CrossRef]

16. Ford, T.W.; Harris, E.; Quiring, S.M. Estimating root zone soil moisture using near-surface observations from SMOS. *Hydrol. Earth Syst. Sci.* **2014**, *18*, 139–154. [CrossRef]

17. Dumedah, G.; Walker, J.P.; Merlin, O. Root-zone soil moisture estimation from assimilation of downscaled Soil Moisture and Ocean Salinity data. *Adv. Water Res.* **2015**, *74*, 14–22. [CrossRef]

18. Mishra, V.; Lee, E.W.; Markert, K.N.; Limaye, A.S. Performance evaluation of soil moisture profile estimation through entropy-based and exponential filter models. *Hydrol. Sci. J.* **2020**, *65*, 1036–1048. [CrossRef]

19. Manfreda, S.; Brocca, L.; Moramarco, T.; Melone, F.; Sheffield, J. A physically based approach for the estimation of root-zone soil moisture from surface measurements. *Hydrol. Earth Syst. Sci.* **2014**, *18*, 1199–1212. [CrossRef]

20. Maggioni, V.; Reichle, R.H.; Anagnostou, E.N. The efficiency of assimilating satellite soil moisture retrievals in a land data assimilation system using different error models. *J. Hydrometeorol.* **2013**, *14*, 368–374. [CrossRef]

21. Tian, S.; Tregoning, P.; Renzullo, L.J.; van Dijk, A.I.J.M.; Walker, J.P.; Pauwels, V.R.N.; Allgeyer, S. Improved water balance component estimates through joint assimilation of GRACE water storage and SMOS soil moisture retrievals. *Water Resour. Res.* **2017**, *53*, 1820–1840. [CrossRef]

22. Reichle, R.H.; Liu, Q.; Koster, R.D.; Crow, W.T.; De Lannoy, G.J.M.; Kimball, J.S.; Ardizzone, J.V.; Bosch, D.; Colliander, A.; Cosh, M.; et al. Version 4 of the SMAP Level-4 Soil Moisture Algorithm and Data Product. *J. Adv. Model Earth Syst.* **2019**, *11*, 3106–3130. [CrossRef]

23. Kornelsen, K.C.; Coulibaly, P. Root-zone soil moisture estimation using data-driven methods. *Water Resour. Res.* **2014**, *50*, 2946–2962. [CrossRef]

24. Grillakis, M.G.; Koutroulis, A.G.; Alexakis, D.D.; Polykretis, C.; Daliakopoulos, I.N. Regionalizing root-zone soil moisture estimates from ESA CCI Soil Water Index using machine learning and information on soil, vegetation, and climate. *Water Resour. Res.* **2021**, *57*, e2020WR029249. [CrossRef]

25. Li, Q.; Wang, Z.; Shangguan, W.; Li, L.; Yao, Y.; Yu, F. Improved Daily SMAP Satellite Soil Moisture Prediction over China using deep learning model with transfer learning. *J. Hydrol.* **2021**, *600*, 126698. [CrossRef]

26. Dumedah, G.; Walker, J.P. Evaluation of model parameter convergence when using data assimilation in soil moisture estimation. *J. Hydrometeorol.* **2014**, *15*, 359–375. [CrossRef]

27. Sabater, J.M.; Jarlan, L.; Calvet, J.C.; Bouyssel, F.; De Rosnay, P. From Near-Surface to Root-Zone Soil Moisture Using Different Assimilation Techniques. *J. Hydrometeorol.* **2007**, *8*, 194. [CrossRef]

28. Walker, J.P.; Willgoose, G.R.; Kalma, J.D. One-dimensional soil moisture profile retrieval by assimilation of near-surface observations: A comparison of retrieval algorithms. *Adv. Water Resour.* **2001**, *24*, 631–660. [CrossRef]

29. Clark, M.P.; Rupp, D.E.; Woods, R.A.; Zheng, X.; Ibbita, R.P.; Slater, A.G.; Schmida, J.; Uddstroma, 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]

30. Reichle, R.H.; De Lannoy, G.J.M.; Liu, Q.; Ardizzone, J.V.; Colliander, A.; Conaty, A.; Crow, W.; Jackson, T.J.; Jones, L.A.; Kimball, J.S.; et al. Assessment of the SMAP Level-4 Surface and Root-Zone Soil Moisture Product Using In Situ Measurements. *J. Hydrometeorol.* **2017**, *18*, 2621–2645. [CrossRef]

31. Reichle, R.H.; De Lannoy, G.J.M.; Liu, Q.; Koster, R.D.; Kimball, J.S.; Crow, W.T.; Ardizzone, J.V.; Chakraborty, P.; Collins, D.W.; Conaty, A.L.; et al. Global Assessment of the SMAP Level-4 Surface and Root-Zone Soil Moisture Product Using Assimilation Diagnostics. *J. Hydrometeorol.* **2017**, *18*, 3217–3237. [CrossRef] [PubMed]

32. Reichle, R.H.; De Lannoy, G.; Koster, R.D.; Crow, W.T.; Kimball, J.S.; Liu, Q. SMAP L4 Global 3-hourly 9 km EASE-Grid Surface and Root Zone Soil Moisture Analysis Update, Version 5 [Indicate Subset Used]; NASA National Snow and Ice Data Center Distributed Active Archive Center: Boulder, CO, USA, 2020. [CrossRef]

33. Muñoz Sabater, J. ERA5-Land Hourly Data from 1981 to Present: Copernicus Climate Change Service (C3S) Climate Data Store (CDS). 2019. Available online: https://cds.climate.copernicus.eu/cdsapp#!/dataset/10.24381/cds.e2161bac?tab=overview (accessed on 20 August 2021).

34. Andini, N.; Kim, D.; Chun, J.A. Operational soil moisture modeling using a multi-stage approach based on the generalized complementary principle. *Agric. Water Manag.* **2020**, *231*, 106026. [CrossRef]

35. Brocca, L.; Hasenauer, S.; Lacava, T.; Moramarco, T.; Wagner, W.; Dorigo, W.; Matgen, P.; Fernández, J.M.; Llorens, P.; Latron, J.; et al. Soil moisture estimation through ASCAT and AMSR-E sensors: An intercomparison and validation study across Europe. *Remote Sens. Environ.* **2011**, *115*, 3390–3408. [CrossRef]

36. Wagner, W.; Lemoine, G.; Rott, H. A Method for Estimating Soil Moisture from ERS Scatterometer and Soil Data. *Remote Sens. Environ.* **1999**, *70*, 191–207. [CrossRef]

37. Albergel, C.; Rüdiger, C.; Pellarin, T.; Calvet, J.C.; Fritz, N.; Froissard, F.; Suquia, D.; Petitta, A.; 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]

38. Pablos, M.; Gonzalez-Zamora, Á.; Sánchez, N.; Martínez-Fernández, J. Assessment of Root Zone Soil Moisture Estimates from SMAP, SMOS and MODIS Observations. *Remote Sens.* **2018**, *10*, 981. [CrossRef]
39. Stefan, V.-G.; Indrio, G.; Escorihuela, M.-J.; Quintana-Seguí, P.; Villar, J.M. High-Resolution SMAP-Derived Root-Zone Soil Moisture Using an Exponential Filter Model Calibrated per Land Cover Type. Remote Sens. 2021, 13, 1112. [CrossRef]

40. Ceballos, A.; Scipal, K.; Wagner, W.; Martínez-Fernández, J. Validation of ERS scatterometer-derived soil moisture data in the central part of the Duero Basin, Spain. Hydrolog. Process. 2005, 19, 1549–1566. [CrossRef]

41. Li, J.; Cui, J.; Sui, P.; Yue, S.; Yang, J.; Lv, Z.; Wang, D.; Chen, X.; Sun, B.; Ran, M.; et al. Valuing the synergy in the water-energy-food nexus for cropping systems: A case in the North China Plain. Ecol. Indic. 2021, 127, 107741. [CrossRef]

42. Mo, K.C.; Lettenmaier, D.P. Objective drought classification using multiple land surface models. J. Hydrometeorol. 2013, 15, 990–1010. [CrossRef]

43. Yang, Y.; Zhang, J.; Bao, Z.; Ao, T.; Wang, G.; Wu, H.; Wang, J. Evaluation of Multi-Source Soil Moisture Datasets over Central and Eastern Agricultural Area of China Using In Situ Monitoring Network. Remote Sens. 2021, 13, 1175. [CrossRef]

44. O’Neill, P.E.; Chan, S.K.; Njoku, E.G.; Jackson, T.; Bindlish, R.; Chaubell, J. SMAP Enhanced L3 Radiometer Global Daily 9 km EASE-Grid Soil Moisture, Version 4 [Indicate Subset Used]; NASA National Snow and Ice Data Center Distributed Active Archive Center: Boulder, CO, USA, 2020. [CrossRef]

45. Reichle, R.; Crow, W.T.; Koster, R.T.; Kimball, J.; De Lannoy, G. Algorithm Theoretical Basis Document (ATBD) SMAP Level 4 Surface and Root Zone Soil Moisture (L4_SM) Data Product. Available online: https://nsidc.org/sites/nsidc.org/files/files/data/smap/pdfs/l4_sm_initrel_v1_9.pdf (accessed on 25 May 2020).

46. Alt, H. The Computational Geometry of Comparing Shapes. In Conference on Efficient Algorithms; Institut fuer Informatik, Freie Universitaet: Berlin, Germany, 2009.

47. Xie, D.; Li, F.; Phillips, J.M. Distributed trajectory similarity search. Proc. VLDB Endow. 2017, 10, 1478–1489. [CrossRef]

48. Lange, R.D.; Beck, R.; Van, D.; Friesen, J.; Wit, A.D.; Wagner, W. Scatterometer-derived soil moisture calibrated for soil texture with a one-dimensional water-flow model. IEEE Trans. Geosci. Electron. 2008, 46, 4041–4049. [CrossRef]

49. Liu, S.; Xing, B.; Yuan, G.; Mo, X.; Lin, Z. Relationship analysis between soil moisture in root zone and top-most layer in China. Chin. J. Plant Ecol. 2013, 37, 1–17, (In Chinese with an English Abstract). [CrossRef]

50. Albergel, C.; Calvet, J.C.; Mahfouf, J.F.; Rudiger, C.; Barbu, A.L.; Lafont, S.; Roujean, J.L.; Walker, J.P.; Crapeau, M.; Wigneron, J.P. Monitoring of water and carbon fluxes using a land data assimilation system: A case study for southwestern France. Hydrol. Earth Syst. Sci. 2010, 14, 1109–1124. [CrossRef]

51. Barbu, A.L.; Calvet, J.C.; Mahfouf, J.F.; Albergel, C.; Lafont, T.S. Assimilation of Soil Wetness Index and Leaf Area Index into the ISBA-A-gs land surface model: Grassland case study. Biogeosciences 2011, 8, 1971–1986. [CrossRef]

52. Entekhabi, D.; Yueh, S.; O’Neill, P.; Kellogg, K.; Allen, A.; Bindlish, R.; Crow, W.T. SMAP Handbook–Soil Moisture Active Passive: Mapping Soil Moisture and Freeze/Thaw from Space; Jet Propulsion Laboratory (JPL) Publication: Pasadena, CA, USA, 2014; 180p.

53. Anderson, G.; Bell, R. Wheat grain-yield response to lime application: Relationships with soil pH and aluminium in Western Australia. Crop Pasture Sci. 2019, 70, 295–305. [CrossRef]

54. Markus, R.; Gustau, C.-V.; Bjorn, S.; Martin, J.; Joachim, D.; Prabhat, C.N. Deep learning and process understanding for data-driven Earth system science. Nature 2019, 566, 195–204.