



Article

Random Forest-Based Reconstruction and Application of the GRACE Terrestrial Water Storage Estimates for the Lancang-Mekong River Basin

Senlin Tang^{1,2}, Hong Wang^{1,*}, Yao Feng¹, Qinghua Liu³, Tingting Wang¹, Wenbin Liu¹ and Fubao Sun^{1,2,4,5}

¹ Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; tangsl.19b@igsrr.ac.cn (S.T.); fengy@igsrr.ac.cn (Y.F.); wangtt@igsrr.ac.cn (T.W.); liuw@igsrr.ac.cn (W.L.); sunfb@igsrr.ac.cn (F.S.)

² College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100190, China

³ Institute of Remote Sensing and Geographic Information System, Peking University, Beijing 100871, China; liuqh_insar@pku.edu.cn

⁴ State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China

⁵ Akesu National Station of Observation and Research for Oasis Agro-Ecosystem, Akesu 843300, China

* Correspondence: wanghong@igsrr.ac.cn



Citation: Tang, S.; Wang, H.; Feng, Y.; Liu, Q.; Wang, T.; Liu, W.; Sun, F. Random Forest-Based Reconstruction and Application of the GRACE Terrestrial Water Storage Estimates for the Lancang-Mekong River Basin. *Remote Sens.* **2021**, *13*, 4831. <https://doi.org/10.3390/rs13234831>

Academic Editors: Yang Hong, Aondover Tarhule, Emad Hasan and Guoqiang Tang

Received: 12 October 2021

Accepted: 26 November 2021

Published: 28 November 2021

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



Copyright: © 2021 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/>).

Abstract: Terrestrial water storage (TWS) is a critical variable in the global hydrological cycle. The TWS estimates derived from the Gravity Recovery and Climate Experiment (GRACE) allow us to better understand water exchanges between the atmosphere, land surface, sea, and glaciers. However, missing historical (pre-2002) GRACE data limit their further application. In this study, we developed a random forest (RF) model to reconstruct the monthly terrestrial water storage anomaly (TWSA) time series using Global Land Data Assimilation System (GLDAS) and Climatic Research Unit (CRU) data for the Lancang-Mekong River basin. The results show that the RF-built TWSA time series agrees well with the GRACE TWSA time series for 2003–2014, showing that correlation coefficients (R) of 0.97 and 0.90 at the basin and grid scales, respectively, which demonstrates the reliability of the RF model. Furthermore, this method is used to reconstruct the historical TWSA time series for 1980–2002. Moreover, the discharge can be obtained by subtracting the evapotranspiration (ET) and RF-built terrestrial water storage change (TWSC) from the precipitation. The comparison between the discharge calculated from the water balance method and the observed discharge showed significant consistency, with a correlation coefficient of 0.89 for 2003–2014 but a slightly lower correlation coefficient (0.86) for 1980–2002. The methods and findings in this study can provide an effective means of reconstructing the TWSA and discharge time series in basins with sparse hydrological data.

Keywords: Lancang-Mekong River basin; random forest; terrestrial water storage; water balance; discharge

1. Introduction

The Lancang-Mekong River basin is the most important transnational water system in Asia, flowing through China, Laos, Myanmar, Thailand, Cambodia, and Vietnam and eventually into the South China Sea. The study of terrestrial water storage (TWS) is thus critical for shipping, hydro-energy, irrigation, and water protection in countries along the river [1–5]. It provides a new perspective for tracking global water resources and has been widely applied in monitoring drought, flood potential, and groundwater changes [6–8].

Before GRACE, few TWS data could be used to understand global water resources [9–12]. The Global Land Data Assimilation System (GLDAS) land surface models (LSMs) and global hydrological models (GHMs) can simulate long-term TWS [10,13]. The TWS is composed of the following components: snow water equivalent (SWE), canopy water storage (CWS), surface water storage (SWS), soil moisture storage (SMS), and groundwater

storage (GWS). However, LSMs and GHMs are unusually not calibrated against GRACE measurements, which may lead to large deviations and a failure to adequately depict the TWS [14–17]. Nevertheless, these models contain complicated physical processes and have more parameters and greater computational complexity [17]. Therefore, there is an urgent need to develop an efficient data-driven algorithm for TWS reconstruction.

With the advent of remote sensing, machine learning and data mining techniques have emerged in meteorology and hydrology [18]. Ndehedehe and Ferreira [19] predicted the TWS using a partial least squares regression (PLSR) model and suggested that the TWS in some river basins is weakly associated with climate. Huang, et al. [20] quantified the effects of climate change and human activities on vegetation dynamics using a support vector machine (SVM) model, providing some guidance for ecological restoration on the Loess Plateau. Nguyen, et al. [21] proposed an artificial neural network to reconstruct long-term and high-resolution precipitation, which has been widely used in the geosciences. Li, et al. [22] combined climate variables and catchment attributes to estimate annual discharge using a random forest (RF) model. RF is an ensemble method combining several weak learners to produce a strong learner that yields the optimum results. Additionally, each variable can be sorted by importance, making this model more explanatory [23–25]. Combined with LSMs and meteorological forcing data, RF provides a comprehensive perspective on TWSA reconstruction.

Effective discharge monitoring is important for water management and utilization, as well as the scientific development of dispatch plans [26–28]. However, due to the variation in regional precipitation frequency and intensity, the estimation of discharge remains challenging. Encouragingly, the water balance method describes the equilibrium of water revenue and expenditure, which provides an effective way to estimate discharge [29–31]. The terrestrial water storage change (TWSC) is negligible in the water balance at the annual level but is essential at the sub-annual level [32–34]. The water balance method allows us to better understand the water cycle, and discharge can be obtained by subtracting the evapotranspiration (ET) and the TWSC from precipitation [35,36]. Thus, discharge in the Lancang-Mekong River basin can be determined using this method with multi-source remote sensing products [37].

In this study, we built monthly TWSA time series using an RF model in the Lancang-Mekong River basin for 2003–2014 and reconstructed the TWSA from 1980 to 2002. Moreover, discharge can be effectively estimated using the water balance method incorporating the RF-built TWSA data, which justifies the reliability and applicability of the RF model. The rest of this article is organized as follows. Sections 2 and 3 describe the study area, data resources, and methods. The reconstruction of the TWSA time series and estimated discharge is presented in Section 4. The uncertainty in multi-source products is discussed in Section 5, followed by conclusions in Section 6.

2. Study Area and Data Resources

2.1. Study Area

The Lancang-Mekong River is the twelfth longest river in the world and the seventh longest river in Asia; it is well known as the most important transnational water system in Asia. It originates in Yushu Tibetan Autonomous Prefecture, Qinghai Province, China. The mainstream spans 4350 km and covers an area of over 795,000 km². The Lancang-Mekong River basin consists of two parts: the Lancang River in China and the Mekong River in mainland Southeast Asia. The Lancang River Basin covers 164,800 km², accounting for 21% of the area, with an average annual flow rate of 2140 m³/s and an average annual outbound water volume of 76.5 billion m³ [1,38,39]. The Mekong River Basin is dominated by croplands, and the Southern Vietnamese region is one of the world's most famous crop growing areas, known as the "Rice Bowl of Vietnam" [40] (Figure 1b).

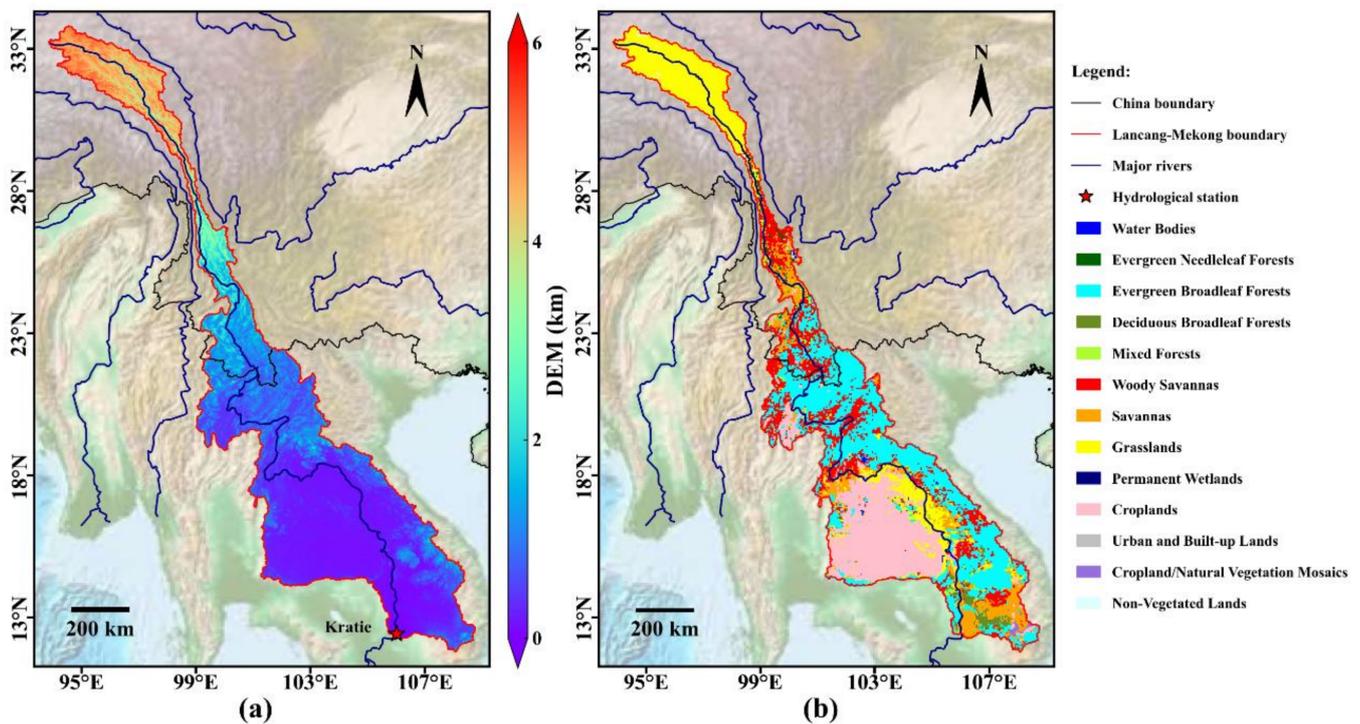


Figure 1. The Global Multi-Resolution Terrain Elevation Data (a) and the Terra and Aqua combined Moderate Resolution Imaging Spectroradiometer (MODIS) Land Cover Type (b) in the Lancang-Mekong River basin.

The Kratie hydrological station is located on the mainstream in the lower Mekong River, which controls the entire red border range in Figure 1. The elevation of the basin declines from northwest to southeast, with an average over 4000 m in China, and the terrain is relatively flat downstream [41] (Figure 1a). Located in the center of the tropical monsoon regions in Asia, the Lancang-Mekong River has great differences in water flow during dry and flooding periods [1]. The Indian summer monsoon brings rich moisture from September to October, with a peak flow of 757,000 m³/s. It is relatively dry from January to February, with a minimum flow rate of 1250 m³/s [39]. Precipitation and snowmelt are the primary sources of water, and the Lancang-Mekong River basin is known to be largely a rain-fed rather than snow-fed river [42]. In the Lancang River Basin, snowmelt mainly comes from the Tibetan plateau, while precipitation is the main discharge source of the Mekong River [39,43].

2.2. Data Sources

2.2.1. Terrestrial Water Storage (TWS)

The Gravity Recovery and Climate Experiment (GRACE) is a collaboration between the National Aeronautics and Space Administration (NASA) and the German Aerospace Center (DLR). It uses a pair of twin satellites to monitor changes in the Earth's gravitational field and study the Earth's water resource geology and climate [44,45]. Due to the rough spatial resolution, it is especially suitable for a study area larger than 200,000 km² [36,46].

In this study, we chose the 0.25° × 0.25° GRACE RL06 product (Table 1) from the Center for Space Research (CSR-RL06, http://www2.csr.utexas.edu/grace/RL06_mascons.html (accessed on 27 November 2021)) to analyze the TWS. The mass concentration blocks have superior performance compared with the standard spherical harmonic approach, which can significantly increase the amplitude and spatial localization of the recovered TWSA data [47]. Moreover, land and ocean signals can be better separated, and the data require no additional striping or smoothing. To ensure a consistent spatial resolution, the data were

resampled to $0.5^\circ \times 0.5^\circ$. Data of the missing months were filled in by interpolation [40]. The TWSA and TWSC are considered Equations (1) and (2).

$$TWSA_t = TWS_t - \overline{TWS}_{2004-2009} \quad (1)$$

$$TWSC_t = \frac{(TWSA_{t+1} - TWSA_{t-1})}{2} \quad (2)$$

where t denotes the month, the TWSA data are calculated by subtracting the average TWS from 2004–2009 from the monthly TWS, and the TWSC is defined as the mean of the variation in the TWS for every other month [48].

Table 1. Details of the multi-source products used in this study.

Product Name	Datasets	Spatial Resolution	Temporal Resolution	Temporal Coverage	Reference
GRACE CSR RL06 Macon	TWS	0.25°	Monthly	2003–2014	[47]
GLDAS NOAH (v2.0)	SWE, SMS, CWS	0.5°	Monthly	1980–2014	[27]
CRU TS v4.03	T, DTR, P, VAP, WET, CLD, FRS, TMN, TMX, PET	0.5°	Monthly	1980–2014	[49]
Discharge	Q	-	Daily	1980–2014	[4]
MSWEP	Precipitation	0.1°	Monthly	1980–2014	[50]
GPCC		0.5°	Monthly	1980–2014	[51]
PERSIANN-CDR		0.25°	Monthly	2003–2014	[21]
TRRM		0.25°	Monthly	2003–2014	[52]
CRU		0.5°	Monthly	1980–2014	[49]
GLEAM v3.3a/b	ET	0.25°	Monthly	1980–2014, 2003–2014	[53]
PML-V2		0.05°	8 days	2003–2014	[54]
GLDAS NOAH/VIC/CLSM (v2.0)		$1^\circ/0.5^\circ$	Monthly	1980–2014	[35]

2.2.2. Global Land Data Assimilation System

NASA and the National Oceanic and Atmospheric Administration (NOAA) co-developed the Global Land Data Assimilation System (GLDAS) [27]. The GLDAS Version 2.0 Noah product was forced by the Princeton meteorological datasets (<https://disc.gsfc.nasa.gov/> (accessed on 27 November 2021)) [55], and the resolution was resampled from $1^\circ \times 1^\circ$ to $0.5^\circ \times 0.5^\circ$ for GLDAS VIC and CLSM ET products. In this study, the Noah TWS is used for comparison with the GRACE TWS. Moreover, the Noah TWS (including the soil moisture storage (SMS, 0–10, 10–40, 40–100, 100–200 cm), snow water equivalent (SWE), and canopy water storage (CWS)) is used as the input for the RF model (Table 1) [56].

2.2.3. Meteorological Data

Monthly meteorological data with $0.5^\circ \times 0.5^\circ$ spatial resolution were collected from the Climatic Research Unit gridded Time Series Version 4 (CRU) [49]. These data are interpolated from global meteorological observation datasets and updated annually. The CRU consists of 10 components: the mean 2 m temperature (TMP), diurnal 2 m temperature range (DTR), precipitation rate (P), vapor pressure (VAP), wet days (WET), cloud cover (CLD), frost days (FRS), minimum 2 m temperature (TMN), maximum 2 m temperature (TMX), potential evapotranspiration (PET). Meteorological elements disturbing the TWS are used as the forcing data in the RF model (Table 1) [57].

To constrain the error caused by uncertainty in the precipitation data, we obtained five gridded precipitation products: the Multi-Source Weighted-Ensemble Precipitation (MSWEP) [50], the Global Precipitation Climatology Centre V2018 (GPCC) [51], the Precip-

itation Estimation from Remotely Sensed Information using Artificial Neural Networks Climate Data Record (PERSIANN-CDR) [21], the Tropical Rainfall Measurement Mission (TRMM) [52], and the CRU precipitation products. Moreover, we obtained six ET products from the Global Land Evaporation Amsterdam Model v3.3a/b (GLEAM) [53], the Penman-Monteith-Leuning Version 2 (PML-V2) [54,58], and the GLDAS ET products [35]. All the precipitation and ET products were interpolated to $0.5^\circ \times 0.5^\circ$ (Table 1).

2.2.4. Discharge Data

Daily discharge observations from 1980–2014 were collected from the Kratie hydrological station in Cambodia (Table 1), which is provided by the Mekong River Commission (MRC) [3,4]. To ensure consistency with each component in the water balance equation, daily discharge data are aggregated into monthly discharge data.

3. Methods

3.1. Reconstruction of the TWSA Time Series with the RF Model

The RF model, proposed by Breiman [25], is an end-to-end machine learning algorithm containing multiple decision trees; it has been widely used in geoscience [22]. The core idea of the RF model is ensemble learning, and the basic unit is a decision tree. Ensemble learning uses multiple learners and integrates learning methods through specific rules to achieve better results than a single learner. The bootstrap aggregating (bagging) method is essential for assembling weak regressors into strong regressors [23]. Random samples are selected with replacement, and each sample is trained to build the model. The average of these multiple models is considered the final result. One advantage of the RF model is the important function of evaluating input characteristics [59]. Each decision tree has one-third of the out-of-bag (OOB) datasets, which can be used for performance evaluation, and the OOB error is calculated. By adding noise to the OOB datasets, a new OOB error can be recalculated. Comparing the error of the datasets before and after adding noise, the variable's importance is ranked. The RF model is given by Equation (3), where the TWSA time series is a function of its associated variables, including the SWE, SMS, and CWS from GLDAS making up the TWSA and T, DTR, P, VAP, WET, CLD, FRS, TMN, TMX, PET from CRU affecting the TWSA.

$$\text{TWSA} = f(\text{SWE}, \text{SMS}, \text{CWS}, \text{T}, \text{DTR}, \text{P}, \text{VAP}, \text{WET}, \text{CLD}, \text{FRS}, \text{TMN}, \text{TMX}, \text{PET}) \quad (3)$$

We reconstruct the TWSA from the basin and grid scales following the flowchart in Figure 2. The technical details are briefly introduced below:

- (1) The GLDAS and CRU datasets are organized into an $N \times M$ matrix, with N and M as the number of variables and months. The prediction target is the TWSA.
- (2) Applying the bootstrap method, a new set of k samples is randomly selected, and a regression tree k is established. Each time, the data that are not pumped are called an OOB.
- (3) The j th variable $X(j)$ and its values are selected as the segmentation variables, and the cut-off point, which defines the following two regions, is selected.

$$R_1(j, s) = \{x | x_j \leq s\}, R_2(j, s) = \{x | x_j > s\} \quad (4)$$

where the optimal parameters are determined by Equation (5).

$$\min_{j,s} \left[\min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2 \right] \quad (5)$$

We can determine the optimal cut-off points from the input variable j as in Equation (6).

$$\hat{c}_1 = \text{ave}(y_i | x_i \in R_1(j, s)), \hat{c}_2 = \text{ave}(y_i | x_i \in R_2(j, s)) \quad (6)$$

The above division is repeated for the two regions until the stop condition is met, and the least-two-times regression tree is obtained.

(4) The final step is to retrieve the ensemble average of all the individual regression trees as the reconstructed TWSA.

To obtain a reliable and stable model, leave-one-year-out cross-validation is adopted [22]. All the forcing data from 2003 to 2014 except for one year are used for RF modeling, and the excluded year is used for the model’s predictions. Rotation estimation of the TWSA year by year achieves more convincing results (Figure 2c). Moreover, the most important variables are gradually added to optimize the RF model by eliminating redundant variables through the normalized root mean square error. Finally, the RF model is extended to the historical period from 1980 to 2002.

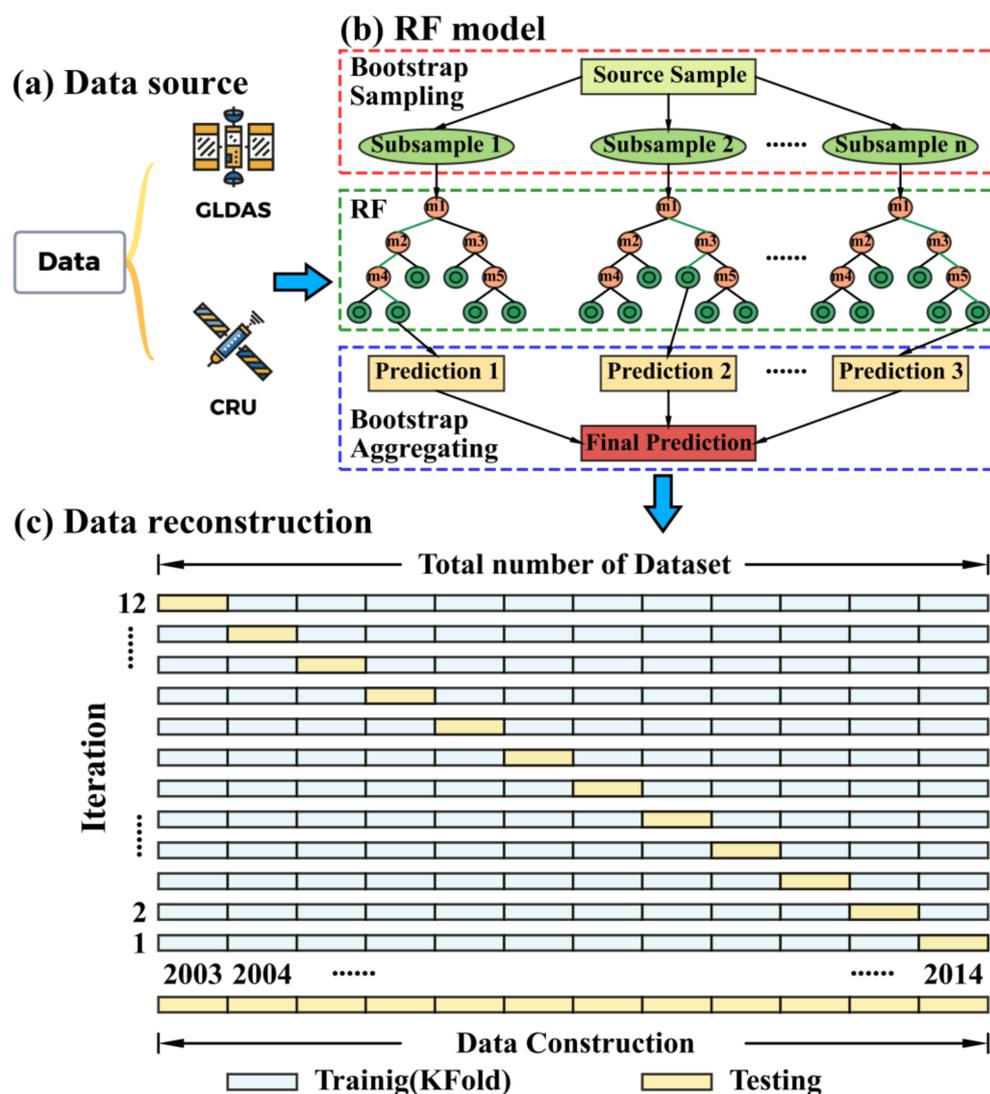


Figure 2. The reconstruction of the TWSA data for the RF model. (a) Forcing datasets; (b) Schematic of random forest algorithm; (c) Reconstruct TWSA using K-fold cross-validation.

3.2. Estimation Based on the Water Balance Method

The water balance equation (Equation (7)) was employed to calculate the monthly discharge in the Lancang-Mekong River basin for 1980–2014 [26,29,30,35–37].

$$Q = P - ET - TWSA \tag{7}$$

where ET , P , $TWSC$, and Q are the monthly evapotranspiration (ET, mm), precipitation (mm), terrestrial water storage change (mm), and discharge (mm), respectively. In this study, P was calculated from the MSWEP, GPCC, PERSIANN-CDR, TRMM, and CRU precipitation products, the ET was calculated using the GLEAM v3.3a/b, PML-V2, and GLDAS NOAH/VIC/CLSM data, and the $TWSC$ was calculated using the GRACE TWS dataset. The observed discharge from the Kratie hydrological station is used for model validation. Limited by the data availability, only three precipitation products (MSWEP, GPCC, CRU) and four ET products (GLEAM v3.3a, GLDAS NOAH/VIC/CLSM) were used in the water balance method for 1980–2002.

3.3. Uncertainty Analysis for Discharge Estimation

According to Equation (7) and the theory of uncertainty propagation, the uncertainty mainly comes from the uncertainty in the precipitation, ET , and $TWSC$ [26,37,57].

$$U_R = \sqrt{U_P^2 + U_{ET}^2 + U_{TWSC}^2} \quad (8)$$

where U_R , U_P , U_{ET} , and U_{TWSC} represent the uncertainty in the discharge, precipitation, ET , and $TWSC$, respectively. Uncertainty is inevitable in multi-source datasets, and the uncertainties in the precipitation and ET are qualitatively estimated using a 95% confidence interval. Measurement and leakage errors exist in the GRACE TWS datasets [10]. Leak errors were assumed to be negligible, and measurement errors can be approximately calculated by the root mean square (RMS) of the TWS residuals, which were acquired from seasonal trend decomposition using the loess method (STL) [57].

$$TWS_t = T_t + S_t + R_t \quad (9)$$

where T_t , S_t , and R_t represent the trend, seasonal, and residual parts of the TWS, respectively. It is important to note that the residuals decomposed by the STL algorithm contain sub-seasonal signals and noise.

3.4. Evaluation Metrics

The RF-built TWSA parameter for 2003–2014 is evaluated using the GRACE TWSA, and the simulated discharge from 1980 to 2014 is evaluated using the observed discharge from the Kratie hydrological station. In this study, a Taylor diagram is a useful tool for qualitative analysis. Moreover, the correlation coefficient (R), Nash–Sutcliffe model efficiency coefficient (NSE), and normalized root mean square error ($NRMSE$) are used for model evaluation. R is used to measure the correlation between the predicted and observed datasets. The NSE is calculated as the magnitude of the simulated dataset's error variance, which is a normalized statistic. The $NRMSE$ is mainly used to evaluate the amplitude of errors. These indicators are calculated as follows.

$$R = \frac{\sum_{i=1}^m (\hat{x}_i - \bar{\hat{x}}_i)(x_i - \bar{x}_i)}{\sqrt{\sum_{i=1}^m (\hat{x}_i - \bar{\hat{x}}_i)^2 \sum_{i=1}^m (x_i - \bar{x}_i)^2}} \quad (10)$$

$$NSE = 1 - \frac{\sum_{i=1}^m (\hat{x}_i - x_i)^2}{\sum_{i=1}^m (x_i - \bar{x}_i)^2} \quad (11)$$

$$NRMSE = \sqrt{\frac{\sum_{i=1}^m (\hat{x}_i - x_i)^2}{\sum_{i=1}^m x_i^2}} \quad (12)$$

where m represents the total number of months in the datasets, x_i and \hat{x}_i represent the observed and predicted datasets, respectively, and \bar{x}_i and $\bar{\hat{x}}_i$ represent the mean values of the observed and predicted datasets.

4. Results

4.1. Evaluation of the RF-Built TWSA Time Series

4.1.1. Evaluation at the Grid Scale

From Equation (3) and Figure 2, we reconstruct the TWSA time series at the grid scale for 2003–2014. Soil moisture storage is the most important variable for reconstructing the TWSA time series (Figure 3a). The SMS, TMX, CLD, T, PET, CWS, and P can be an optimal combination of the RF model (NRMSE = 0.49). Figure 3b presents a boxplot of the R, NSE, and NRMSE between the GLDAS and RF models in the Lancang-Mekong River basin. Comparing the minimum, median, third quartile, and maximum values of the boxplots, the RF model is superior to the GLDAS model with higher R and NSE, and NRMSE values indicating minor errors. Figure 4a,b intuitively shows the distribution of the scatter plot at the grid scale, and the fitted line of the RF model is closer to 1:1 than that of the GLDAS model. Moreover, the RF model is also superior to the GLDAS model in the Taylor diagram (Figure 4c), with R, NSE, and NRMSE values of 0.83, 0.66, and 0.58 for the GLDAS model, which are improved by the RF model to 0.90, 0.80, and 0.44, respectively.

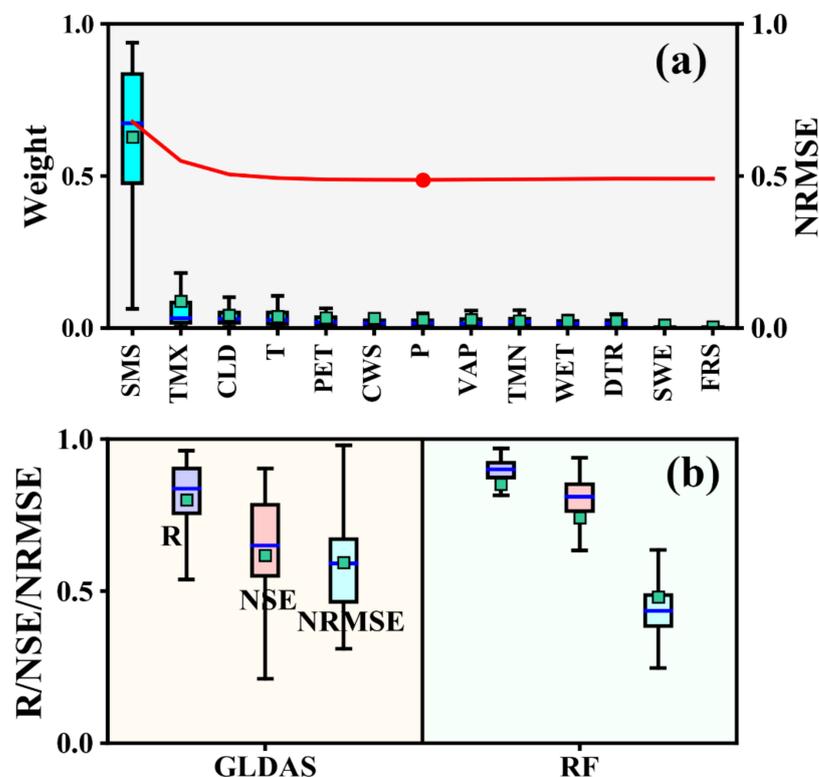


Figure 3. Grid-scale analysis in the Lancang-Mekong River basin for 2003–2014. (a) Relative weight of the variables in the RF models; the red line represents the model error. (b) Comparison between the GLDAS and RF models; the left (right) panel shows the boxplots of R, NSE, and NRMSE calculated from the GRACE TWSA and GLDAS TWSA time series (RF-built TWSA time series).

Figure 5 presents the magnitudes and spatial patterns of the R, NSE, and NRMSE of the GLDAS and RF models for 2003–2014. The GLDAS and RF models exhibit similar patterns, but the RF model is significantly superior to the GLDAS model. It is worth noting that the simulation performance of the TWSA time series in the northern Qinghai–Tibet Plateau region is poor. Compared with the GLDAS model, R increased from 0.80 to 0.90 for the RF model in the southern croplands. Figure 6 shows the cumulative distribution functions (CDFs) of the R, NSE, and NRMSE and those of the RF model show a clear improvement over those of the GLDAS model. Moreover, 56.7% of the grid cells show an NSE larger than 0.8 in the RF model, while the percentage is 21.3% in the GLDAS model.

Overall, the RF model shows remarkable performance at the grid scale. Furthermore, the optimized model is extended to 1980–2002.

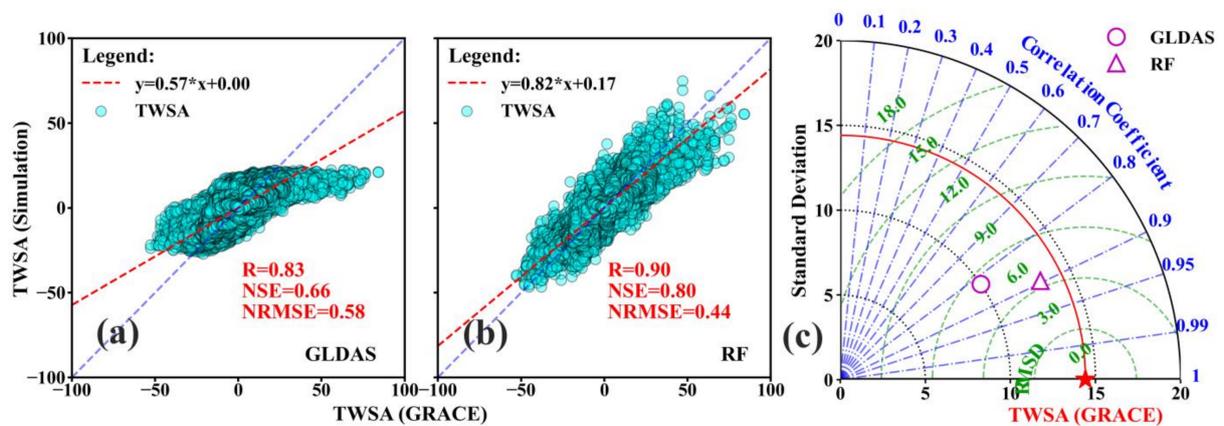


Figure 4. Performance comparison of the GRACE TWSA and simulated TWSA (GLDAS and RF-built TWSA) at the grid scale for 2003–2014. The 1:1 line is blue, and the fitted line is red. (a) Scatter plot of the GLDAS TWSA and GRACE TWSA time series; (b) Scatter plot of the RF-built TWSA and GRACE TWSA series; (c) Taylor diagram for comparing the GRACE TWSA data and the simulated TWSA data.

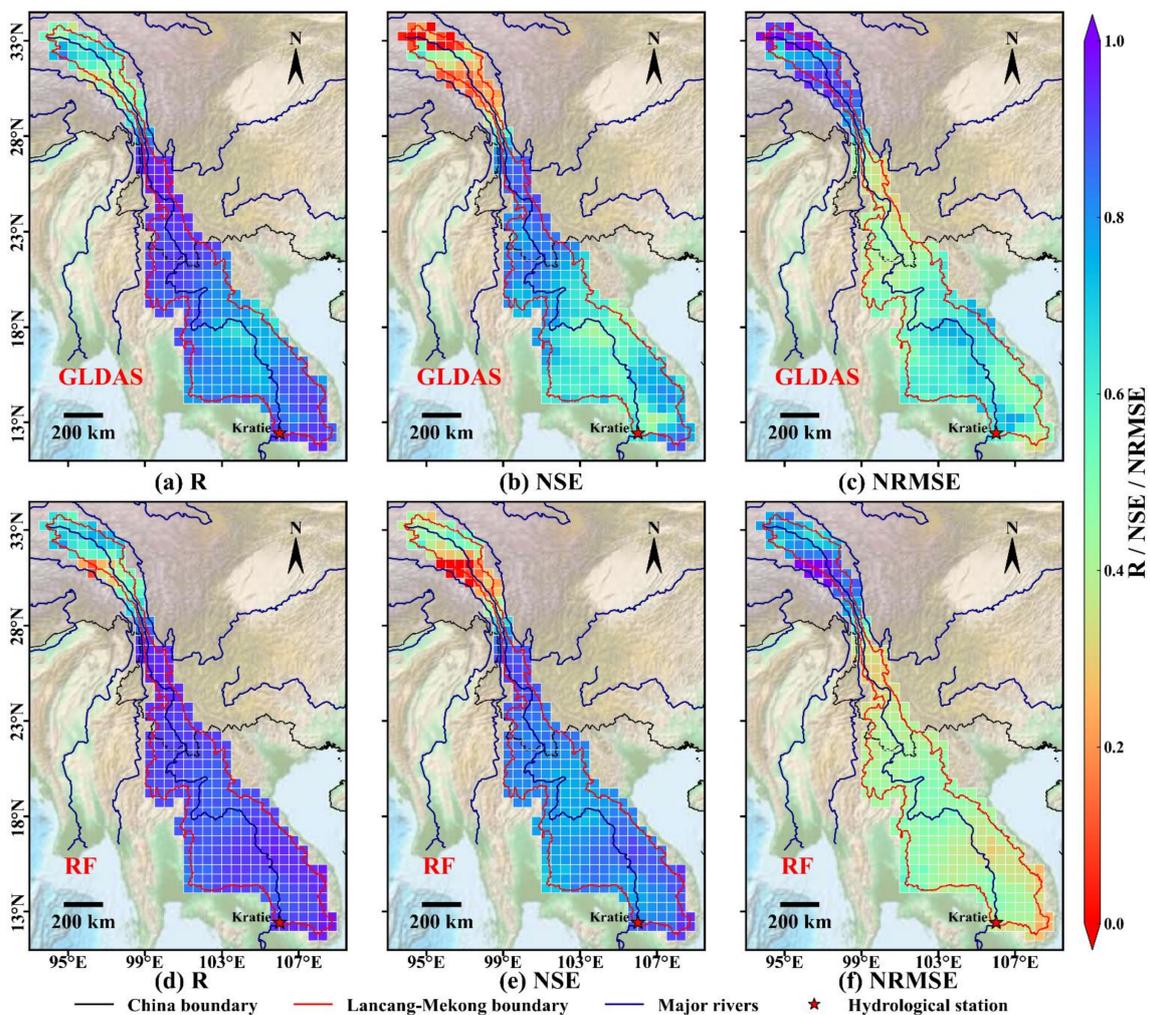


Figure 5. Spatial distribution of the R, NSE, and NRMSE values calculated from the GRACE TWSA and simulated TWSA (GLDAS and RF-built TWSA) for 2003–2014. (a–c) represent the analysis results of the GLDAS model, and (d–f) represent the RF model.

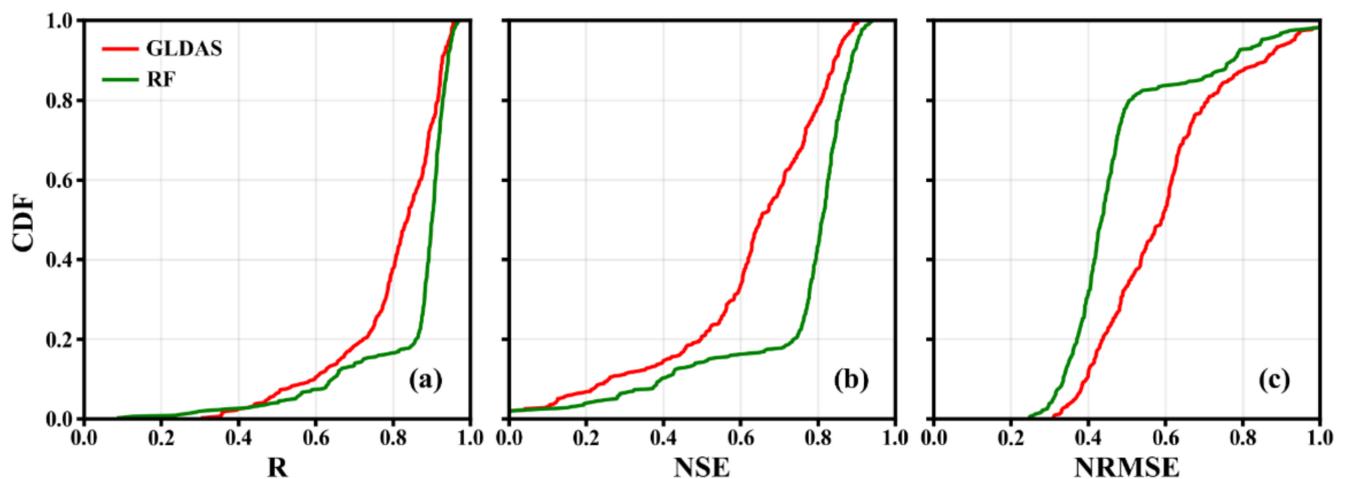


Figure 6. (a–c) represent the Cumulative distribution functions (CDFs) of the R, NSE, and NRMSE values derived from the Figure 5. Red indicates the GLDAS model, and green indicates the RF model.

4.1.2. Evaluation at the Basin Scale

The basin-scale analysis provides a more comprehensive evaluation of the RF-built TWSA time series. The reliability of the RF model has been verified at the grid scale, and the grid-averaged TWSA time series can be used in Lancang-Mekong River basin analysis. Figure 7 shows a comparison between the GRACE and the RF-built TWSA, as well as the reconstructed historical TWSA. The Kfold cross-validation shows that they are well adapted for every year during 2002–2014. The R between GRACE and the RF-built TWSA ranges from 0.96 to 0.99, the NSE ranges from 0.82 to 0.98, and the NRMSE ranges from 0.15 to 0.41. The RF-built TWSA for each year is credible. Overall, the monthly TWSA time series can be accurately depicted by the RF model, and the seasonal fluctuations are consistent with those from the GRACE TWSA time series. Encouragingly, the historical TWSA time series in the Lancang-Mekong River basin was also reconstructed for 1980–2002 (Figure 7).

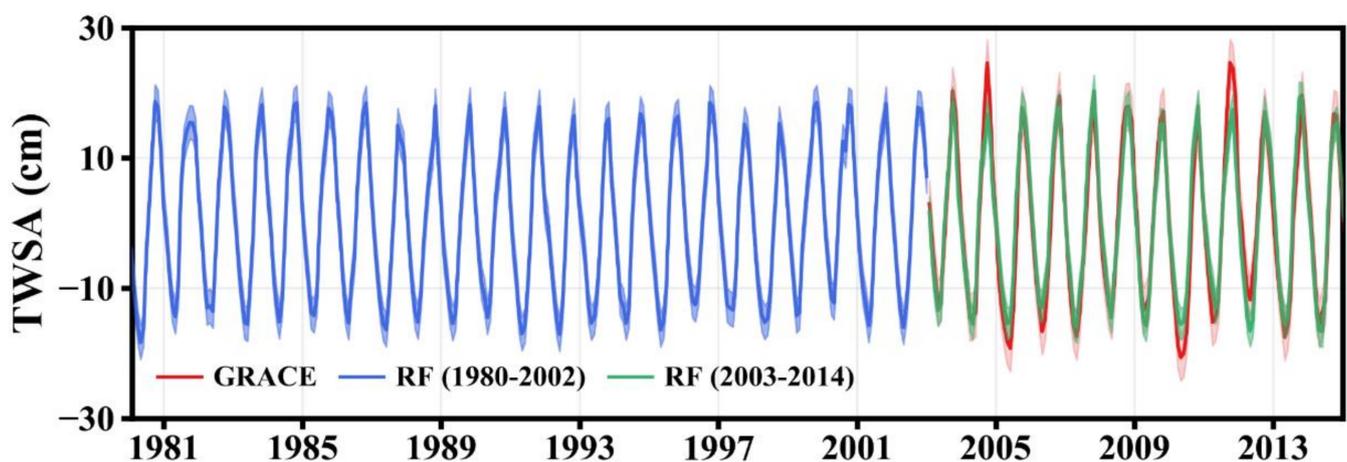


Figure 7. Comparisons of the monthly TWSA data in the Lancang-Mekong River basin for 1980–2014 (red: GRACE TWSA, green/blue: RF-built TWSA), and the shade represents the uncertainty in TWSA data.

Figure 8a,b suggest the superior performance of the RF model compared to the GLDAS model for 2003–2014. Figure 8c clearly shows the advantages and disadvantages of the RF and GLDAS models. The simulated TWSA that agrees well with the GRACE TWSA lies closest to the red star and red arc. In contrast, the simulated TWSA time series from the GLDAS model shifts away from the GRACE TWSA time series. Quantitative analysis at the basin scale shows better simulation results by the RF model ($R = 0.97$, $NSE = 0.93$,

NRMSE = 0.26) and less desirable results ($R = 0.93$, $NSE = 0.81$, $NRMSE = 0.44$) by the GLDAS model. Although the GLDAS model has good correlation, the scatter plot exhibits large deviations, while the RF-built TWSA time series is more convincing. Moreover, the performance of the TWSA time series is significantly improved at the basin scale compared to that at the grid scale.

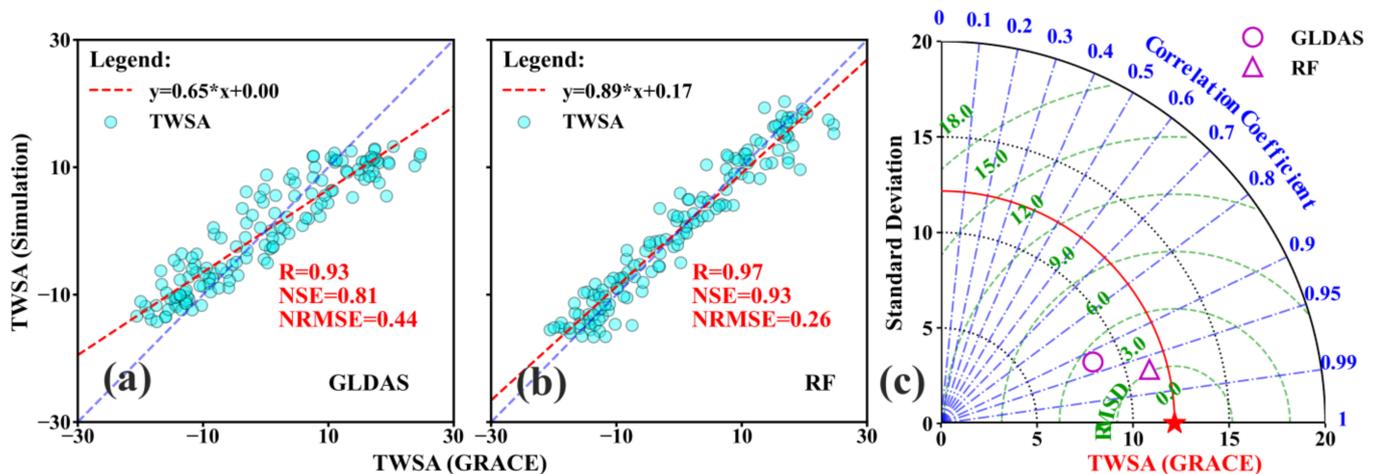


Figure 8. Performance comparison of the GRACE TWSA and simulated TWSA (GLDAS and RF-built TWSA) at the basin scale for 2003–2014. The 1:1 line is blue, and the fitted line is red. (a) Scatter plot of the GLDAS TWSA and GRACE TWSA time series; (b) Scatter plot of the RF-built TWSA and GRACE TWSA series; (c) Taylor diagram for comparing the GRACE TWSA data and the simulated TWSA data.

4.2. Estimated Discharge by the Water Balance Method

The discharge calculated by both the water balance method (Equation (7)) with the GRACE TWSC (WB-GRACE) and the RF-based TWSC (WB-RF) achieves good agreement with the field measurements (Figure 9a,b). The shades represent the uncertainty in the precipitation and ET, and the range of the monthly average 95% confidence interval errors is 14.64 mm. Figure 9c,d present the results of simulated discharge, and the scatter plots obtained by the WB-GRACE and WB-RF are closer to the 1:1 line. The Taylor diagram (Figure 9e) indicates that the discharge estimated by the WB-RF is superior to that estimated by the WB-GRACE for 2003–2014, and the estimated discharge achieves similar performance for 1980–2002.

Quantitative analysis shows that the discharge calculated from the WB-RF performed better for 2003–2014 ($R = 0.89$, $NSE = 0.78$, $NRMSE = 0.32$), followed by that calculated from the WB-GRACE ($R = 0.86$, $NSE = 0.73$, $NRMSE = 0.35$). The estimated discharge from the WB-RF outperformed the discharge (2003–2014) calculated by the WB-GRACE; thus, the reliability of the estimated discharge is demonstrated directly. The discharge calculated by the WB-RF also achieved desirable performance for 1980–2002 ($R = 0.84$, $NSE = 0.70$, $NRMSE = 0.37$), and the expandability of the estimated discharge was further demonstrated. The estimated discharge for 1980–2002 achieved lower performance than the discharge for 2003–2014, which was partially caused by the limited frequency of the multi-source datasets, as only part of the P and ET are used for 1980–2002. Moreover, the discharge calculated by the WB-RF in this research seems to have better metrics than the results from Xie, et al. [26], with an NSE of 0.66.

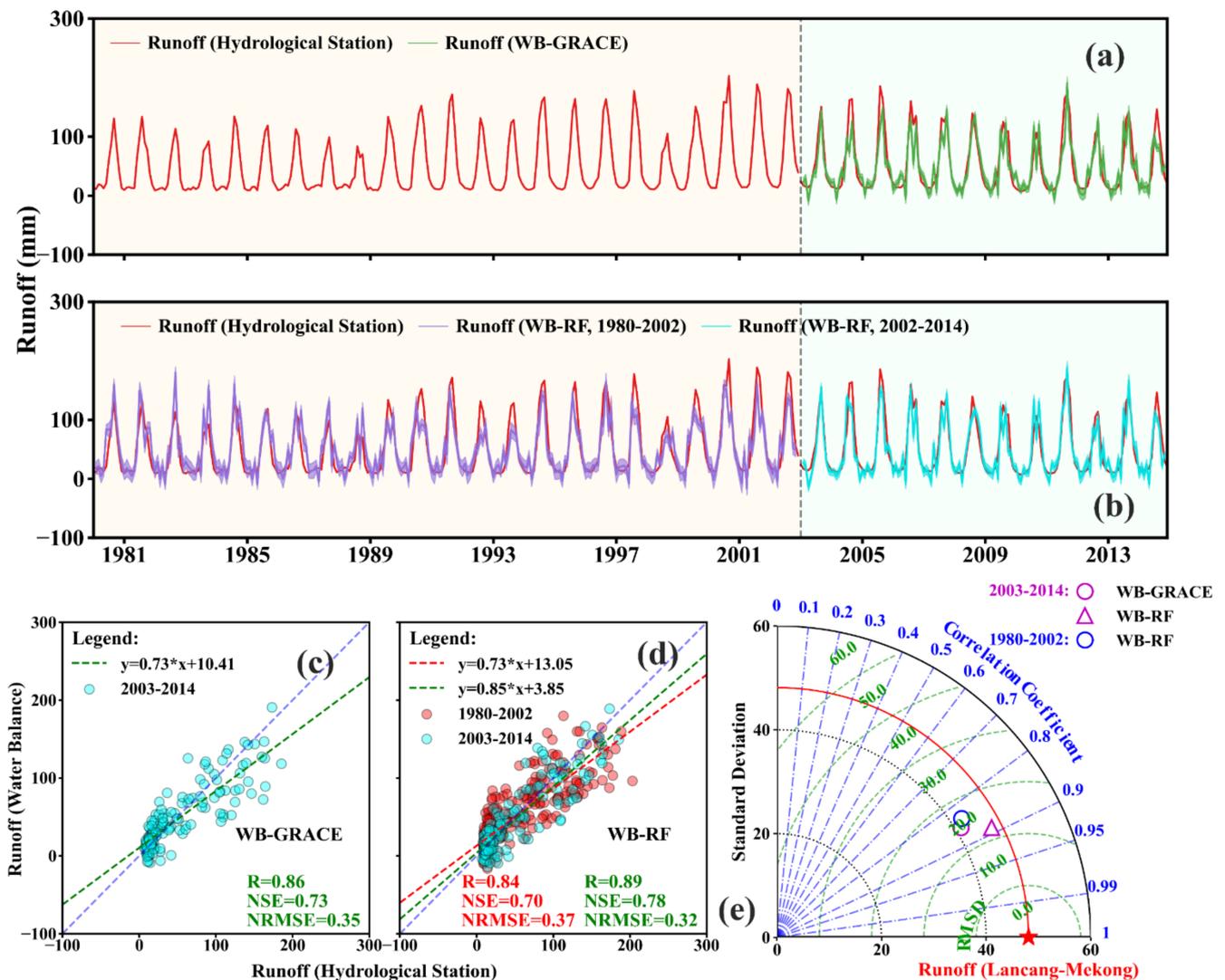


Figure 9. Comparison between the estimated discharge (WB-GRACE, WB-RF) and the observed discharge in the Lancang-Mekong River basin during 1980–2014. (a,b) refer to monthly estimated discharge time series using WB-GRACE and WB-RF, respectively, and the shades represent the uncertainty of discharge. (c,d) are the scatter plots of the estimated discharge and observed discharge. The 1:1 line is blue, and the fitted lines are red (1980–2002) and green (2003–2014). (e) Taylor diagram for comparing the estimated discharge and observed discharge.

5. Discussion

5.1. Reliability and Uncertainty in the RF Model

Although data-driven methods are less explanatory than physical methods, they can achieve higher performance and execution efficiency [17]. With the enhancement of the acquisition capacity of remote sensing data, the era of big data has arrived. Moreover, with the rapid interdisciplinary development, the application of machine learning and data mining in the field of meteorology and hydrology is increasing [60]. Compared with physical methods, data-driven methods can achieve faster calculation speed and higher precision [9]. Furthermore, more advanced machine learning algorithms, such as recursive neural networks, should be explored in the future.

Unfortunately, there are limited datasets available. Only the SWE, SMS, and CWS in the GLDAS model are adopted, and the CRU meteorological forcing data are also used to improve the performance of the model [9,57]. Figure 5 shows that the grid cells in the northern Qinghai-Tibet Plateau region, with lower performance in terms of an increasing glacier melt in the Lancang-Mekong River basin, are not taken into consideration [61]. The performance of machine learning methods is largely influenced by the model's input

variables. In the future, we will fully account for other explanatory variables affecting the TWSA, such as surface water storage and groundwater storage. The discharge estimated by the WB-RF in Figure 9 is well illustrated, with NSE exceeding 0.7, and the seasonal frequency is well captured. Compared with other parts, there is a significant discrepancy in peak discharge, which leads to a decrease in the prediction accuracy. It is suspected that the RF model does not contain sequence information, and effective integration of multi-year information could further improve the prediction accuracy.

5.2. Uncertainties in the Water Balance Equation

Figure 10a–k present the magnitudes and spatial patterns of precipitation and ET products. Precipitation and ET present similar spatial distributions, decreasing from the southeast to the northwest. Figure 10l,m present the differences among multi-source products, and the fluctuation in the multi-source ET is larger than that of the precipitation products. The ET from the GLDAS-CLSM data is the largest, while that from the GLDAS-VIC data is the smallest. Figure 11a,b qualitatively depict the 95% confidence interval error of precipitation and ET, respectively. The uncertainty in the precipitation is 16.57 mm and that in the ET is higher than that in the precipitation at 27.01 mm. In addition, the uncertainty in the TWSA can be calculated by Equation (9), which reaches 36.05 mm.

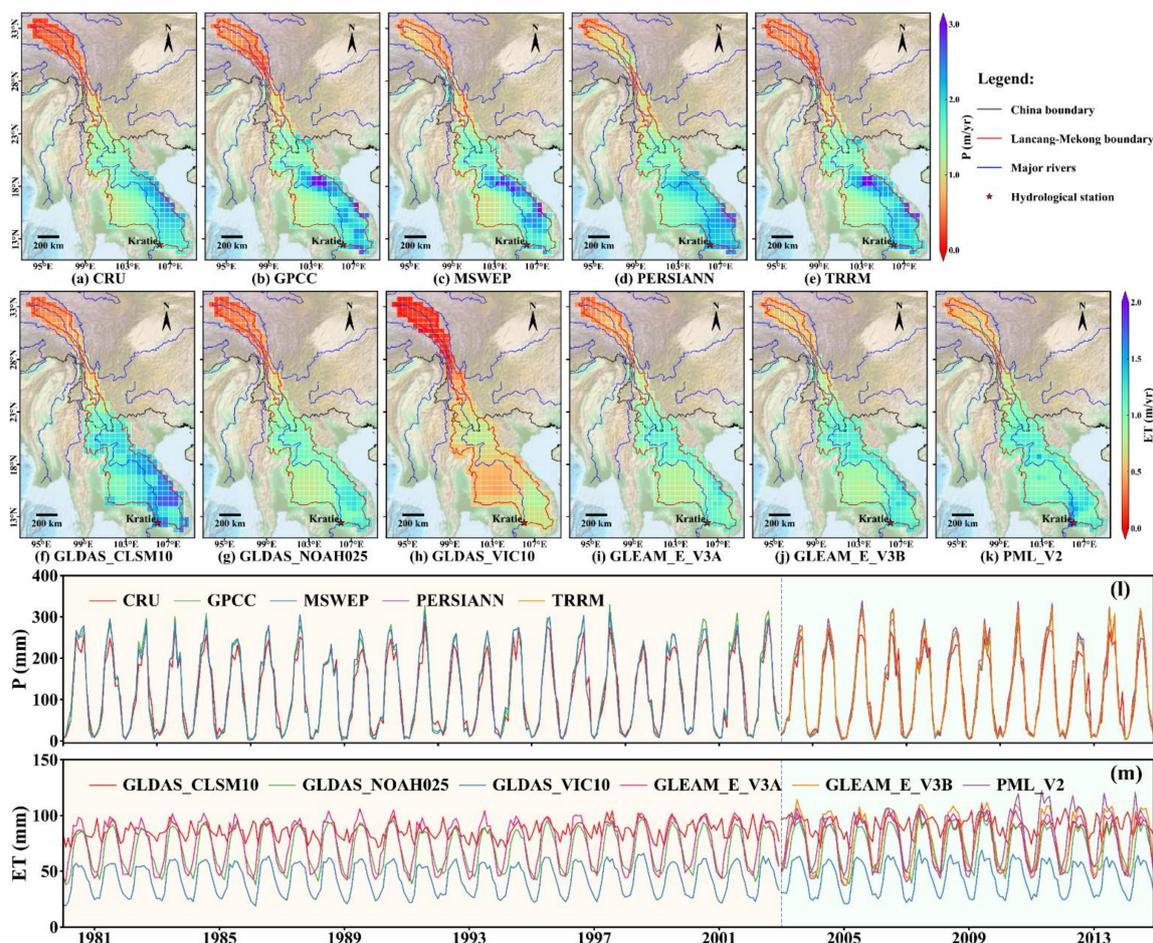


Figure 10. The spatial and temporal distribution of precipitation and ET products. (a–e) correspond to the spatial distribution of the CRU, GPCC, MSWEP, PERSIANN, and TRMM precipitation products, respectively; (f–k) correspond to the spatial distribution of the GLDAS-CLSM, GLDAS-Noah, GLDAS-VIC, GLEAM-V3A, GLEAM-V3B, and PML-V2 ET products, respectively; (l) corresponds to the monthly precipitation time series from (a–e); (m) corresponds to the monthly ET time series from (f–k). Limited by the data availability, only three precipitation products (MSWEP, GPCC, CRU) and four ET products (GLEAM v3.3a, GLDAS NOAH/VIC/CLSM) were used in the water balance method for 1980–2002.

Figure 12 quantitatively presents the effects of precipitation and ET on the estimated discharge during 2003–2014, which are visually expressed in the heatmap. The estimation of discharge from the WB-RF ($R = 0.74\text{--}0.89$, $NSE = 0.26\text{--}0.75$, $NRMSE = 0.34\text{--}0.59$) was slightly superior to that from the WB-GRACE ($R = 0.66\text{--}0.86$, $NSE = 0.23\text{--}0.70$, $NRMSE = 0.37\text{--}0.60$). For different combinations of precipitation and ET, the results vary greatly. In this study, five precipitation products and six ET products are used to reduce the uncertainty caused by a single data product [26,29,30]. Moreover, the uncertainty in the TWSC caused by leakage and measurement errors is calculated by Equation (9). The rigorous uncertainty in the GRACE TWS is still ongoing, with an average uncertainty value of 2 cm [11]. The coarser spatial resolution of GRACE TWSA introduces greater errors into the water balance equation, and machine learning is a valuable tool for the assimilation of observational and modeled information to achieve higher spatial and temporal resolutions for GRACE TWSA [9,62], thus promoting the application of the GRACE satellite.

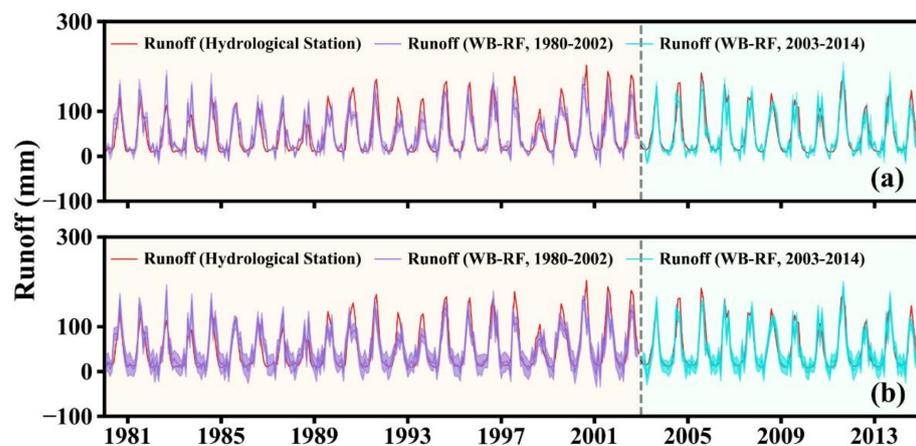


Figure 11. Monthly estimated discharge using the WB-RF. (a,b) correspond to the uncertainty in the precipitation and ET components in the water balance method, respectively.

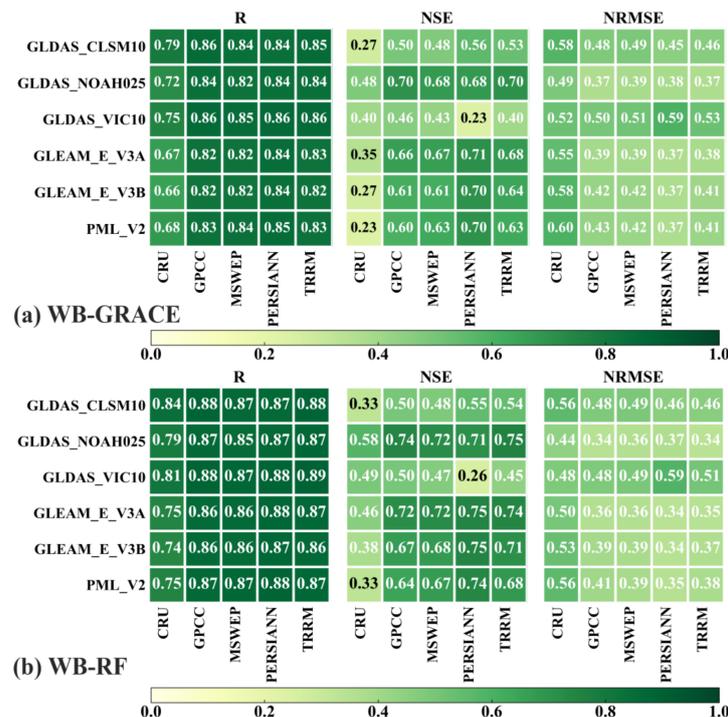


Figure 12. The quantitative performance of the multi-source precipitation and ET products on the estimated discharge. (a,b) refer to the estimated discharge using WB-GRACE and WB-RF, respectively.

5.3. Outlook of the Data-Driven Method

With rapid interdisciplinary development, the application of machine learning and data mining in the field of meteorology and hydrology is increasing [60]. Although data-driven methods are less explanatory than physical methods, they can achieve higher performance and execution efficiency [9,17]. With the enhancement of the acquisition capacity of remote sensing data, the era of big data has arrived. Combining physical process models with data-based methods will guide machine learning in the future [63]. Although the RF model has performed well in TWSA simulation, it can be improved in the following aspects. A more persuasive physical process (characteristics of snowmelt and the effect of human activity such as reservoir storage changes, etc.) should be considered in the data-driven approach. Furthermore, the RF is an ensemble learning method without consideration of temporal correlation. More advanced machine learning algorithms, such as recursive neural networks, should be explored in the future.

6. Conclusions

The TWSC is an indispensable component in the estimated discharge at the sub-annual level. The missing historical GRACE data limit its further application, and a method to fill in the gaps is urgently needed. To help address the limitations, the TWSA time series is reconstructed by the RF model based on GLDAS and CRU data, and the performance of this model is optimized through the NRMSE. Furthermore, the water balance method is employed to estimate discharge in the Lancang-Mekong River basin, combining the average precipitation (MSWEP, GPCC, PERSIANN-CDR, TRMM and CRU), ET (GLEAM v3.3a/b, PML-V2 and GLDAS NOAH/VIC/CLSM), and RF-built TWSA data. The primary conclusions are summarized as follows.

In general, the RF model can effectively construct the monthly TWSA time series and fill in missing historical data. The performance of the TWSA time series constructed by the RF model is better than that of the GLDAS TWSA at grid and basinscales. The R, NSE, and NRMSE values are 0.90, 0.80, and 0.44, respectively, for the RF model at the grid scale and 0.97, 0.93, and 0.26, respectively, at the basin scale for 2003–2014. Better performance at the basin scale is caused by wide spatial coverage. The RF-built TWSA time series achieves low performance in the northern Qinghai–Tibet Plateau region without considering the snow melting process. Moreover, the historical GRACE TWSA time series from 1980–2002 was effectively reconstructed. The GLDAS land surface model data and CRU meteorological data have greatly enriched the hydrological database, and the RF model can effectively integrate and utilize effective information to reconstruct the GRACE time series. As an ensemble learning method in machine learning, the RF model has the potential to explore the complex relationship between big data in the hydrological field.

The RF-built TWSA time series, combined with the water balance method, can be used to successfully estimate discharge in the Lancang-Mekong River basin. Overall, the discharge calculated by the WB-RF agrees well with the observed discharge ($R = 0.89$, $NSE = 0.78$, $NRMSE = 0.32$). In addition, desirable results are also obtained for 1980–2002 ($R = 0.84$, $NSE = 0.70$, $NRMSE = 0.37$). The uncertainty in the estimated discharge is quantitatively assessed, and the average uncertainties of the precipitation, ET, and TWSC are 16.57, 27.01, and 36.05 mm, respectively. Surface discharge is an important part of the water cycle, and the estimation of discharge is difficult in areas without hydrological stations. This study combines the water balance equation and multi-source remote sensing products, which confirms the reliability of the estimated discharge and provides guidance for sustainable water resource management in the Lancang-Mekong River basin. For complex terrains, harsh climates, or underdeveloped areas, this method can be used as an effective discharge monitoring method.

Author Contributions: Conceptualization, S.T., H.W. and W.L.; Formal analysis, S.T.; Supervision, T.W.; Visualization, S.T.; Writing—original draft, S.T.; Writing—review and editing, H.W., Y.F., Q.L., W.L. and F.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Sciences Foundation of China (42001015 and 42001031), the Program for the “Kezhen-Bingwei” Youth Talents (2020RC004) from the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, the China Postdoctoral Science Foundation (2018M640173, 2020M670432, 2021T140657, and 2020T130646), and the Top-Notch Young Talents Program of China (Fubao Sun).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: This study was performed based on public-access data. Data utilized in this research can be requested by contacting the first author at a reasonable request.

Acknowledgments: We would like to thank the University of Texas at Austin for providing the CSR RL06 GRACE mascon solutions. We are grateful to the Goddard Earth Sciences Data and Information Center (GES DISC) for providing the GLDAS data. We sincerely appreciate the meteorological data provided by organizations and individuals. We also sincerely thank the Mekong River Commission (MRC) for providing the in-situ river discharge measurements. We also wish to thank the CRU, MSWEP, GPCC, TRRM, GLEAM v3.3a/b and PML-V2, for the various datasets used in this work. Finally, we greatly acknowledge the anonymous referees for their valuable constructive suggestions and comments, which were helpful in improving the quality of our manuscript.

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

References

- Jing, W.; Zhao, X.; Yao, L.; Jiang, H.; Xu, J.; Yang, J.; Li, Y. Variations in terrestrial water storage in the Lancang-Mekong river basin from GRACE solutions and land surface model. *J. Hydrol.* **2020**, *580*, 124258. [[CrossRef](#)]
- Lu, Y.; Tian, F.; Guo, L.; Borzi, I.; Patil, R.; Wei, J.; Liu, D.; Wei, Y.; Yu, D.J.; Sivapalan, M. Socio-Hydrologic Modeling of the Dynamics of Cooperation in the Transboundary Lancang-Mekong River. *Hydrol. Earth Syst. Sci. Discuss.* **2020**, *25*, 1883–1903. [[CrossRef](#)]
- Mohammed, I.N.; Bolten, J.D.; Srinivasan, R.; Lakshmi, V. Satellite observations and modeling to understand the Lower Mekong River Basin streamflow variability. *J. Hydrol.* **2018**, *564*, 559–573. [[CrossRef](#)]
- Mohammed, I.N.; Bolten, J.D.; Srinivasan, R.; Meechaiya, C.; Spruce, J.P.; Lakshmi, V. Ground and satellite based observation datasets for the Lower Mekong River Basin. *Data Brief* **2018**, *21*, 2020–2027. [[CrossRef](#)]
- Yun, X.; Tang, Q.; Wang, J.; Liu, X.; Zhang, Y.; Lu, H.; Wang, Y.; Zhang, L.; Chen, D. Impacts of climate change and reservoir operation on streamflow and flood characteristics in the Lancang-Mekong River Basin. *J. Hydrol.* **2020**, *590*, 125472. [[CrossRef](#)]
- Thomas, A.C.; Reager, J.T.; Famiglietti, J.S.; Rodell, M. A GRACE-based water storage deficit approach for hydrological drought characterization. *Geophys. Res. Lett.* **2014**, *41*, 1537–1545. [[CrossRef](#)]
- Reager, J.T.; Thomas, B.F.; Famiglietti, J.S. River basin flood potential inferred using GRACE gravity observations at several months lead time. *Nat. Geosci.* **2014**, *7*, 588–592. [[CrossRef](#)]
- Thomas, B.F.; Famiglietti, J.S.; Landerer, F.W.; Wiese, D.N.; Molotch, N.P.; Argus, D.F. GRACE Groundwater Drought Index: Evaluation of California Central Valley groundwater drought. *Remote Sens. Environ.* **2017**, *198*, 384–392. [[CrossRef](#)]
- Sun, Z.L.; Long, D.; Yang, W.T.; Li, X.Y.; Pan, Y. Reconstruction of GRACE Data on Changes in Total Water Storage Over the Global Land Surface and 60 Basins. *Water Resour. Res.* **2020**, *56*, 21. [[CrossRef](#)]
- Scanlon, B.R.; Zhang, Z.; Save, H.; Sun, A.Y.; Muller Schmied, H.; van Beek, L.P.H.; Wiese, D.N.; Wada, Y.; Long, D.; Reedy, R.C.; et al. Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E1080–E1089. [[CrossRef](#)] [[PubMed](#)]
- Wiese, D.N.; Landerer, F.W.; Watkins, M.M. Quantifying and reducing leakage errors in the JPL RL05M GRACE mascon solution. *Water Resour. Res.* **2016**, *52*, 7490–7502. [[CrossRef](#)]
- Pellet, V.; Aires, F.; Papa, F.; Munier, S.; Decharme, B. Long-term total water storage change from a Satellite Water Cycle reconstruction over large southern Asian basins. *Hydrol. Earth Syst. Sci.* **2020**, *24*, 3033–3055. [[CrossRef](#)]
- Suzuki, K.; Park, H.; Makarieva, O.; Kanamori, H.; Hori, M.; Matsuo, K.; Matsumura, S.; Nesterova, N.; Hiyama, T. Effect of Permafrost Thawing on Discharge of the Kolyma River, Northeastern Siberia. *Remote Sens.* **2021**, *13*, 4389. [[CrossRef](#)]
- Xu, L.; Chen, N.C.; Zhang, X.; Chen, Z.Q. Spatiotemporal Changes in China’s Terrestrial Water Storage From GRACE Satellites and Its Possible Drivers. *J. Geophys. Res.-Atmos.* **2019**, *124*, 11976–11993. [[CrossRef](#)]
- Zhang, L.; Dobslaw, H.; Stacke, T.; Güntner, A.; Dill, R.; Thomas, M. Validation of terrestrial water storage variations as simulated by different global numerical models with GRACE satellite observations. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 821–837. [[CrossRef](#)]
- Humphrey, V.; Gudmundsson, L.; Seneviratne, S.I. A global reconstruction of climate-driven subdecadal water storage variability. *Geophys. Res. Lett.* **2017**, *44*, 2300–2309. [[CrossRef](#)]
- Humphrey, V.; Gudmundsson, L. GRACE-REC: A reconstruction of climate-driven water storage changes over the last century. *Earth Syst. Sci. Data* **2019**, *11*, 1153–1170. [[CrossRef](#)]

18. Jing, W.L.; Di, L.P.; Zhao, X.D.; Yao, L.; Xia, X.L.; Liu, Y.X.; Yang, J.; Li, Y.; Zhou, C.H. A data-driven approach to generate past GRACE-like terrestrial water storage solution by calibrating the land surface model simulations. *Adv. Water Resour.* **2020**, *143*, 103683. [[CrossRef](#)]
19. Ndehedehe, C.E.; Ferreira, V.G. Assessing land water storage dynamics over South America. *J. Hydrol.* **2020**, *580*, 124339. [[CrossRef](#)]
20. Huang, S.Z.; Zheng, X.D.; Ma, L.; Wang, H.; Huang, Q.; Leng, G.Y.; Meng, E.H.; Guo, Y. Quantitative contribution of climate change and human activities to vegetation cover variations based on GA-SVM model. *J. Hydrol.* **2020**, *584*, 124687. [[CrossRef](#)]
21. Nguyen, P.; Shearer, E.J.; Tran, H.; Ombadi, M.; Hayatbini, N.; Palacios, T.; Huynh, P.; Braithwaite, D.; Updegraff, G.; Hsu, K.; et al. The CHRS Data Portal, an easily accessible public repository for PERSIANN global satellite precipitation data. *Sci. Data* **2019**, *6*, 180296. [[CrossRef](#)] [[PubMed](#)]
22. Li, M.; Zhang, Y.; Wallace, J.; Campbell, E. Estimating annual runoff in response to forest change: A statistical method based on random forest. *J. Hydrol.* **2020**, *589*, 125168. [[CrossRef](#)]
23. Jing, W.L.; Zhao, X.D.; Yao, L.; Di, L.P.; Yang, J.; Li, Y.; Guo, L.Y.; Zhou, C.H. Can Terrestrial Water Storage Dynamics be Estimated From Climate Anomalies? *Earth Space Sci.* **2020**, *7*, 19. [[CrossRef](#)]
24. Jing, W.; Zhang, P.; Zhao, X.; Yang, Y.; Jiang, H.; Xu, J.; Yang, J.; Li, Y. Extending GRACE terrestrial water storage anomalies by combining the random forest regression and a spatially moving window structure. *J. Hydrol.* **2020**, *590*, 125239. [[CrossRef](#)]
25. Gibson, R.; Danaher, T.; Hehir, W.; Collins, L. A remote sensing approach to mapping fire severity in south-eastern Australia using sentinel 2 and random forest. *Remote Sens. Environ.* **2020**, *240*, 111702. [[CrossRef](#)]
26. Xie, J.K.; Xu, Y.P.; Gao, C.; Xuan, W.D.; Bai, Z.X. Total Basin Discharge From GRACE and Water Balance Method for the Yarlung Tsangpo River Basin, Southwestern China. *J. Geophys. Res.-Atmos.* **2019**, *124*, 7617–7632. [[CrossRef](#)]
27. Wang, F.; Wang, Z.M.; Yang, H.B.; Di, D.Y.; Zhao, Y.; Liang, Q.H. Utilizing GRACE-based groundwater drought index for drought characterization and teleconnection factors analysis in the North China Plain. *J. Hydrol.* **2020**, *585*, 124849. [[CrossRef](#)]
28. Zhao, M.; Geruo, A.; Zhang, J.; Velicogna, I.; Liang, C.; Li, Z. Ecological restoration impact on total terrestrial water storage. *Nat. Sustain.* **2021**, *4*, 56–62. [[CrossRef](#)]
29. Liu, W.B.; Wang, L.; Zhou, J.; Li, Y.Z.; Sun, F.B.; Fu, G.B.; Li, X.P.; Sang, Y.F. A worldwide evaluation of basin-scale evapotranspiration estimates against the water balance method. *J. Hydrol.* **2016**, *538*, 82–95. [[CrossRef](#)]
30. Liu, W.B.; Sun, F.B.; Li, Y.Z.; Zhang, G.Q.; Sang, Y.F.; Lim, W.H.; Liu, J.H.; Wang, H.; Bai, P. Investigating water budget dynamics in 18 river basins across the Tibetan Plateau through multiple datasets. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 351–371. [[CrossRef](#)]
31. Goncalves, R.D.; Stollberg, R.; Weiss, H.; Chang, H.K. Using GRACE to quantify the depletion of terrestrial water storage in Northeastern Brazil: The Urucuia Aquifer System. *Sci. Total Environ.* **2020**, *705*, 135845. [[CrossRef](#)]
32. Cavalcante, R.B.L.; Pontes, P.R.M.; Tedeschi, R.G.; Costa, C.P.W.; Ferreira, D.B.S.; Souza, P.W.M.; de Souza, E.B. Terrestrial water storage and Pacific SST affect the monthly water balance of Itacaiunas River Basin (Eastern Amazonia). *Int. J. Climatol.* **2020**, *40*, 3021–3035. [[CrossRef](#)]
33. Han, J.T.; Yang, Y.T.; Roderick, M.L.; McVicar, T.R.; Yang, D.W.; Zhang, S.L.; Beck, H.E. Assessing the Steady-State Assumption in Water Balance Calculation Across Global Catchments. *Water Resour. Res.* **2020**, *56*, 16. [[CrossRef](#)]
34. Chen, J.; Tapley, B.; Rodell, M.; Seo, K.W.; Wilson, C.; Scanlon, B.R.; Pokhrel, Y. Basin-Scale River Runoff Estimation From GRACE Gravity Satellites, Climate Models, and In Situ Observations: A Case Study in the Amazon Basin. *Water Resour. Res.* **2020**, *56*, e2020WR028032. [[CrossRef](#)]
35. Pascolini-Campbell, M.A.; Reager, J.T.; Fisher, J.B. GRACE-based Mass Conservation as a Validation Target for Basin-Scale Evapotranspiration in the Contiguous United States. *Water Resour. Res.* **2020**, *56*, 18. [[CrossRef](#)]
36. Wan, Z.M.; Zhang, K.; Xue, X.W.; Hong, Z.; Hong, Y.; Gourley, J.J. Water balance-based actual evapotranspiration reconstruction from ground and satellite observations over the conterminous United States. *Water Resour. Res.* **2015**, *51*, 6485–6499. [[CrossRef](#)]
37. Lv, M.X.; Ma, Z.G.; Yuan, X.; Lv, M.Z.; Li, M.X.; Zheng, Z.Y. Water budget closure based on GRACE measurements and reconstructed evapotranspiration using GLDAS and water use data for two large densely-populated mid-latitude basins. *J. Hydrol.* **2017**, *547*, 585–599. [[CrossRef](#)]
38. Burbano, M.; Shin, S.; Khanh, N.; Pokhrel, Y. Hydrologic changes, dam construction, and the shift in dietary protein in the Lower Mekong River Basin. *J. Hydrol.* **2020**, *581*, 124454. [[CrossRef](#)]
39. Chen, A.F.; Chen, D.L.; Azorin-Molina, C. Assessing reliability of precipitation data over the Mekong River Basin: A comparison of ground-based, satellite, and reanalysis datasets. *Int. J. Climatol.* **2018**, *38*, 4314–4334. [[CrossRef](#)]
40. Liu, S.A.; Li, X.; Chen, D.; Duan, Y.Q.; Ji, H.Y.; Zhang, L.P.; Chai, Q.; Hu, X.D. Understanding Land use/Land cover dynamics and impacts of human activities in the Mekong Delta over the last 40 years. *Glob. Ecol. Conserv.* **2020**, *22*, e00991. [[CrossRef](#)]
41. Gesch, D.B.; Brock, J.; Parrish, C.E.; Rogers, J.N.; Wright, C.W. Introduction: Special issue on advances in topobathymetric mapping, models, and applications. *J. Coast. Res.* **2016**, 1–3. [[CrossRef](#)]
42. Ruiz-Barradas, A.; Nigam, S. Hydroclimate Variability and Change over the Mekong River Basin: Modeling and Predictability and Policy Implications. *J. Hydrometeorol.* **2018**, *19*, 849–869. [[CrossRef](#)]
43. Wang, S.; Zhang, L.; She, D.; Wang, G.; Zhang, Q. Future projections of flooding characteristics in the Lancang-Mekong River Basin under climate change. *J. Hydrol.* **2021**, *602*, 126778. [[CrossRef](#)]
44. Landerer, F.W.; Swenson, S.C. Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resour. Res.* **2012**, *48*, 11. [[CrossRef](#)]

45. Save, H.; Bettadpur, S.; Tapley, B.D. High-resolution CSR GRACE RL05 mascons. *J. Geophys. Res.-Solid Earth* **2016**, *121*, 7547–7569. [[CrossRef](#)]
46. Xie, Z.; Huete, A.; Cleverly, J.; Phinn, S.; McDonald-Madden, E.; Cao, Y.; Qin, F. Multi-climate mode interactions drive hydrological and vegetation responses to hydroclimatic extremes in Australia. *Remote Sens. Environ.* **2019**, *231*, R713–R715. [[CrossRef](#)]
47. Scanlon, B.R.; Zhang, Z.Z.; Save, H.; Wiese, D.N.; Landerer, F.W.; Long, D.; Longuevergne, L.; Chen, J. Global evaluation of new GRACE mascon products for hydrologic applications. *Water Resour. Res.* **2016**, *52*, 9412–9429. [[CrossRef](#)]
48. Dangar, S.; Mishra, V. Natural and anthropogenic drivers of the lost groundwater from the Ganga river basin. *Environ. Res. Lett.* **2021**, *16*, 114009. [[CrossRef](#)]
49. Harris, I.; Osborn, T.J.; Jones, P.; Lister, D. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci. Data* **2020**, *7*, 109. [[CrossRef](#)]
50. Beck, H.E.; Wood, E.F.; Pan, M.; Fisher, C.K.; Miralles, D.G.; van Dijk, A.; McVicar, T.R.; Adler, R.F. MSWEP V2 Global 3-Hourly 0.1 degrees Precipitation: Methodology and Quantitative Assessment. *Bull. Amer. Meteorol. Soc.* **2019**, *100*, 473–502. [[CrossRef](#)]
51. Schamm, K.; Ziese, M.; Becker, A.; Finger, P.; Meyer-Christoffer, A.; Schneider, U.; Schröder, M.; Stender, P. Global gridded precipitation over land: A description of the new GPCC First Guess Daily product. *Earth Syst. Sci. Data* **2014**, *6*, 49–60. [[CrossRef](#)]
52. Beck, H.E.; Vergopolan, N.; Pan, M.; Levizzani, V.; van Dijk, A.; Weedon, G.P.; Brocca, L.; Pappenberger, F.; Huffman, G.J.; Wood, E.F. Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 6201–6217. [[CrossRef](#)]
53. Martens, B.; Miralles, D.G.; Lievens, H.; van der Schalie, R.; de Jeu, R.A.M.; Fernandez-Prieto, D.; Beck, H.E.; Dorigo, W.A.; Verhoest, N.E.C. GLEAM v3: Satellite-based land evaporation and root-zone soil moisture. *Geosci. Model Dev.* **2017**, *10*, 1903–1925. [[CrossRef](#)]
54. Zhang, Y.Q.; Kong, D.D.; Gan, R.; Chiew, F.H.S.; McVicar, T.R.; Zhang, Q.; Yang, Y.T. Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002–2017. *Remote Sens. Environ.* **2019**, *222*, 165–182. [[CrossRef](#)]
55. He, Q.; Chun, K.P.; Fok, H.S.; Chen, Q.; Dieppois, B.; Massei, N. Water storage redistribution over East China, between 2003 and 2015, driven by intra- and inter-annual climate variability. *J. Hydrol.* **2020**, *583*, e2020WR027392. [[CrossRef](#)]
56. Deng, H.J.; Chen, Y.N. Influences of recent climate change and human activities on water storage variations in Central Asia. *J. Hydrol.* **2017**, *544*, 46–57. [[CrossRef](#)]
57. Jing, W.L.; Yao, L.; Zhao, X.D.; Zhang, P.Y.; Liu, Y.X.Y.; Xia, X.L.; Song, J.; Yang, J.; Li, Y.; Zhou, C.H. Understanding Terrestrial Water Storage Declining Trends in the Yellow River Basin. *J. Geophys. Res.-Atmos.* **2019**, *124*, 12963–12984. [[CrossRef](#)]
58. Zhang, Y.Q.; Chiew, F.H.S.; Liu, C.M.; Tang, Q.H.; Xia, J.; Tian, J.; Kong, D.D.; Li, C.C. Can Remotely Sensed Actual Evapotranspiration Facilitate Hydrological Prediction in Ungauged Regions Without Runoff Calibration? *Water Resour. Res.* **2020**, *56*, 15. [[CrossRef](#)]
59. Pham, L.T.; Luo, L.; Finley, A.O. Evaluation of Random Forest for short-term daily streamflow forecast in rainfall and snowmelt driven watersheds. *Hydrol. Earth Syst. Sci. Discuss.* **2020**, *25*, 2997–3015. [[CrossRef](#)]
60. Yan, J.B.; Jia, S.F.; Lv, A.F.; Zhu, W.B. Water Resources Assessment of China's Transboundary River Basins Using a Machine Learning Approach. *Water Resour. Res.* **2019**, *55*, 632–655. [[CrossRef](#)]
61. Lutz, A.F.; Immerzeel, W.W.; Shrestha, A.B.; Bierkens, M.F.P. Consistent increase in High Asia's runoff due to increasing glacier melt and precipitation. *Nat. Clim. Chang.* **2014**, *4*, 587–592. [[CrossRef](#)]
62. Vishwakarma, B.D.; Zhang, J.; Sneeuw, N. Downscaling GRACE total water storage change using partial least squares regression. *Sci. Data* **2021**, *8*, 95. [[CrossRef](#)] [[PubMed](#)]
63. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195–204. [[CrossRef](#)] [[PubMed](#)]