A Novel Standardized Drought and Flood Potential Index Based on Reconstructed Daily GRACE Data

JINGHUA XIONG,[a] SHENGLIAN GUO,[a] ABHISHEK,[b] JUN LI,[c] AND JIABO YIN[a]

[a] *State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan, China*
[b] *School of Environment and Society, Tokyo Institute of Technology, Yokohama, Japan*
[c] *School of Civil Engineering and Transportation, State Key Laboratory of Subtropical Building Science, South China University of Technology, Guangzhou, China*

(Manuscript received 21 January 2022, in final form 13 June 2022)

ABSTRACT: Multiple indicators derived from the Gravity Recovery and Climate Experiment (GRACE) satellite have been used in monitoring floods and droughts. However, these measures are constrained by the relatively short time span (~20 years) and coarse temporal resolution (1 month) of the GRACE and GRACE Follow-On missions, and the inherent decay mechanism of the land surface system has not been considered. Here we reconstructed the daily GRACE-like terrestrial water storage anomaly (TWSA) in the Yangtze River basin (YRB) during 1961–2015 based on the Institute of Geodesy at Graz University of Technology (ITSG)-Grace2018 solution using the random forest (RF) model. A novel antecedent metric, namely, standardized drought and flood potential index (SDFPI), was developed using reconstructed TWSA, observed precipitation, and modeled evapotranspiration. The potential of SDFPI was evaluated against in situ discharge, VIC simulations, and several widely used indices such as total storage deficit index (TSDI), self-calibrated Palmer drought severity index (sc-PDSI), and multiscale standardized precipitation evapotranspiration index (SPEI). Daily SDFPI was utilized to monitor and characterize short-term severe floods and droughts. The results illustrate a reasonably good accuracy of ITSG-Grace2018 solution when compared with the hydrological model output and regional water balance estimates. The RF model presents satisfactory performances for the TWSA reconstruction, with a correlation coefficient of 0.88 and Nash–Sutcliffe efficiency of 0.76 during the test period 2011–15. Spatiotemporal propagation of the developed SDFPI corresponds well with multiple indices when examined for two typical short-term events, including the 2003 flood and 2013 drought. A total of 22 submonthly exceptional floods and droughts were successfully detected and featured using SDFPI, highlighting its outperformance and capabilities in providing inferences for decision-makers and stakeholders to monitor and mitigate the short-term floods and droughts.

KEYWORDS: Asia; Atmosphere-land interaction; Machine learning

1. Introduction

Terrestrial water storage (TWS), the sum of water stored in canopies, rivers and lakes, ice and snow, soil, and groundwater, is a pivotal component of the global water cycle and energy budget (Rodell et al. 2018). TWS plays an essential role in multiple interconnected phenomena such as socioeconomic manifestation, determining freshwater availability, and modulating the various hydrological processes (Abhishek et al. 2021; Tapley et al. 2019). Furthermore, TWS changes are significantly related to the formation of hydrological extremes, including floods and droughts (Houborg et al. 2012; Thomas et al. 2014). However, TWS variations are poorly understood due to inadequate in situ observations worldwide and substantial uncertainties in hydrological and land surface simulations attributed to the inadequate model physics or the inevitable uncertainty propagation from the forcing datasets (Abhishek and Kinouchi 2021; Scanlon et al. 2018).

Launched initially in March 2002, the Gravity Recovery and Climate Experiment (GRACE) twin satellites project and its successor GRACE Follow-on (both jointly referred to as GRACE) have become an unprecedented tool to measure the large-scale TWS anomalies (TWSA) from space (Tapley et al. 2004). The GRACE mission has provided monthly global TWSA data with a spatial resolution of approximately 150 000 km^2. Given the inherent linkage between TWS and major hydrological fluxes such as precipitation, runoff, and evapotranspiration, GRACE-based TWSA and subsequently derived indicators have been widely used in large-scale drought and flood monitoring studies (Abhishek and Kinouchi 2021; Reager and Famiglietti 2009; Reager et al. 2014; Long et al. 2014). By taking into account the terrestrial water capacity in the runoff generation, Reager and Famiglietti (2009) proposed the flood potential index (FPI) based on GRACE TWSA and precipitation for large-scale flood monitoring. Long et al. (2014) improved the FPI by considering the return period of floods and used the improved FPI in assessing the 2011 exceptional flood in the Yun-Gui Plateau, China. Similarly, the threshold values of FPI for different intensity floods have been determined based on the threshold of discharge for various percentiles (Gupta and Dhanya 2020). In general, the FPI and improved FPI have demonstrated great potential in flood monitoring worldwide (Sun et al. 2017;

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JHM-D-22-0011.s1.

Corresponding author: Shenglian Guo, slguo@whu.edu.cn

DOI: 10.1175/JHM-D-22-0011.1

X. Chen et al. 2018; Idowu and Zhou 2019; Yang et al. 2021). Alternatively, due to the sensitivity of GRACE TWSA to changes in surface water, soil moisture, and groundwater, many GRACE-based drought indices have been established to assess hydrological droughts and agricultural droughts (Ramillien et al. 2008; Yirdaw et al. 2008; Frédéric and Guillaume 2018; Li et al. 2019). More specifically, Yirdaw et al. (2008) initially proposed the total storage deficit index (TSDI) and proved its effectiveness in characterizing the 2002/03 Canadian Prairie droughts. Nie et al. (2018) reformed the TSDI and examined it using several severe droughts worldwide, including the 2005 drought in Africa, the 2006 drought in China, and the 2010 drought in Brazil. Further, the modified TSDI was introduced because the original TSDI was not efficient for regions with intensive anthropogenic impacts such as irrigation and water diversion (Hosseini-Moghari et al. 2019, 2020). Apart from the TSDI, other well-known drought metrics such as the GRACE-drought severity index (GRACE-DSI) (Zhao et al. 2017), water storage deficit index (WSDI) (Thomas et al. 2014), GRACE-based hydrological drought index (GHDI) (Yi and Wen 2016), combined climatologic deviation index (CCDI) (Sinha et al. 2019), and the most recent drought potential index (DPI) (Abhishek et al. 2021) have been developed and studied in various river basins globally. Although the abovementioned GRACE-based indicators can effectively monitor drought or flood events, they merely consider the terrestrial wetness/dryness conditions at the current stage, thus neglecting the influences of TWSA in previous periods caused by the inherent decay mechanism of the land system.

Despite the great potential in regional and global flood and drought monitoring, the relatively short observation period of GRACE satellites of about 20 years makes them difficult to quantify long-term TWSA changes, thus affecting the reliability and applicability of GRACE-based flood and monitoring indices. Moreover, there is an 11-month data gap between the two generations of GRACE satellites. To overcome these shortcomings, numerous studies have attempted to generate longer-term TWSA series beyond the period of the GRACE missions to allow for the monitoring of floods and droughts (Humphrey et al. 2017; Humphrey and Gudmundsson 2019; Sun et al. 2021). For instance, Long et al. (2014) initially applied the artificial neural network model combined with precipitation, air temperature, and modeled TWSA to extend GRACE TWSA back to 1979 in southwest China. Further, Sun et al. (2020) employed three machine learning models to reconstruct TWSA derived from six GRACE solutions for 60 major global river basins and global land, and comprehensively evaluated their performance at basin and grid scales. A recent study successfully adopted the state-of-the-art automated machine learning approach to generate long-term continuous TWSA series over the conterminous United States (Sun et al. 2021). To sum up, many data-driven methods, including random forest, long short-term memory, and support vector machine approaches, have presented the satisfactory ability to predict GRACE-like TWSA globally (Yang et al. 2018; Jing et al. 2020; F. Wang et al. 2021). Given their added value, extended GRACE datasets have recently become preferential options to assess long-term dryness and wetness conditions in hydrology and climate study (Humphrey and Gudmundsson 2019).

Another limitation of GRACE-based flood and drought monitoring indices is the coarse time scale (monthly), which is inadequate to capture submonthly events. In such cases, short-term droughts at the submonthly scale and most floods that generally develop within a short period from hours to days cannot be monitored by GRACE-inferred metrics. Therefore, a few studies have developed improved GRACE data with temporal resolution from a day to 10 days using daily GRACE Level-1B observations and advanced interpolation methods (Sakumura et al. 2016; Croteau et al. 2020; Schindelegger et al. 2021). Indeed, the improved daily GRACE data showed a high correlation with runoff measurements during floods in, e.g., the Ganges–Brahmaputra delta, highlighting the efficient flood monitoring ability (Gouweleeuw et al. 2018). A recent study also revealed that the derived daily GRACE data evidenced the 2020 monsoon season due to premature soil saturation (Han et al. 2021) temporally and spatially. In addition, daily GRACE solutions illustrated reasonably consistent relationships with different hydrometeorological fluxes such as evapotranspiration and runoff from global atmospheric reanalyzes, suggesting the potential to detect short-term extremes such as floods and droughts (Eicker et al. 2020; Sharma et al. 2020).

This study extends previous hydrological applications of GRACE by developing a more robust and reliable approach for flood and drought monitoring using improved GRACE data. In detail, we aim to 1) reconstruct daily GRACE-like TWSA during the period 1961–2015 using the random forest method; 2) develop an antecedent daily-scale flood and drought monitoring index, that is, the standardized drought and flood potential index (SDFPI), by considering the time lag of terrestrial response to climate factors based on extended GRACE TWSA and precipitation data; and 3) characterize drought and flood events including onset, duration, peak severity, average severity, affected area, and cessation using the SDFPI. This study attempts to enrich indices for effective monitoring of short-term floods and droughts in data-sparse regions and provide useful information such as the start, development, and decay of such events for stakeholders and decision-makers.

2. Study area

Located within the range of 24°28′–35°46′N and 90°32′–121°55′E, the Yangtze River basin (YRB) is chosen as our study region due to four reasons (see Fig. 1). First, it is featured the subtropical and tropical humid monsoon climate, leading to abundant precipitation and frequent typhoons during the wet season. The distribution of yearly precipitation is uneven, with more than 70% of annual precipitation being received in the six rainy months between May and October, which favors the formation of hydrological extremes such as floods and droughts (Li et al. 2018). Second, characterized by high drainage density and flat riverbeds, the midstream and downstream of YRB are expected to face high flood risks (Zhang et al. 2020). Third, considering that the YRB accounts for more than

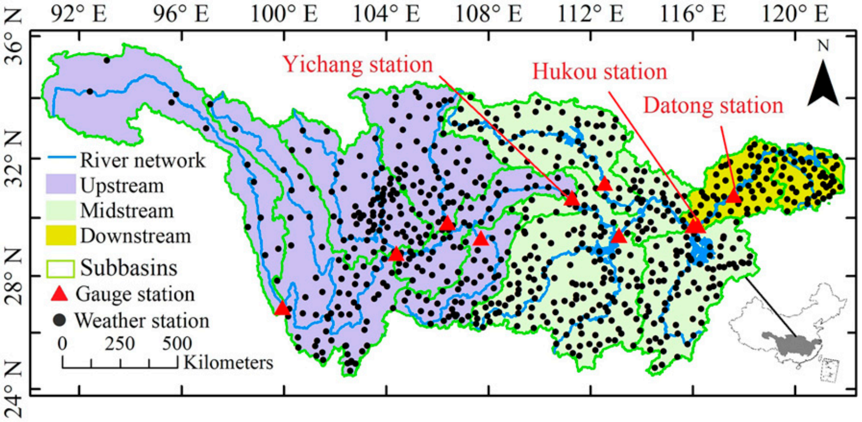

FIG. 1. Location of the Yangtze River basin. The Yichang, Hukou, and Datong gauge stations are the outlet of the upstream, midstream, and downstream reaches of the YRB, respectively. Various subbasins have been delineated by green polygons and the 739 weather stations are shown by solid black circles.

50% and 30% of the gross domestic product and the total population of China, respectively, the water-related extremes are likely to cause huge damage to infrastructures, public services, and even loss of life (Xia and Chen 2021). Fourth, GRACE data are more reliable in humid areas than in dry areas because of relatively higher signal-to-noise ratios, which gives us confidence for the reconstruction of GRACE TWSA (Long et al. 2015).

3. Data

a. *Daily GRACE TWSA*

We utilized the ITSG-Grace2018 product to reconstruct the daily TWSA series from 1961 to 2015 in the Yangtze River basin. The latest daily ITSG-Grace2018 gravity field solution, solved by the Institute of Geodesy at the Graz University of Technology, was processed from Level-1B Release 03 observations of the GRACE mission between 2002 and 2016 from NASA's Jet Propulsion Laboratory (Ellmer et al. 2018). The ITSG-Grace2018 solution has followed the standard signal definition of official monthly GRACE products. In particular, geophysical background model simulations have been removed to carry out dealiasing, meaning that the ITSG-Grace2018 solution reflects the cumulative mass changes in the hydrosphere, cryosphere, glacial isostatic adjustment, earthquakes, residual atmosphere, and ocean signal. Furthermore, the terrestrial water storage component has been isolated by reducing the gravitational effect of atmospheric mass changes from the daily gravity field (Kvas et al. 2019). The ITSG-Grace2018 dataset was derived from spherical harmonic coefficients expansion up to the degree/order of 40 and released as the gridded TWSA with a spatial resolution of 1°. To restore daily TWSA signals from limited GRACE observations, a Kalman filter was used to fit the expected temporal variations of the gravity field, and a three-order autoregressive model was utilized to express the spatiotemporal correlations between epochs (Kurtenbach et al. 2012). Therefore, spatial filtering is unnecessary because the spatially correlated noise has been reduced by the Kalman filter. Other processing steps widely used in processing monthly GRACE data such as the geocentric correction, replacement of coefficient C_2^0, and removal of glacial isostatic adjustment have been omitted to focus on the submonthly TWSA changes (Swenson and Wahr 2006).

b. *VIC simulations*

Daily simulations of TWSA during 1961–2015 were used for the reconstruction of GRACE-like TWSA, which were collected from outputs by the Variable Infiltration Capacity (VIC-4.2d) model from Miao and Wang (2020). Moreover, VIC-based evapotranspiration was applied to estimate TWS changes based on the water balance equation for comparison with ITSG-Grace2018 results. The modeled soil moisture and runoff are used for comparisons with SDFPI during droughts and floods, respectively. Their spatial resolutions have been averaged from 0.25° to 1° to be consistent with GRACE data. The VIC model can effectively capture the terrestrial water cycle by simulating the canopy interception water, water stored in snow, surface water stored in ponding, runoff, and soil moisture within the depth of three soil layers (up to a depth of 200 cm). The VIC model has been widely used to analyze terrestrial water storage changes at both regional and global scales (Hao and Singh 2015; Hao et al. 2018). Detailed information about the forcing, configurations, and validations of the VIC model can be found in Miao and Wang (2020).

c. *In situ measurements*

Daily in situ precipitation and air temperature between 1961 and 2015 were collected from 739 weather stations over the YRB to reconstruct daily TWSA (see Fig. 1). These weather observations are temporally continuous and have thoroughly undergone quality control. To be consistent with GRACE data and VIC outputs, they are spatially interpolated into a 1° scale using the inverse distance weighting method (Gong et al. 2014). To compare with SDFPI during drought and flood events, we collected daily discharge data from 10 gauge stations located at the outlet of the main subbasins of the YRB (Fig. 1). These hydrological observations are standardized by subtracting the mean value and dividing by the standard deviation during the period 2003–15 for better comparison. Discharge observations during the period 2003–15 of the Yichang, Hukou, and Datong gauge stations, were alternatively used to obtain water balance estimates of terrestrial water storage changes, whose unit was transformed from water volume (m^3) to height (mm) by dividing the catchment area of the upper, middle, and lower reaches of the YRB, respectively.

4. Methods

We performed the comparison between daily GRACE solution and water balance estimates of TWSA based on the in situ measurement and VIC simulations. Thereafter, daily GRACE-like TWSA between 1961 and 2015 was reconstructed using the random forest method using several observed and simulated variables as the predictors. Furthermore, the SDFPI series was established using reconstructed TWSA and observed precipitation for spatiotemporal monitoring of short-term floods and droughts.

a. Random forest

The random forest (RF) model is a machine learning model proposed by Breiman (2001) to solve both the nonlinear regression and classification problems. As a flexible ensemble algorithm, the RF model consists of multiple decision trees, and each decision tree is generated using the random factor sampled independently from the training dataset. Each decision tree has many nodes, and the training dataset is segmented at each node using a random subset of predictors. This split process is repeated recursively on each decision tree until the splitting no longer improves the model predictions. The final result of the RF model is obtained from all decision trees by voting for classification or averaging for regression (Pelletier et al. 2016). The RF model is able to handle large differentiations within different hydrometeorological predictors and does not require an understanding of the prior data distribution. Another advantage of the model is that the risk of overfitting is effectively reduced by introducing the randomness in the selection of samples and attributes when growing each regression tree and the integration of numerous decorrelated decision trees (Hutengs and Vohland 2016; Probst et al. 2019). However, we do not convey that other algorithms will necessarily degrade the performance of TWSA reconstructions in the YRB due to the regional differences of divergent models (Sun et al. 2021). For example, some state-of-the-art recurrent neural networks [e.g., long short-term memory (LSTM)], which also consider the antecedent data into account, can be selected based on their performance and multimodel sensitivity analysis in the target study regions. In this study, the RF model was employed to directly reconstruct daily TWSA time series between 1961 and 2015, with three predictors including, in situ precipitation, air temperature, and VIC-modeled TWSA. Unlike the conventional way to remove and read linear trends and/or seasonal cycles of TWSA prior to reconstruction (Humphrey and Gudmundsson 2019), we straight predict the full TWSA because of the missing knowledge of TWSA outside the GRACE era (i.e., 1961–2002) and the need of the whole TWSA as input for the development of SDFPI. Daily TWSA from the ITSG-Grace2018 solution during 2003–15 served as the response variable of the RF model. In situ precipitation and air temperature are used as they are the main factors driving the terrestrial water and energy cycles (Trenberth 2011). VIC-modeled TWSA is also used due to the significant correlation with GRACE TWSA (CC = 0.86). Approximately 80% of the samples (2003–12) were used as the training set and the remaining 20% (2013–15) as the test set. A fivefold cross-validation method is employed to further split the training data into two subsets for training (80%) and validation (20%), respectively. A fivefold cross-validation method is a popular resampling procedure to estimate the robustness of the model on a limited dataset, which randomly shuffles the dataset and splits it into five groups to test the model skill in a specific group while keeping the other four groups for training. This approach is commonly applied in the machine learning discipline due to the simplicity and efficiency in testing model performance for independent data (G. Chen et al. 2018; Wei et al. 2019). The evaluation results are then used for model screening. There are three important parameters affecting the predictive skill of the RF model, namely, "ntree," "mtry," and "nodesize," which indicate the number of trees of the RF model, the number of variables tried at each split of the decision tree, and the minimum number of observations in a terminal node, respectively (Zhong et al. 2019; Maimaitijiang et al. 2020). Several metrics, including correlation coefficient (CC), root-mean-square error (RMSE), and Nash–Sutcliffe model efficiency coefficient (NSE), were used for performance evaluation.

b. Water balance equation

The water balance equation was used to estimate terrestrial water storage changes (ΔS) for evaluation of the ITSG-Grace2018 solution, which can be written as follows:

$$\Delta S_t = P_t - \mathrm{ET}_t - R_t \tag{1}$$

where P_t and R_t are observed precipitation and runoff, respectively, for the specific time t, and ET_t is the evapotranspiration modeled by the VIC model. Alternatively, changes in TWS can be expressed using GRACE data as follows:

$$\Delta S_t = \mathrm{TWSA}_t - \mathrm{TWSA}_{t-1}, \tag{2}$$

where TWSA_t is derived from the ITSG-Grace2018 data. The water balance method has been proven as an effective way to assess the accuracy of GRACE data (Rodell et al. 2018).

c. Rationale and formulation of SDFPI

The SDFPI series during 1961–2015 was developed by standardizing the antecedent terrestrial water storage deficit to monitor short-term hydrological extremes, i.e., floods and droughts, using TWSA reconstructions, VIC-modeled evapotranspiration, and in situ precipitation. To begin with, the storage deficit as the water that can be held in storage before achieving the historic storage anomaly was calculated (Reager and Famiglietti 2009):

$$S_t^{\text{def}} = S_{\max} - S_{t-1}, \tag{3}$$

where $S_{\max}$ and S_{t-1} are the historical daily maximum TWSA and that of the previous day, respectively. A higher S_t^{def} implies a higher water storage capacity for the current day. Then, the drought and flood potential amount (DFPA) was derived by subtracting the precipitation and evapotranspiration from the S_t^{def} for the specific day (t) as follows:

$$\text{DFPA}_t = P_t - \text{ET}_t - S_t^{\text{def}}, \tag{4}$$

where P_t is the observed precipitation for the day. ET_t denotes the simulated evapotranspiration from the VIC model. DFPA_t represents the difference between incoming water flux and the terrestrial storage deficit for the current day. A higher/lower DFPA can translate into the greater occurrence of a flood/drought. Furthermore, it is worth noting that the possibility of flood/drought on a certain day is not only related to the precipitation, evapotranspiration, and terrestrial deficit in the day, but also those in several previous days. For example, if a heavy rainstorm occurred in the present day, TWS will increase gradually in the following days until reaching the terrestrial storage capacity, followed by a slow decrease to the normal level. This phenomenon is due to the inherent decay mechanism of the land surface system including multiple soil horizons and underlying aquifers (Lu 2009; Lu et al. 2014). Hence, to take the time lag effect into account, we developed the antecedent DFPA of a certain day as follows (Li et al. 2020):

$$\text{DFPA}'_t = \sum_{n=0}^{N} \beta^n \times \text{DFPA}_n, \tag{5}$$

where N is the number of the earlier days used in the calculation, which can be determined from the decay constant parameter β and the precision of the truncation ε with the following formulation:

$$\beta^N = \varepsilon, \tag{6}$$

where ε indicates the contribution fraction of the DFPA on the last day. In this study, ε is taken as 1%, and β is set to 0.5. A β close to 1 means taking the average value of the previous N days, while a value closer to 0 indicates a higher weight to the present day. The N is correspondingly determined as 50 (days), which roughly equals the response time of the terrestrial system to coming precipitation in the YRB (H. Wang et al. 2015). The sensitivity analysis for different combinations of β and ε (N) is provided in the discussion, suggesting the selected group of $\beta = 0.5$ and $\varepsilon = 0.01$ ($N = 50$) is favorably robust in the YRB. Then, we used the breaks for additive seasonal and trend (BFAST) algorithm to remove the secular trend and seasonal components in DFPA'_t to avoid the obscurity of short-term extremes (Jong et al. 2012). The BFAST algorithm requires inputting the whole time series of DFPA'_t, which was derived from the TWSA reconstructions, precipitation measurements, and ET simulations beforehand. It adopts an addictive decomposition model to input series as follows:

$$\text{DFPA}'_t = \text{DFPA}'_{t_T} + \text{DFPA}'_{t_S} + \text{DFPA}'_{t_R}, \tag{7}$$

where DFPA'_{t_T} and DFPA'_{t_S} are the trend and seasonal components of the original DFPA'_t, and DFPA'_{t_R} denotes the remainder detrended and deseasonalized component. The detrended and deseasonalized DFPA'_t has passed the augmented Dickey–Fuller test at a confidence level of 0.01, showing the stationarity of the partitioned time series. Moreover, the insignificant linear regression ($p = 0.99$, $R^2 = 0$) fitting results during the whole period 1961–2015 also indicate the removal of secular trends in DFPA'_t. We also compared the BFAST results with the seasonal-trend decomposition based on LOESS (locally estimated scatterplot smoothing; STL) method, and a significant correlation (CC = 0.85, $p < 0.01$) was discovered, presenting the robustness of our time series decomposition results. Finally, to standardize detrended and deseasonalized DFPA'_t, three probability distributions including normal, generalized extreme value (GEV), and logistic distributions were tested because they can be used with the negative value of input time series. Multiple metrics such as Akaike information criterion (AIC), Bayesian information criterion (BIC), RMSE, and NSE were used for screening the best distribution (Li et al. 2021). AIC and BIC can efficiently represent the goodness-of-fit of different distributions, and RMSE and NSE can well reflect the error between empirical and theoretical frequency. The lower are the AIC, BIC, and RMSE, the better is the performance of the fitted distribution. The NSE closer to 1 means a higher accuracy of the distribution. In addition, Kolmogorov–Smirnov (KS) test and chi-squared (CS) test were used to evaluable the significance of the potential probability distribution. A lower p value indicates a more reliable probability distribution (Li et al. 2020). Based on the best performing probability distribution (i.e., logistic distribution, Table 1), the SDFPI can be derived as follows:

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\text{SDFPI}} \exp(-x^2/2)dx = \int_{-\infty}^{\text{DFPA}} f(x)dx, \tag{8}$$

where $f(x)$ represents the selected probability distribution function. For assessment of the monitoring skill of SDFPI, other widely used indices, including TSDI, self-calibrated Palmer drought severity index (sc-PDSI), and multiscale (1, 3, 6, 9, 12 months) standardized precipitation evapotranspiration index (SPEI), were analyzed. These metrics have demonstrated robust capability in flood and drought monitoring (W. Wang et al. 2015), thus a strong correlation with them can indicate the robustness in SDFPI as well, though at monthly scales. TSDI was computed using the reconstructed TWSA series following Nie et al. (2018), and SPEI and sc-PDSI were provided by the Climatic Research Unit of the University of East Anglia (Vicente-Serrano et al. 2010a,b).

TABLE 1. The goodness of fit and test results of marginal distributions of grid-based DFPA$'_t$ in the YRB. The metrics indicating the best goodness of fit are shown in bold fonts.

Probability distribution	AIC ($\times10^4$)	BIC ($\times10^4$)	RMSE (mm)	NSE	p value (KS)	p value (CS)
Generalized extreme value	18.46	18.47	0.02	0.99	0.00	0.00
Logistic	**18.41**	**18.41**	**0.01**	**1.00**	0.02	0.00
Normal	18.43	18.43	**0.01**	**1.00**	0.00	0.00

We characterized the start, end, maximum and average intensity, and affected areas of short-term floods and droughts in the YRB between 1961 and 2015 using developed SDFPI. The onset of a drought/flood event is defined as the first day for which SDFPI is higher/lower than the specific threshold, and the cessation is identified as the first day SDFPI returns to lower/higher than the threshold. Thus, the duration can be determined as the difference between the onset and cessation of a drought or flood. The droughts and floods with five different categories each (i.e., exceptional, extreme, severe, moderate, and abnormal) were classified using SDFPI based on the U.S. Drought Monitor (USDM) (Kuwayama et al. 2019). The maximum and average intensity was taken as the maximum and mean SDFPI value during the event. We estimated the percentage area affected by droughts/floods by calculating the ratio of grids with SDFPI lower/higher than the certain threshold to the total of 182 grid cells in the YRB. Therefore, the overall severity score was estimated as the multiplication of duration (days), average intensity, and the percentage of the affected area as described in Shah and Mishra (2020).

5. Results

a. Accuracy of daily GRACE data

The accuracy of daily GRACE data is crucial to the reconstruction of TWSA and the construction of SDFPI. Therefore, we compared the terrestrial water storage changes (ΔS) derived from the ITSG-Grace2018 solution, VIC model, and water balance estimates at both daily and monthly scales in the upper, middle, and lower reaches of the YRB between 2003 and 2015. For the upstream of the YRB, daily GRACE data presents reasonably good consistency with VIC modeling results with the CC of 0.47 ($p < 0.01$), which is relatively poorer for water balance estimates (CC = 0.39). Moreover, comparably better performances of the GRACE data are revealed in the midstream and downstream parts of the basin (see Fig. 2). Figure 3 demonstrates that monthly changes in TWS from the ITSG-Grace2018 product show better performance than daily results, which is reflected by higher correlations with both VIC simulations (CC = 0.58–0.81, $p < 0.01$) and water balance estimations (CC = 0.34–0.53, $p < 0.01$) in the different regions of the YRB. Differences between GRACE and VIC output indicate the uncertainty in daily GRACE solutions, which can be attributed to the multiple postprocessing techniques such as Kalman filtering and spherical harmonic truncation. In addition, the comparably weak correlation with water balance estimates further reveals the inherent bias of the daily GRACE solution, even though the consistency is stronger at the monthly scale. However, we note the incomplete representation of the TWS components in the VIC model (e.g., deep groundwater) can also contribute to the differences with GRACE data. Moreover, the propagation of uncertainties in precipitation, discharge, and VIC-modeled evapotranspiration may also contribute to the discrepancies between GRACE and water balance estimates. Generally, the ITSG-Grace2018 product presents fairly good accuracy by comparing with the VIC model and water balance results, highlighting the potential of establishing SDFPI using reconstructed TWSA.

b. Performance of the RF model

The RF model was used to reconstruct daily TWSA in the YRB between 1961 and 2015. Figure 4 shows original GRACE TWSA and RF-simulated results from 2003 to 2015. It is discovered that RF-simulated TWSA is underestimated in the dry season (November–April), especially during the test period (2011–15) and overestimated in the wet season (May–October) from 2003 to 2005. Such differences can be derived from the uncertainty in the RF model and the bias in VIC-simulated TWSA (Miao and Wang 2020), which can subsequently propagate to the reconstructed TWSA and the resulting SDFPI. The satisfactory predictive skill during the training period, while the model performance in the test period is relatively poorer because of the underestimation of TWSA from 2013 to 2015, resulting in the CC, NSE, and RMSE of 0.88, 0.76, and 23.95 mm, respectively. Further, Fig. S1 in the online supplemental material illustrates the probability density function of various metrics of the RF model in the test period at the grid scale. It can be observed that 60% of grid cells have CCs higher than the average value (0.74), and the percentage is 57% (38%) for the areas having NSE/RMSE higher (lower) than the mean value [0.54 (46.34) mm], respectively.

c. Development of SDFPI

Based on the reconstructed TWSA, observed precipitation, and simulated evapotranspiration, the antecedent DFPA was derived with the parameters $\beta = 0.95$ and $N = 90$. Furthermore, the SDFPI was developed by fitting the antecedent DFPA using an appropriate probability distribution. Figure 5 shows the fitting results of three selected probability distributions including the GEV, logistic, and normal distributions. It is clearly seen that all the candidate probability distributions can reasonably well fit the empirical probability of the antecedent DFPA (see Table 1). This is evident by p values lower than 0.05 in KS and CS test results of three optional distributions. However, multiple metrics simultaneously indicate that the logistic probability distribution is most suitable to standardize the antecedent DFPA because of the lowest AIC, BIC, and RMSE together with the highest NSE. Therefore, it

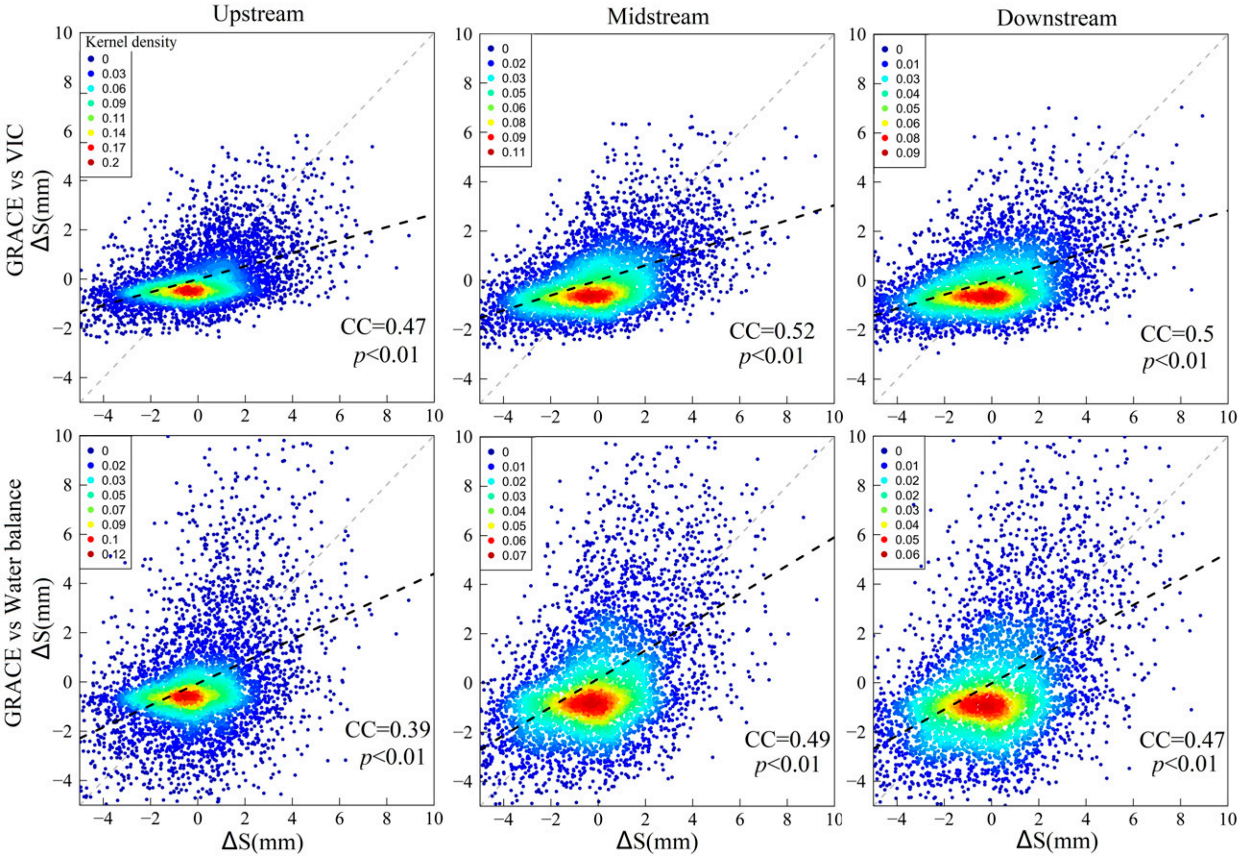

FIG. 2. Scatterplots between daily Δ*S* from the ITSG-Grace2018 product and (top) VIC model and (bottom) water balance estimates in the upstream, midstream, and downstream of the YRB.

is used to develop SDFPI combined with reconstructed TWSA for the YRB. In addition, distribution selection and significance test results on the grid scale are shown in Fig. S2. KS and CS test results indicate that all the grid cells have satisfied the 0.05 significance level, while other indices including AIC, BIC, RMSE, and NSE together imply that the logistic probability distribution is relatively more appropriate to construct the SDFPI spatially.

Table 2 lists the categories and corresponding percentiles of SDFPI for different classes of droughts and floods. Exceptional flood/drought events will be monitored when the SDFPI is higher (lower) than 2.05 (−2.05), the threshold values are ±1.64, ±1.28, ±0.84, and ±0.52 for the extreme, severe, moderate, and abnormal drought or flood events, respectively. Similarly, the definitions and classifications of multiple indices (i.e., TDSI, scPDSI, and SPEI) have been summarized for comparison (see Table S1).

Figure 6 illustrates the daily SDFPI series in the YRB during the period 1961–2015. The SDFPI roughly fluctuates between −3 and 3 with apparent seasonal characteristics with high values in the wet season and low values in the dry season. The YRB witnessed alternative droughts and floods over several decades according to SDFPI. For example, the 1998 basinwide catastrophic flood from July to September with a return period of 100 years (in terms of flooding water volume over; Wang and Plate 2002) is detected by SDFPI. This flood event inundated approximately 210 000 km^2 of land area and caused an economic loss of more than 120 billion yuan and at least 3000 deaths. Another two severe floods lasting from July to August in 2010 (Sun et al. 2017) and 1983 (Liu et al. 2020) over the midstream YRB are also well observed. An abrupt increase in SDFPI during the wet season of 2015 is also witnessed, consistent with the precipitation-induced flood in the lower reach of the YRB. We note several mild floods occurred during the wet season of 2003 over the upper and middle reaches of the YRB due to continuous precipitation extremes are detected by SDFPI (Jiang et al. 2006), indicating the capacity of SDFPI to assess antecedent precipitation influences. Alternatively, two well-documented severe large-scale droughts over several decades in south China are confirmed by SDFPI, including the compound drought and heat wave attacking the upper reach of the YRB in the 2006 summer (Li et al. 2021) and the 2011 drought striking the whole of south China in the spring and summer (Long et al. 2014). The 2011 drought left more than 1300 lakes dried up in the YRB, which threatened the water security of 4.2 million people (CMA 2012). An extreme drought during the wet season of 1978 over the middle reach of the YRB is also reported by

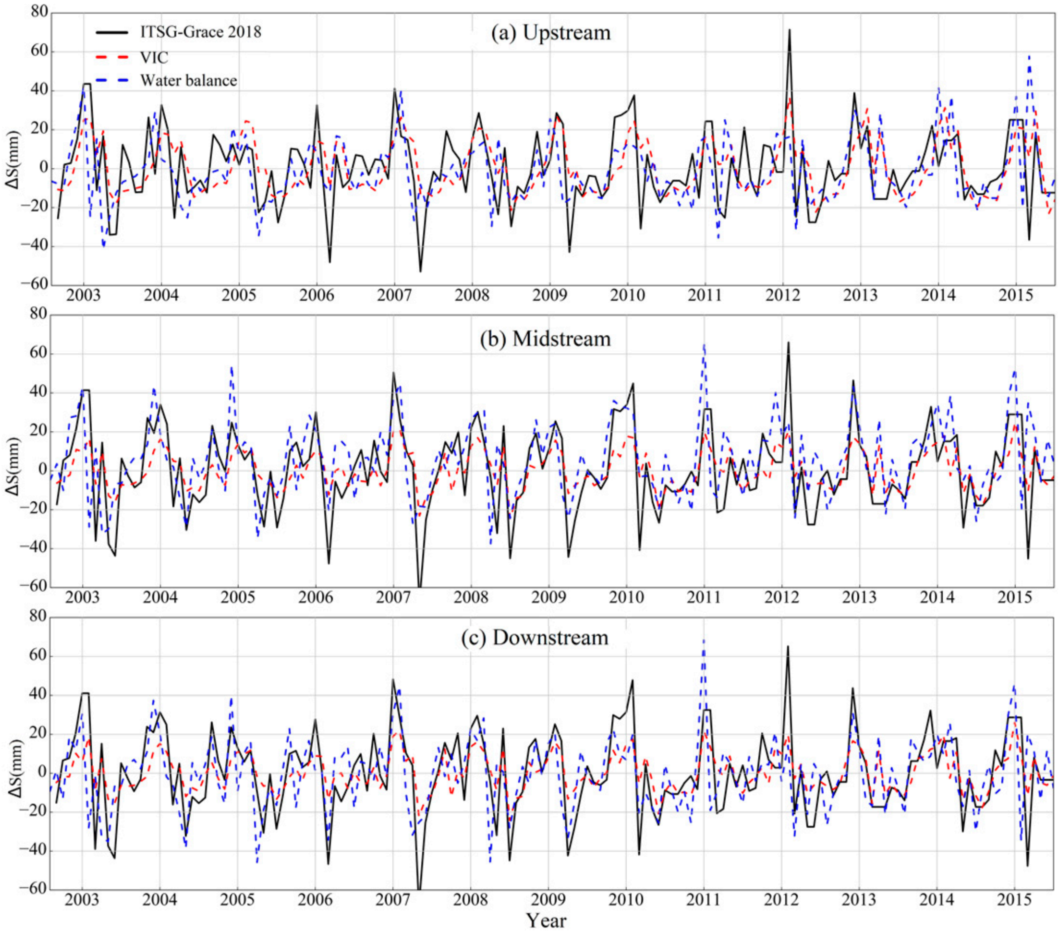

FIG. 3. Monthly time series of TWSA derived from ITSG-Grace2018, VIC, and water balance estimates in the (a) upstream, (b) midstream, and (c) downstream of the YRB during the period 2003–15.

SDFPI (Chai et al. 2019). In addition, a flash drought in the upper and middle reaches of the YRB during the summer of 2013 is apparently evidenced, which damaged more than 2 million hectares of crops over south China (mostly in the YRB) (Yuan et al. 2015), highlighting high effectiveness of SDFPI in capturing the short-term droughts.

d. *Evaluation of SDFPI*

To evaluate the effectiveness of SDFPI in flood and drought monitoring, we used several widely used monthly indices consisting of TSDI, scPDSI, and 1-, 3-, 6-, 9 , and 12-month SPEI for intercomparison and evaluation of the effectiveness of SDFPI in flood and drought monitoring. The monthly SDFPI was computed by averaging the daily values in each month. Figure 7 presents the monthly time series of SDFPI, TSDI, scPDSI, and the change range of multiscale SPEI. SDFPI shows a relatively higher correlation with TSDI (CC = 0.78, $p < 0.01$) than scPDSI (CC = 0.63, $p < 0.01$). In addition, 6-month SPEI shows the most significant correlation with SDFPI with the CC of 0.65 among those of multiscale SPEI, while the 1-month SPEI shows the lowest CC (0.32) with SDFPI. This correlation hints at the effects of antecedent climate factors and TWSA on the current status of land systems. The spatial distribution of CC between SDFPI and these indices over the YRB is depicted in Fig. S3. TSDI presents the strongest correlation with SDFPI among multiple indices, with the CC generally higher than 0.7 in the middle and lower reaches of the YRB and relatively lower in the upper reach, which might arise from the data uncertainty due to the scarce weather stations in the upstream YRB. Similarly, scPDSI demonstrates relatively higher CC (0.5–0.8) in the middle and lower catchments of the YRB than in the upstream basin. Such spatial patterns are detected in 6- and 9-month SPEI, but the 1-, 3-, and 12-month SPEI show the CC lower than 0.5 over the majority of the YRB. Generally, SDFPI presents reasonably good consistency with these popular indicators (e.g., TSDI, scPDSI, and SPEI) temporally and spatially, highlighting the potential of SDFPI for floods and droughts monitoring at the monthly time scale. However, these widely used indices are not able to illustrate the submonthly evolution of floods and droughts, showing the potential outperformance of our proposed SDFPI. For example, several basinwide floods in 1983, 1998, and 2015 are definitely detected, and some large-scale droughts in 1978, 2006, and 2011 are also seen by these indices.

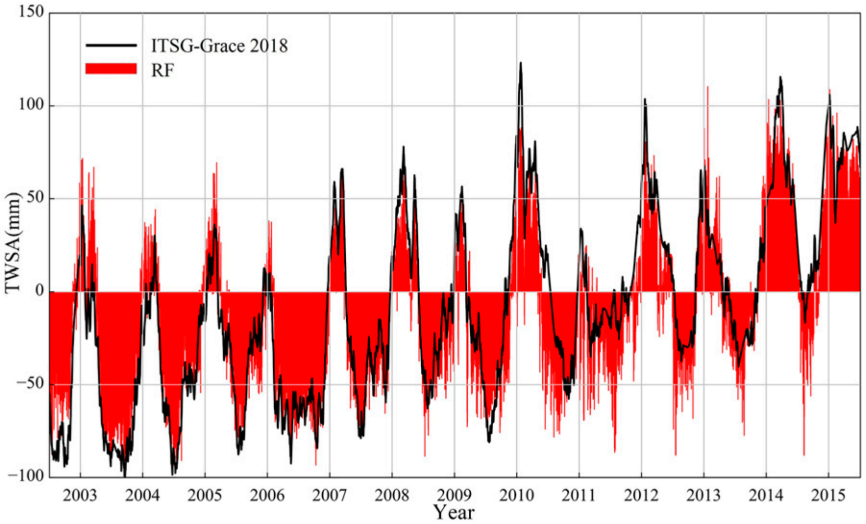

FIG. 4. Daily TWSA series from the ITSG-Grace2018 solution and the RF model in the YRB from 2003 to 2015.

However, both TSDI and SPEI largely underestimate, for example, the flash drought in the summer of 2013 compared with SDFPI due to relatively coarse temporal resolution (e.g., monthly). While the event is observed by scPDSI, it cannot provide the evolutions at a finer time scale. Moreover, the mild flood in the wet season of 2003 triggered by successive precipitation is not detected by TSDI, scPDSI, and SPEI. It may be because they neglect the antecedent condition of TWSA saturation. Therefore, these two typical short-term events with relatively shorter duration and less intensity are further studied as case studies to validate the capability of SDFPI in capturing their short-term dynamics and antecedent status by comparing with other indices (i.e., TSDI, scPDSI, and 6-month SPEI), VIC-modeled soil moisture and runoff, and in situ discharge.

Figure 8a shows the daily evolution of SDFPI and alternative hydrological fluxes during the flood process in June 2003. The continuous net precipitation gains from 19 to 30 June lead to the maximum TWSA of 19.5 mm, meaning the saturation of the land system due to the replenishment from increasing net precipitation. However, both precipitation and runoff peaks on 25 June with the highest of SDFPI (1.96) on the next day, the obvious delay between the actual flood process and land system indicate the potentially mistaken forecast of the onset of floods due to the delayed response of TWSA to net precipitation. Then, the decreasing net precipitation from

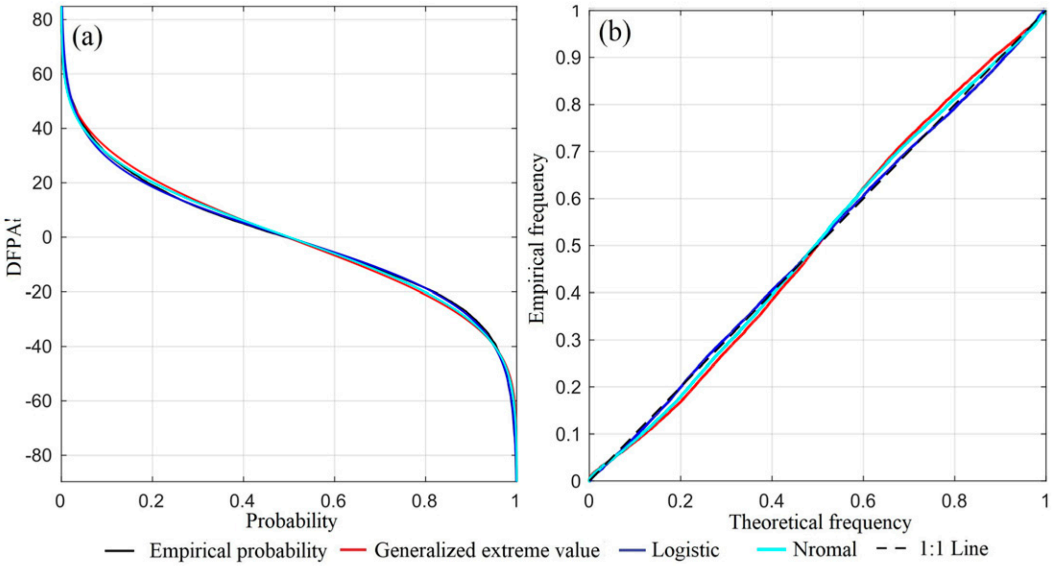

FIG. 5. (a) Fitted empirical and theoretical probabilities of marginal distribution for DFPA$'_t$ and (b) P–P plots of different theoretical distributions.

TABLE 2. Flood and drought categories of SDFPI. The percentiles for positive and negative SDFPI were calculated independently.

SDFPI	Category	Description	Percentile
[2.05, ∞)	W5	Exceptional flood	2%
[1.64, 2.05)	W4	Extreme flood	5%
[1.28, 1.64)	W3	Severe flood	10%
[0.84, 1.28)	W2	Moderate flood	20%
[0, 0.52)	W1	Abnormally wet	30%
(−0.52, 0]	D1	Abnormally dry	30%
(−1.28, −0.84]	D2	Moderate drought	20%
(−1.64, −1.28]	D3	Severe drought	10%
(−2.05, −1.64]	D4	Extreme drought	5%
(−∞,−2.05]	D5	Exceptional drought	2%

26 to 28 June also results in the rapid decline of SDFPI, followed by a consecutive recovery until the end of June (Fig. 8). Changes in SDFPI present a significant correlation (CC = 0.69, $p < 0.05$) with runoff from the VIC model, indicating the effective monitoring ability for the short-term floods of SDFPI. We further illustrate the spatial evolution of SDFPI, VIC runoff anomalies, and standardized in situ discharge during the flood (Fig. 9). SDFPI suggests that the flood originated from the upstream YRB on 16 June and gradually developed on 22 June. After peaking across the midstream and downstream basins on 26 June, the flood slowly alleviated until 4t July. In addition to the short-term flood, the consistent wetness was observed in the lower reach of the YRB, which might be caused by the river routing and regional precipitation. Similar spatial patterns are also detected by VIC-simulated runoff, which illustrated comprehensive runoff deficits over the middle and lower reaches on 16 June followed by obvious runoff gains in the northern and southwest parts of the YRB. The positive runoff subsequently spread over the downstream with the maximum runoff anomaly of 46.07 mm on 26 June. Afterward, the negative runoff anomalies appeared again from 29 June to 4 July in the middle and lower basins, implying the cessation of the flood (Fig. 9). Similarly, we discovered clear change patterns during the flood according to in situ discharge collected at the outlets of the 10 main subbasins, which correspondingly changes from −0.42 to 7.11, especially in the middle region (i.e., Wujiang River basin). We note a persistent flood that occurred in the upper reach of the YRB was also evident in June according to SDFPI, which can be derived from the consistently positive runoff anomalies and in situ discharge. The spatial distributions of these variables and other widely used indices (e.g., TSDI, scPDSI, and SPEI) in August 2003 are shown in Fig. S4. All the indexes suggest the high wetness in the upper reach of the YRB, which is also witnessed by observed discharge. The consistency among different metrics at the monthly scale suggest the monitoring ability of SDFPI for larger flood events.

Alternatively, Fig. 8b presents the daily series of SDFPI and related hydrological variables during the 2013 flash drought. The abrupt decrease in net precipitation that occurred from 11 to 18 June rapidly caused the water deficit of the land system. The decreasing soil moisture quickly responded to the depleting water supply, which showed a time lag of 1–2 days than TWSA mainly due to the imperfect representation of the soil layer depth in the VIC model (e.g., unsaturated zone and groundwater aquifers). Consequently, SDFPI indicates the peak of drought on 19 June followed by a gradual rebound from 23 to 27 June due to continuous gains in net precipitation. Although soil moisture began to decrease on 27 June caused by the reduction in net precipitation, TWSA and SDFPI still kept increasing, though at slower rates, due to the delayed response of groundwater percolation, further indicating the importance to account for the antecedent information of the land system. The spatial evolution in Fig. 10 has depicted

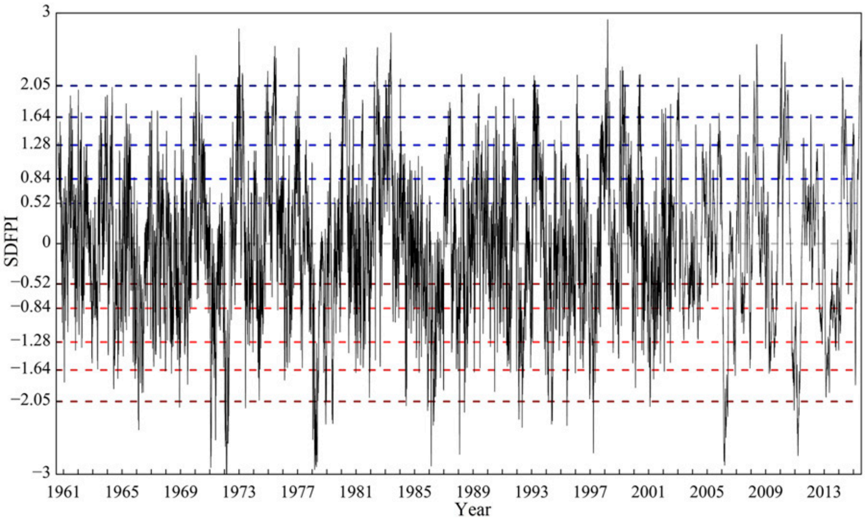

FIG. 6. Daily SDFPI series and the thresholds for different categories of droughts and floods in the YRB between 1961 and 2015. The dashed line represents the thresholds for different classes of droughts and floods, referring to Table 2.

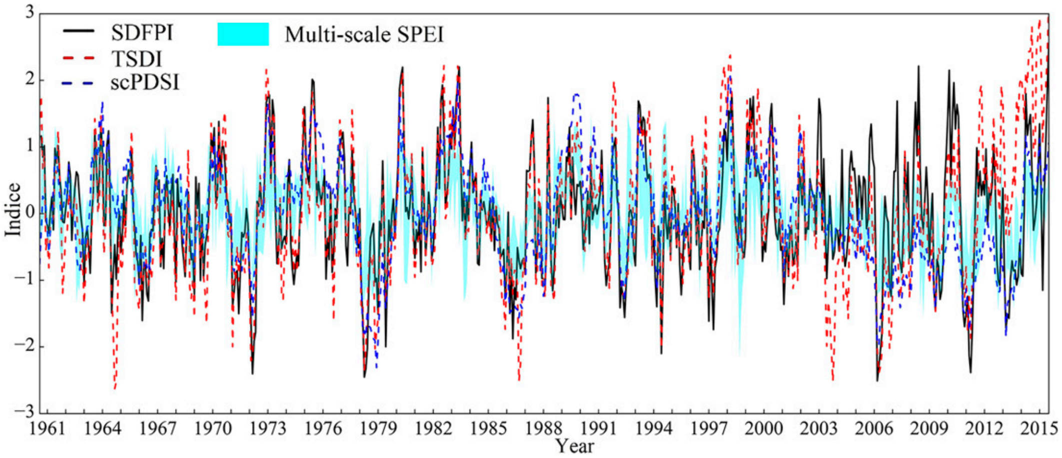

FIG. 7. Monthly series of SDFPI, TSDI, scPDSI, and change range of 1-, 3-, 6-, 9-, and 12-month SPEI in the YRB between 1961 and 2015.

the detailed process of onset, evolution, propagation, and cessation of the drought event. Although almost the whole basin shows the positive SDFPI on 10 June, the southwest of the middle basin rapidly shows a drying tendency with SDFPI ranging from −2 to −1 locally on 14 June (Fig. 10). Further, the drought peaked in the southern midstream and nearly propagated to the upper reach on 19 June. Slowly rebounded SDFPI were discovered over the whole YRB until 1 July, with some clear positive values in the western middle basin. Similar patterns are well compared with soil moisture anomalies from the VIC model and observed discharge from different gauge stations, especially for the short-term drought event in the

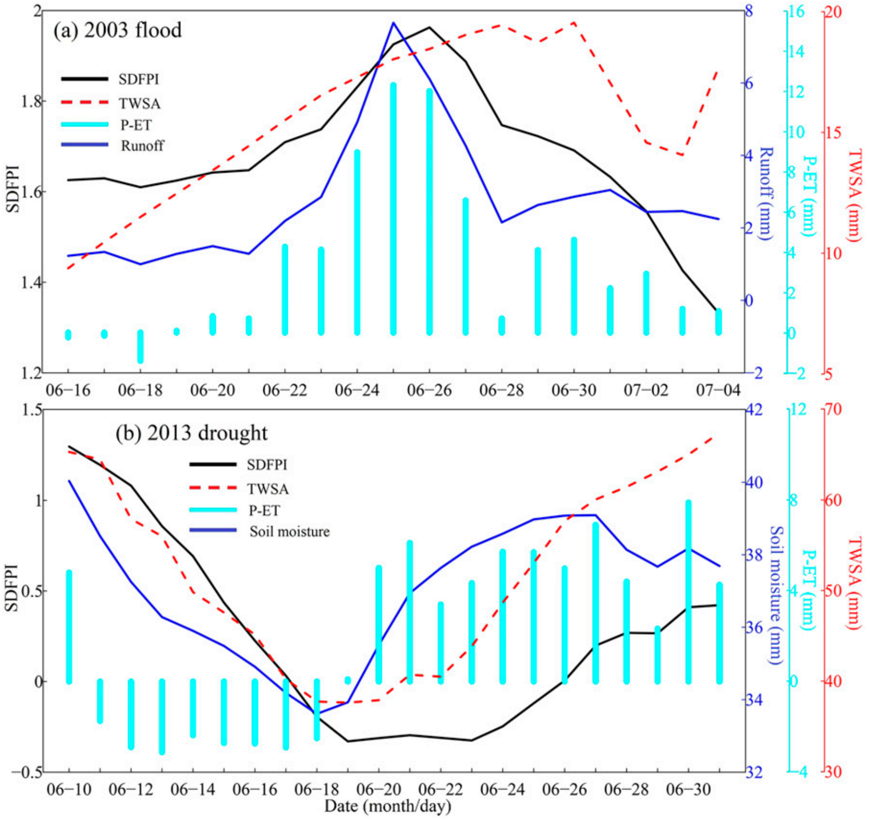

FIG. 8. (a) Daily time series of SDFPI, VIC runoff, net precipitation (P − ET), and TWSA during the 2003 flood. (b) Daily series of SDFPI, VIC soil moisture, net precipitation (P − ET), and TWSA during the 2013 drought.

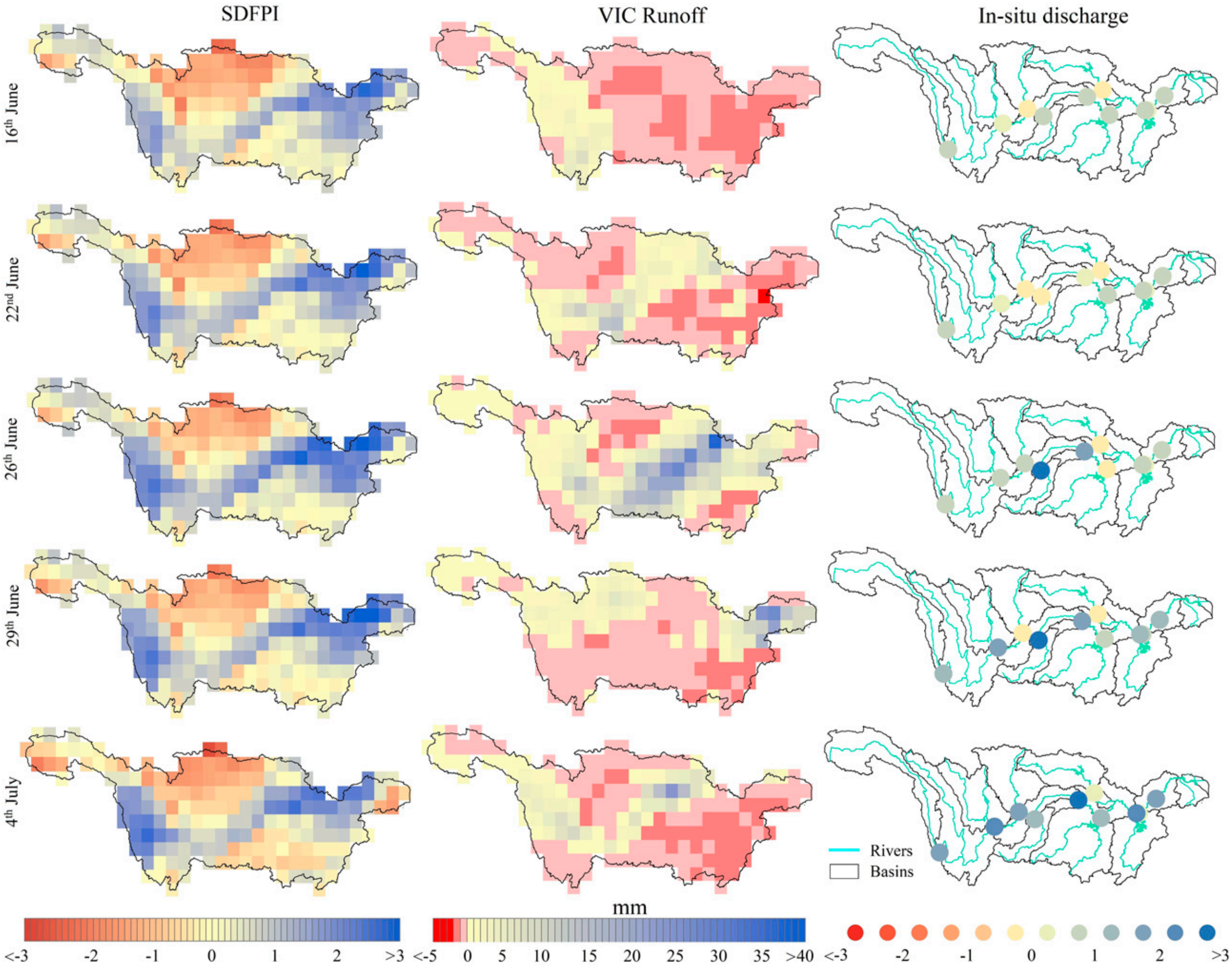

FIG. 9. Spatial evolution of (left) SDFPI, (center) VIC runoff, and (right) in situ discharge over the YRB during the 2003 flood. Note that the VIC runoff represents the anomaly with respect to the average value during 1961–2015. The in situ discharge has been standardized for better comparison. Since the fixed time intervals (e.g., pentad scale) may conceal the actual flood evolution information, the presented dates are manually selected after confirming the various phases (e.g., flood onset, peak, etc.).

south of the midstream YRB with the local flood event that occurred in the western parts on 1 July (see Fig. 10). As shown in Fig. S5, the spatial distribution of SDFPI in June 2013 compares reasonably well with VIC-based soil moisture anomalies and in situ discharge except for some difference in the upper and lower reaches due to distinctive calculation methods and data used, highlighting the potential of SDFPI for the multiscale drought monitoring.

e. *Application of SDFPI*

Given the effectiveness in short-term floods and droughts monitoring by SDFPI as revealed by two short-term events (discussed in the previous section), we applied it to monitor and characterize all the floods (SDFPI > 2.05) and droughts (SDFPI < -2.05) that belong to the "exceptional" class in the YRB during the period 1961–2015 (see Tables 2 and 3). We note only the submonthly events that have durations ranging from one week to one month are considered to focus on more intensive disasters and avoid any false positive/negative extremes. A total of 10 droughts and 12 floods have been reported, of which the severest drought (ID: 2) and flood (ID: 10) events are independently analyzed. The drought event lasting for 19 days began on 20 June and ended on 8 August 1971, affecting 6.33% of the total area in the YRB. The max and mean intensity of the drought were -2.91 and -2.59, respectively (Table 3). In this case, the overall severity score reached a value of 3.12. Alternatively, the submonthly exceptional flood (ID: 10) lasted for 19 days, starting on 28 November and ending on 16 December 1982, and the max and average intensity reached 2.55 and 2.31, respectively (Table 3). Spatially, 16.05% of the area in the YRB was affected by the flooding. Hence, the overall severity score reached 7.06. Moreover, most droughts with longer durations generally have higher intensity and larger affected areas (e.g., ID: 1, 2, 13), whereas there are exceptions with long durations that affected a small proportion of areas (e.g., ID: 17), and vice versa (e.g., ID: 8). Similar patterns are also discovered based on SDFPI-monitored floods, indicating the difficulty to monitor such

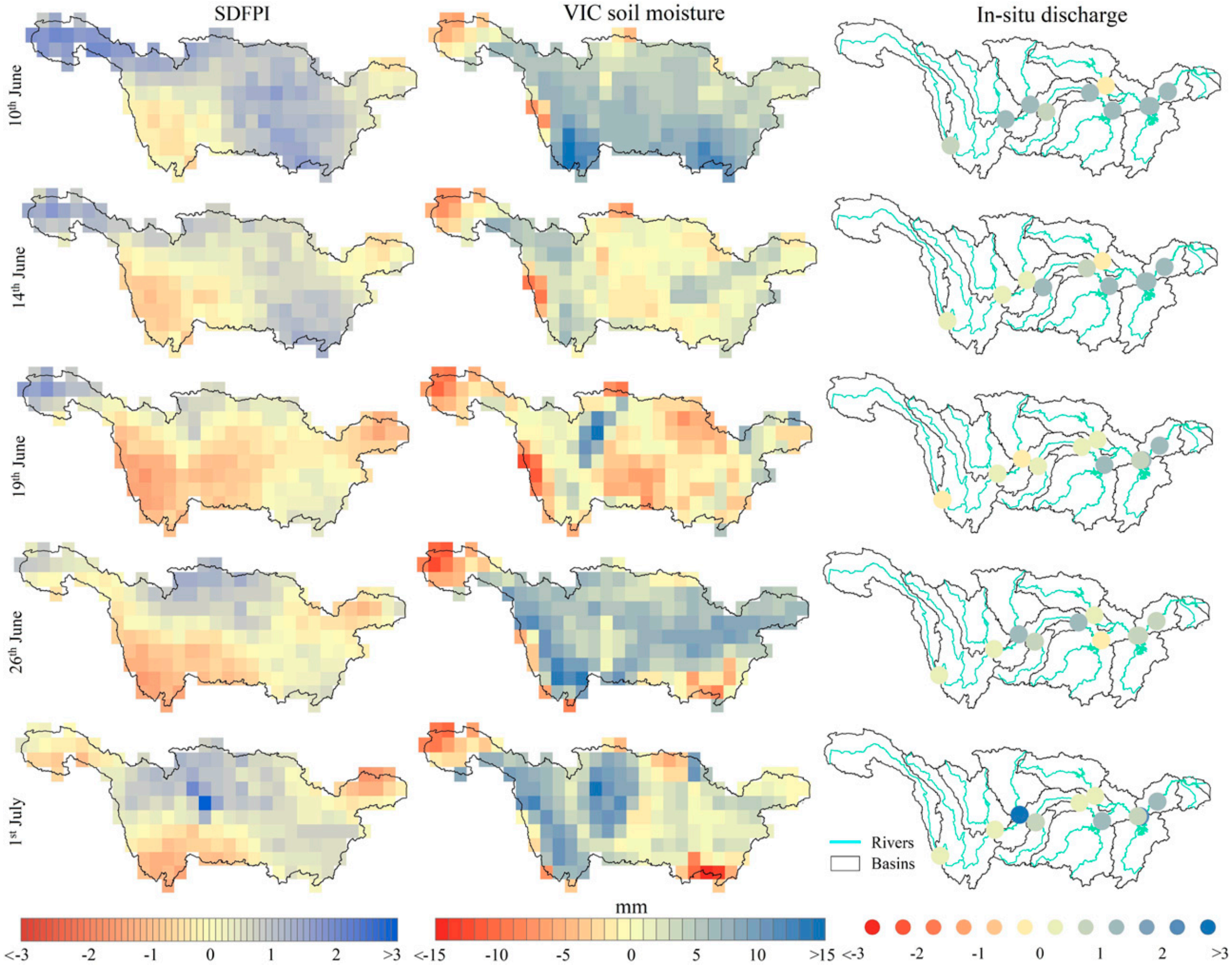

FIG. 10. Spatial evolution of (left) SDFPI, (center) VIC soil moisture, and (right) in situ discharge over the YRB during the 2013 drought. Note that the VIC soil moisture represents the anomaly with respect to the average value during 1961–2015. The in situ discharge has been standardized for better comparison. Selection of dates has been made manually similar to those in Fig. 9.

short-term events in the data-sparse regions (e.g., upper YRB). Furthermore, Fig. 11 illustrates the temporal changes of SDFPI during these submonthly events. It can be seen that most drought and flood events with one peak develop to the maximum SDFPI within two pentads, posing a great challenge for the prediction and warning based on conventional weekly or monthly indices. In addition, some events have multiple peaks within a month (e.g., IDs: 2, 4, 7), implying that the forecasters may mistakenly identify the transient drop in drought/flood intensity as the actual cessation of the event, underscoring the significance and importance of SDFPI. Apart from the exceptional extremes, the monitored droughts and floods belonging to the "extreme" and "severe" classes have been summarized in Tables S2 and S3, among which 56 and 95 extreme and severe events are identified using SDFPI, respectively. To sum up, daily SDFPI successfully monitored and characterized short-term floods and droughts with various intensities.

Last, to get insights into the hotspots of short-term extreme events in the YRB, we analyzed the spatial distribution of the number of short-term exceptional floods and droughts during 1961–2015 (Fig. 12). The hotspots of the flood occurrences are mainly located in the center of midstream YRB (i.e., Wujiang and Dongting Lake basins), including more than 15 floods over there. Another hotspot is in the lower reaches of the YRB (i.e., Yangtze delta), which is one of the most flood-prone regions of China with an expanded population and rapid urbanization. In terms of drought events, the upper (i.e., Jinsha and Yalong River basins) and middle (i.e., Dongting and Poyang Lake basins) reaches of the YRB experienced the most frequent drought events of more than 10 times during 1961–2015. Parts of the north of middle reach are also featured by frequent droughts with approximately 15 events identified. Overall, SDFPI illustrates efficient monitoring and characterization ability for submonthly flood and drought events in the YRB.

TABLE 3. Floods and droughts belonging to the exceptional class in the YRB between 1961 and 2015 monitored by SDFPI. The most severe flood and drought events are shown in bold fonts.

ID	Type	Start	End	Duration (days)	Max intensity	Mean intensity	Affected area (%)	Overall score
1	Drought	10 Aug 1966	24 Aug 1966	7	−2.42	−2.27	6.83	1.09
2	**Drought**	**20 Jul 1971**	**8 Aug 1971**	**19**	**−2.91**	**−2.59**	**6.33**	**3.12**
3	Drought	9 Jul 1972	24 Jul 1972	15	−2.64	−2.33	6.37	2.23
4	Drought	25 Sep 1972	16 Oct 1972	13	−2.60	−2.28	9.38	2.78
5	Flood	20 Jun 1973	17 Jul 1973	17	2.80	2.39	9.28	3.77
6	Flood	16 Sep 1973	30 Sep 1973	12	2.22	2.16	13.05	3.38
7	Flood	11 Nov 1975	22 Dec 1975	26	2.57	2.24	7.35	4.29
8	Drought	25 May 1979	5 Jun 1979	9	−2.35	−2.20	7.81	1.55
9	Drought	5 Nov 1979	29 Nov 1979	15	−2.34	−2.23	6.19	2.07
10	**Flood**	**28 Nov 1982**	**16 Dec 1982**	**19**	**2.55**	**2.31**	**16.05**	**7.06**
11	Drought	3 Oct 1986	25 Oct 1986	11	−2.36	−2.18	10.89	2.61
12	Drought	19 Jul 1988	25 Jul 1988	7	−2.74	−2.35	4.32	0.71
13	Drought	3 Nov 1994	25 Nov 1994	20	−2.35	−2.20	5.27	2.32
14	Flood	27 Aug 1998	14 Sep 1998	16	2.91	2.52	16.00	6.45
15	Flood	18 Sep 1999	24 Sep 1999	7	2.25	2.16	14.13	2.13
16	Flood	29 Oct 2000	4 Nov 2000	7	2.20	2.12	9.03	1.34
17	Drought	2 Nov 2006	18 Nov 2006	17	−2.25	−2.14	2.65	0.96
18	Flood	8 Sep 2007	17 Sep 2007	10	2.20	2.12	5.66	1.20
19	Flood	4 Nov 2008	24 Nov 2008	21	2.59	2.36	10.83	5.36
20	Flood	11 Jul 2010	30 Jul 2010	20	2.72	2.46	10.49	5.17
21	Flood	12 Oct 2010	19 Oct 2010	8	2.36	2.26	3.16	0.57
22	Flood	18 Sep 2014	30 Sep 2014	13	2.16	2.11	10.48	2.87

6. Discussion

a. Sensitivity of β and N

Different parameters of β and N will inevitably influence the development of the SDFPI. Thus, there is an urgent need to carry out a sensitivity analysis. We have tested a total of 81 combinations of parameters with $\beta = 0.1, 0.2, 0.3, \ldots, 0.9$ and $N = 10, 20, 30, \ldots, 90$. Figures S6a–c demonstrate the daily SDFPI series for the YRB between 1961 and 2015 derived from the parameter group of $\beta = 0.5$ and varying N. It is

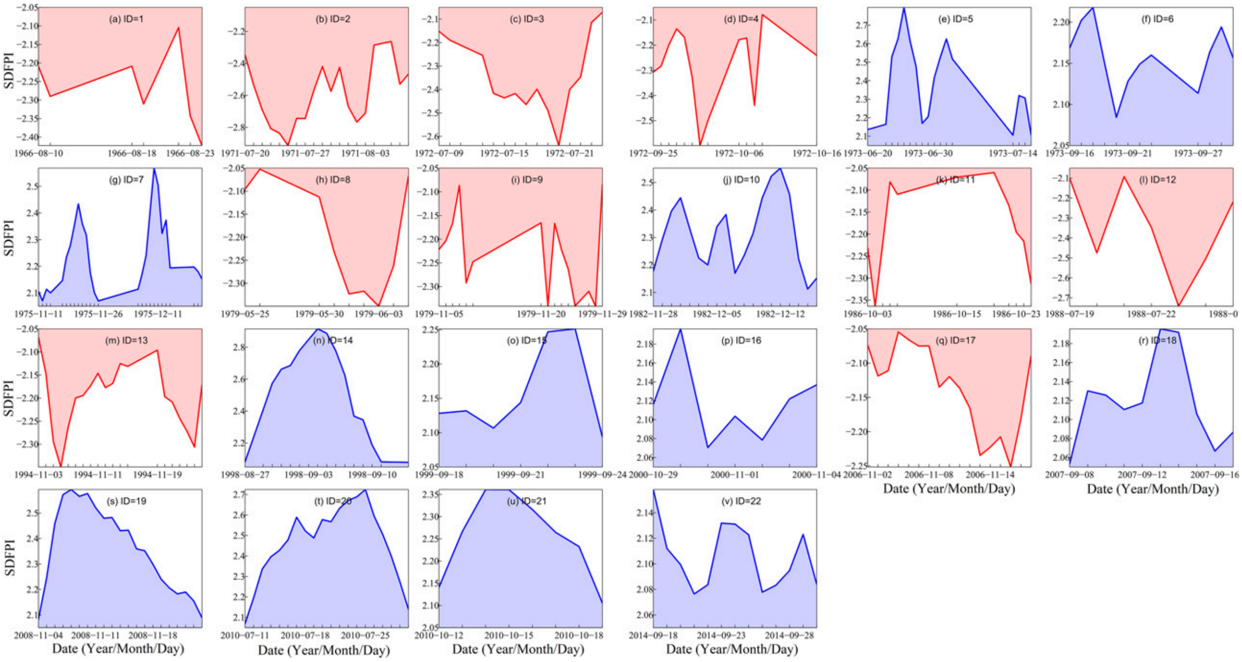

FIG. 11. Temporal evolution of SDFPI over the monitored exceptional floods and droughts. Note that the blue and red shades represent the flood and drought events, respectively. Please refer to Table 3 for details of these events.

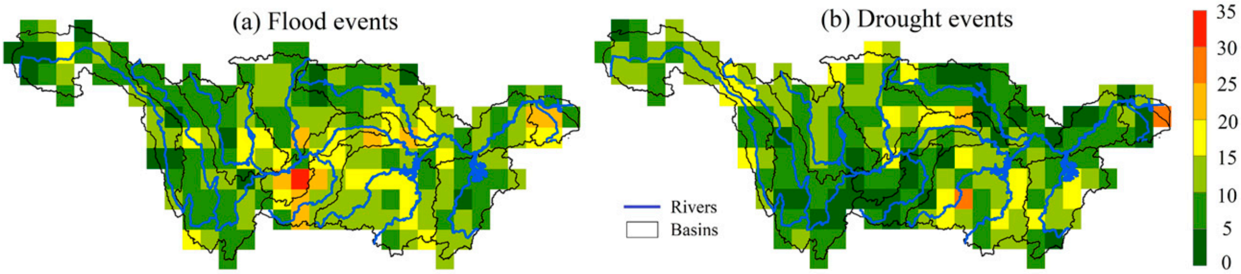

FIG. 12. Spatial distribution of the number of submonthly floods and droughts over the YRB monitored by SDFPI during 1961–2015.

inferred that discrepancies among different SDFPI series could be rarely detected, indicating the influences of different choices of N on SDFPI are relatively small in the YRB. Such patterns are also reflected by the high correlation between the ensemble mean and the standard deviation of different SDFPI series derived from a fixed β of 0.5 and varying N, particularly when N is greater than 30 (see Figs. S7a and S8a). However, they may have substantial impacts on the monitoring skill for short-term floods and droughts over alternative basins, such as those having a faster or slower hydrological responses to natural variability and anthropogenic changes. In terms of the sensitivity of β, an ensemble of the SDFPI series derived from a fixed $N = 50$ and changing β is shown in Figs. S7d–f. Given that a β of 0 indicates a higher weight to the status of the current day, a consistent pattern is discovered indicating a lower β is expected to produce more abrupt estimates during the flood and/or drought events. While a higher value indicates a more gradual SDFPI series as a β of 1 means the direct average of the antecedent information of TWSA. The SDFPI is highly likely to be influenced by the variability of β when it is larger than 0.6 based on the correlation map within SDFPI derived from various β (see Figs. S7b and S8b). Generally, the parameter N is insensitive to the SDFPI, while the parameter β impacts the SDFPI performance relatively more. However, we note different combinations of β and N may fit well in different basins due to distinctive hydrological and climatic conditions, and their sensitivities may also depend on various regions.

b. *Comparisons with previous studies*

Many GRACE-based indicators have been proposed to monitor large-scale floods and droughts, such as the FPI, DPI (Abhishek et al. 2021), the TSDI, and GRACE-DSI, which have been widely validated and used in many catchments worldwide (Abhishek et al. 2021; Reager and Famiglietti 2009; Zhao et al. 2017; Nie et al. 2018). Moreover, some of these studies have assessed the flood and drought monitoring application of GRACE-inferred indexes in the YRB (Sun et al. 2017; Yang et al. 2021). Sun et al. (2017) applied GRACE TWSA and remote sensing precipitation to calculate FPI and identified the exceptional 2010 flood as the most serious disaster during the period 2003–14. In addition, several major floods that occurred in 2010 and 2015 are detected by FPI according to a recent study (Yang et al. 2021). Based on reconstructed GRACE data, Wang and Chen (2021) calculated the FPI and assessed the catastrophic floods in 2010 and 2015. The abovementioned flood events are also detected by our proposed SDFPI (see Fig. 8), highlighting the high effectiveness of flood monitoring for, apart from the added short-term skills, large-scale and for longer duration events. However, previous GRACE-based flood indicators are generated based on monthly GRACE data, which is too infrequent to capture the flooding mechanisms. These indicators are generally calculated using the precipitation at the current month, neglecting the contribution of past precipitation in feeding the land system, especially the soil moisture for the YRB. In this regard, the daily SDFPI is more suitable and robust to monitor the flood events that generally happen within a month, such as the case of the 2003 flood induced by continuous precipitation extremes (discussed in section 5d).

Many papers also used to discuss the drought monitoring ability of different GRACE-based indices. Zhang et al. (2016) initially constructed water storage deficit (WSD) using extended GRACE data to identify drought events in the YRB during 1979–2012, and examinations showed that the severest drought happened during 2003–08 (i.e., the well-known 2006 drought), which is in line with our SDFPI results. This exceptional drought event has also been assessed by Sun et al. (2018) using the WSD-based index derived from original GRACE data from 2003 to 2015. Similarly, Deng et al. (2021) standardized the WSD before using it to assess the spatiotemporal variability of drought events in mainland China in the twenty-first century and obtained similar results. GRACE-DSI was also used to monitor large-scale droughts in the YRB (Cui et al. 2021; Zhang et al. 2021). For instance, the severe drought from 2002 to 2009 was considered the most significant hazard over several years. These studies agree well with the 2006 drought finding based on SDFPI, indicating its reliable drought monitoring skill. However, previously used indicators like WSD and GRACE-DSI neglect the linkage between the atmosphere and land and the decay mechanism of the land system. Such simplifications may impact the drought monitoring skill of these GRACE-based indicators, which is especially true for groundwater droughts that slowly respond to the supplement from precipitation (Wossenyeleh et al. 2021). Additionally, the sensitivity analysis of our proposed SDFPI highlights the importance of the antecedent effects of TWSA on the spatiotemporal evolution and the subsequent monitoring of the short-term events.

c. Implications and limitations

The development of daily-scale SDFPI based on the GRACE satellite is supposed to play an essential role in early warning and prevention of floods and droughts, especially in data-sparse and data-scarce regions globally such as Africa, Southeast Asia, and northwest China (Domeneghetti 2016). Even in developed countries with dense hydrometeorological observation networks, incomplete representation of TWS measurements (e.g., soil moisture and groundwater level) may negatively influence the hydrological application of SDFPI. Therefore, the proposal of SDFPI can contribute a lot to floods and droughts monitoring over alternative basins worldwide, particularly under the commissioning of the GRACE-FO (Kornfeld et al. 2019) and other auxiliary data from the continuously developing hydrological models (Abhishek et al. 2021). In addition, the unprecedented time scale of SDFPI makes it capable to detect short-term hydrological extremes that are characterized by a period of rapid intensification over subseasonal time scales, which have devastating impacts but are notoriously difficult to predict (Christian et al. 2021; Otkin et al. 2021). Alternatively, the ability of SDFPI to detect floods that always occur within a month outperforms previous GRACE indexes like FPI, which can merely be used to detect large-scale floods with long duration and high volumes such as the 2016 and 2020 floods in the YRB (L. Wang et al. 2021).

Although our developed SDFPI shows advanced practical skills in short-term droughts and floods monitoring, some limitations still exist. The SDFPI is developed using RF-reconstructed TWSA, VIC-modeled evapotranspiration, and observed precipitation, which certainly suffer from propagated uncertainty from different data sources, particularly in the RF and VIC models. Specifically, the ITSG-Grace2018 solution is truncated to a degree/order of 40, which is less than that (i.e., 60) of the official GRACE products from NASA JPL. Moreover, the daily values for the months of missing data in the original monthly data may tend toward the trend and annual signal of the static background field. Thus, the daily GRACE signal can be attenuated compared to the monthly solutions (Mayer-Gürr et al. 2018). Moreover, the inherent coarse spatial resolution of the original GRACE product (i.e., ~150 000 km^2) prevents it from detecting the flood and drought mechanisms at local scales (Tapley et al. 2004). Somehow the use of gravity satellites to monitor floods and droughts is an afterwork due to the latency arising from the in-orbit measuring, postprocessing, and product releasing. Such temporal delay urges the need to develop near-real-time GRACE products for early warning and management of floods and droughts (Tapley et al. 2019), which we have attempted to start with SDFPI. Furthermore, the machine learning–based SDFPI is expected to work efficiently for floods and droughts induced by the substantial gains or losses in terrestrial water storage, while relatively less robust for those caused by human activities such as dam cracks and groundwater pumping, particularly in the context of multisource uncertainty from the models, parameters, inputs, and their interactions. Indeed, the validation and application of SDFPI are only carried out in the YRB due to the constraints of in situ hydrometeorological measurements. In this regard, more validations for SDFPI are required to investigate its floods and droughts monitoring potential in other regions worldwide with diversified climate and hydrological conditions (Lehner et al. 2006; Birkmann et al. 2022).

7. Conclusions

In this study, we evaluated the accuracy of the ITSG-Grace2018 solution compared with the VIC simulations and water balance estimates in the YRB. The daily GRACE-like TWSA between 1961 and 2015 was reconstructed based on the RF model combined with multiple hydrometeorological forcing. Thereafter, the antecedent SDFPI was developed using reconstructed TWSA, in situ precipitation, and VIC-modeled evapotranspiration. The categories of SDFPI were determined for different classes of floods and droughts based on the USDM criterion. The performance of SDFPI is subsequently compared with widely used indices including TSDI, scPDSI, and multiscale SPEI spatially and temporally. Two typical short-term flood and drought events in 2003 and 2013 are evaluated against the modeled runoff, soil moisture, in situ discharge, and various indices for validation of SDFPI. All the severe submonthly floods and droughts were then monitored and characterized using SDFPI, of which two severest events (one for drought and flood each) were studied in detail. Spatial hotspots of the short-term floods and droughts are identified. The main conclusions of our study are summarized as follows.

1) The ITSG-Grace2018 product presents favorably good accuracy compared with the VIC model and water balance results in different subbasins of the YRB at both daily and monthly scales. The RF model shows reasonably superior performances for the TWSA reconstructions from 2003 to 2015 with the CC of 0.88, the NSE of 0.76, and the RMSE of 23.95 mm during the test period (2011–15). Similarly, satisfactory accuracy according to multiple metrics is discovered at grid scales, indicating the robustness of the RF model in reconstructing daily TWSA. Based on the reconstructed TWSA, the logistic distribution was selected to establish SDFPI in combination with in situ precipitation and VIC-simulated evapotranspiration.
2) The monthly SDFPI is highly correlated to TSDI and scPDSI, with the significant CC ($p < 0.01$) of 0.78 and 0.63, respectively. Six-month SPEI presents the strongest correlation with SDFPI (CC = 0.65, $p < 0.01$) within multiscale SPEI. The spatial distribution of correlations between monthly SDFPI and TSDI shows higher values in the middle and lower regions while lower values in the upper reach of the YRB. Similar patterns were also detected in correlation maps in scPDSI and 6-month SPEI. SDFPI compared well with in situ discharge measurements and VIC simulations at the daily scale during two typical short-term events, including the 2003 flood and 2013 drought.
3) A total of 22 exceptional historical floods and droughts were detected and characterized using SDFPI. The severest submonthly drought in 1971 and flood in 1982 were included. The drought started on 20 June and ended on 8 August 1971,

and the maximum and average intensity reached −2.91 and −2.59, respectively. A total of 6.33% of the area was affected, and the overall severity score reached 3.12. Alternatively, the flood began on 28 November and ended on 16 December 1982, affecting 16.05% of the total area. The maximum and mean intensity of the flood were 2.55 and 2.31, respectively. In this case, the overall severity score was 7.06. The temporal evolution of these short-term events reveals various patterns of hydrological processes during extreme events. The hotspots of the floods and droughts are mainly located in the middle and upper reach of the YRB, respectively.

4) Since the developed SDFPI accounts for the antecedent TWSA variation via the land decay mechanism, it provides advanced modeling skills jointly for the floods and droughts at unprecedented daily scales. SDFPI can provide reasonable responses to the changes in net precipitation, especially after long-term deficits/gains in TWSA. Successful application of SDFPI in the monitoring and characterization of short-term events uncovers the potential a way for early warning of hydrological extremes based on machine learning and multisource data. Moreover, it can effectively be used for informing the policymakers for mitigation strategies or preempt adaptations (when used in conjunction with the projected precipitation and TWSA) in case of various short-term water extreme events.

Acknowledgments. This study was financially supported by the National Key Research and Development Program of China (2021YFC3200305) and the National Natural Science Foundation of China (U20A20317). The numerical calculations in this paper have been done on the supercomputing system in the Supercomputing Center of Wuhan University.

Data availability statement. VIC simulations are available via the science Data Bank platform (https://www.scidb.cn/en/detail?dataSetId=782627214964293632). ITSG-Grace2018 solution is provided by the Graz University of Technology (https://www.tugraz.at/institute/ifg/downloads/). The in situ meteorological and hydrological data are not publicly available due to the restrictive data sharing policies of the ongoing projects.

REFERENCES

Abhishek, and T. Kinouchi, 2021: Synergetic application of GRACE gravity data, global hydrological model, and in-situ observations to quantify water storage dynamics over Peninsular India during 2002–2017. *J. Hydrol.*, **596**, 126069, https://doi.org/10.1016/j.jhydrol.2021.126069.

——, ——, and T. Sayama, 2021: A comprehensive assessment of water storage dynamics and hydroclimatic extremes in the Chao Phraya River Basin during 2002–2020. *J. Hydrol.*, **603**, 126868, https://doi.org/10.1016/j.jhydrol.2021.126868.

Birkmann, J., and Coauthors, 2022: Understanding human vulnerability to climate change: A global perspective on index validation for adaptation planning. *Sci. Total Environ.*, **803**, 150065, https://doi.org/10.1016/j.scitotenv.2021.150065.

Breiman L., 2001: Random forests. *Mach. Learn.*, **45**, 5–32, https://doi.org/10.1023/A:1010933404324.

Chai, Y., Y. Li, Y. Yang, S. X. Li, W. Zhang, J. Q. Ren, and H. B. Xiong, 2019: Water level variation characteristics under the impacts of extreme drought and the operation of the three gorges dam. *Front. Earth Sci.*, **13**, 510–512, https://doi.org/10.1007/s11707-018-0739-3.

Chen, G., and Coauthors, 2018: A machine learning method to estimate PM2.5 concentrations across China with remote sensing, meteorological and land use information. *Sci. Total Environ.*, **636**, 52–60, https://doi.org/10.1016/j.scitotenv.2018.04.251.

Chen, X., J. Jiang, and H. Li, 2018: Drought and flood monitoring of the Liao River basin in Northeast China using extended GRACE data. *Remote Sens.*, **10**, 1168, https://doi.org/10.3390/rs10081168.

Christian, J., J. B. Basara, E. D. Hunt, J. A. Otkin, J. C. Furtado, V. Mishra, X. Xiao, and R. M. Randall, 2021: Global distribution, trends, and drivers of flash drought occurrence. *Nat. Commun.*, **12**, 6330, https://doi.org/10.1038/s41467-021-26692-z.

CMA, 2012: *Yearbook of Meteorological Disasters of China.* China Meteorological Press, 187 pp.

Croteau, M. J., R. S. Nerem, B. D. Loomis, and T. J. Sabaka, 2020: Development of a daily GRACE mascon solution for terrestrial water storage. *J. Geophys. Res. Solid Earth*, **125**, e2019JB018468, https://doi.org/10.1029/2019JB018468.

Cui, L., C. Zhang, C. Yao, Z. Luo, X. Wang, and Q. Li, 2021: Analysis of the influencing factors of drought events based on GRACE data under different climatic conditions: A case study in Mainland China. *Water*, **13**, 2575, https://doi.org/10.3390/w13182575.

Deng, S., S. Liu, and X. Mo, 2021: Assessment and attribution of China's droughts using an integrated drought index derived from GRACE and GRACE-FO data. *J. Hydrol.*, **603**, 127170, https://doi.org/10.1016/j.jhydrol.2021.127170.

Domeneghetti, A., 2016: On the use of SRTM and altimetry data for flood modeling in data-sparse regions. *Water Resour. Res.*, **52**, 2901–2918, https://doi.org/10.1002/2015WR017967.

Eicker, A., L. Jensen, V. Woehnke, H. Dobslaw, A. Kvas, T. Mayer-Guerr, and R. Dill, 2020: Daily GRACE satellite data evaluate short-term hydro-meteorological fluxes from global atmospheric reanalyses. *Sci. Rep.*, **10**, 4504, https://doi.org/10.1038/s41598-020-61166-0.

Ellmer, M., T. Mayer-Gürr, S. Behzadpour, B. Klinger, A. Kvas, S. Strasser, and N. Zehentner, 2018: ITSG-Grace2018: The new GRACE time series from TU Graz. *2018 Fall Meeting*, Washington, D.C., Amer. Geophys. Union, Abstract G13C-0539, https://agu.confex.com/agu/fm18/meetingapp.cgi/Paper/346797.

Frédéric, F., and R. Guillaume, 2018: Monitoring groundwater storage changes using the Gravity Recovery and Climate Experiment (GRACE) satellite mission: A review. *Remote Sens.*, **10**, 829, https://doi.org/10.3390/rs10060829.

Gong, G., S. Mattevada, and S. E. O'Bryant, 2014: Comparison of the accuracy of kriging and IDW interpolations in estimating groundwater arsenic concentrations in Texas. *Environ. Res.*, **130**, 59–69, https://doi.org/10.1016/j.envres.2013.12.005.

Gouweleeuw, B. T., A. Kvas, C. Gruber, A. K. Gain, T. Mayer-Guerr, F. Flechtner, and A. Guentner, 2018: Daily GRACE gravity field solutions track major flood events in the Ganges-Brahmaputra Delta. *Hydrol. Earth Syst. Sci.*, **22**, 2867–2880, https://doi.org/10.5194/hess-22-2867-2018.

Gupta, D., and C. T. Dhanya, 2020: The potential of GRACE in assessing the flood potential of peninsular Indian river basins.

Int. J. Remote Sens., **41**, 9007–9036, https://doi.org/10.1080/01431161.2020.1797218.

Han, S.-C., K. Ghobadi-Far, I.-Y. Yeo, C. M. McCullough, E. Lee, and J. Sauber, 2021: GRACE follow-on revealed Bangladesh was flooded early in the 2020 monsoon season due to premature soil saturation. *Proc. Natl. Acad. Sci. USA*, **118**, e2109086118, https://doi.org/10.1073/pnas.2109086118.

Hao, Z., and V. P. Singh, 2015: Drought characterization from a multivariate perspective: A review. *J. Hydrol.*, **527**, 668–678, https://doi.org/10.1016/j.jhydrol.2015.05.031.

——, ——, and Y. Xia, 2018: Seasonal drought prediction: Advances, challenges, and future prospects. *Rev. Geophys.*, **56**, 108–141, https://doi.org/10.1002/2016RG000549.

Hosseini-Moghari, S.-M., S. Araghinejad, K. Ebrahimi, and M. J. Tourian, 2019: Introducing modified total storage deficit index (MTSDI) for drought monitoring using GRACE observations. *Ecol. Indic.*, **101**, 465–475, https://doi.org/10.1016/j.ecolind.2019.01.002.

——, ——, ——, Q. Tang, and A. AghaKouchak, 2020: Using GRACE satellite observations for separating meteorological variability from anthropogenic impacts on water availability. *Sci. Rep.*, **10**, 15098, https://doi.org/10.1038/s41598-020-71837-7.

Houborg, R., M. Rodell, B. Li, R. Reichle, and B. F. Zaitchik, 2012: Drought indicators based on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage observations. *Water Resour. Res.*, **48**, https://doi.org/10.1029/2011WR011291.

Humphrey, V., and L. Gudmundsson, 2019: GRACE-REC: A reconstruction of climate-driven water storage changes over the last century. *Earth Syst. Sci. Data*, **11**, 1153–1170, https://doi.org/10.5194/essd-11-1153-2019.

——, ——, and S. I. Seneviratne, 2017: A global reconstruction of climate-driven subdecadal water storage variability. *Geophys. Res. Lett.*, **44**, 2300–2309, https://doi.org/10.1002/2017GL072564.

Hutengs, C., and M. Vohland, 2016: Downscaling land surface temperatures at regional scales with random forest regression. *Remote Sens. Environ.*, **178**, 127–141, https://doi.org/10.1016/j.rse.2016.03.006.

Idowu, D., and W. Zhou, 2019: Performance evaluation of a potential component of an early flood warning system-A case study of the 2012 flood, Lower Niger River Basin, Nigeria. *Remote Sens.*, **11**, 1970, https://doi.org/10.3390/rs11171970.

Jiang, T., Q. Zhang, D. Zhu, and Y. Wu, 2006: Yangtze floods and droughts (China) and teleconnections with ENSO activities (1470–2003). *Quat. Int.*, **144**, 29–37, https://doi.org/10.1016/j.quaint.2005.05.010.

Jing, W., P. Zhang, X. Zhao, Y. Yang, H. Jiang, J. Xu, J. Yang, and Y. Li, 2020: Extending GRACE terrestrial water storage anomalies by combining the random forest regression and a spatially moving window structure. *J. Hydrol.*, **590**, 125239, https://doi.org/10.1016/j.jhydrol.2020.125239.

Jong, R., J. Verbesselt, M. E. Schaepman, and S. Bruin, 2012: Trend changes in global greening and browning: Contribution of short-term trends to longer-term change. *Global Change Biol.*, **18**, 642–655, https://doi.org/10.1111/j.1365-2486.2011.02578.x.

Kornfeld, R. P., B. W. Arnold, M. A. Gross, N. T. Dahya, W. M. Klipstein, P. F. Gath, and S. Bettadpur, 2019: GRACE-FO: The Gravity Recovery and Climate Experiment Follow-On mission. *J. Spacecr. Rockets*, **56**, 931–951, https://doi.org/10.2514/1.A34326.

Kurtenbach, E., A. Eicker, T. Mayer-Guerr, M. Holschneider, M. Hayn, M. Fuhrmann, and J. Kusche, 2012: Improved daily GRACE gravity field solutions using a Kalman smoother. *J. Geodyn.*, **59–60**, 39–48, https://doi.org/10.1016/j.jog.2012.02.006.

Kuwayama, Y., A. Thompson, R. Bernknopf, B. Zaitchik, and P. Vail, 2019: Estimating the impact of drought on agriculture using the US Drought Monitor. *Amer. J. Agric. Econ.*, **101**, 193–210, https://doi.org/10.1093/ajae/aay037.

Kvas, A., S. Behzadpour, M. Ellmer, B. Klinger, S. Strasser, N. Zehentner, and T. Mayer-Gürr, 2019: ITSG-Grace2018: Overview and evaluation of a new GRACE-only gravity field time series. *J. Geophys. Res. Solid Earth*, **124**, 9332–9344, https://doi.org/10.1029/2019JB017415.

Lehner, B., P. Doll, J. Alcamo, T. Henrichs, and F. Kaspar, 2006: Estimating the impact of global change on flood and drought risks in Europe: A continental, integrated analysis. *Climatic Change*, **75**, 273–299, https://doi.org/10.1007/s10584-006-6338-4.

Li, B., and Coauthors, 2019: Global GRACE data assimilation for groundwater and drought monitoring: Advances and challenges. *Water Resour. Res.*, **55**, 7564–7586, https://doi.org/10.1029/2018WR024618.

Li, J., Z. Wang, X. Wu, C.-Y. Xu, S. Guo, and X. Chen, 2020: Toward monitoring short-term droughts using a novel daily scale, standardized antecedent precipitation evapotranspiration index. *J. Hydrometeor.*, **21**, 891–908, https://doi.org/10.1175/JHM-D-19-0298.1.

——, ——, ——, J. Zscheischler, S. Guo, and X. Chen, 2021: A standardized index for assessing sub-monthly compound dry and hot conditions with application in China. *Hydrol. Earth Syst. Sci.*, **25**, 1587–1601, https://doi.org/10.5194/hess-25-1587-2021.

Li, S., X. Chen, V. P. Singh, and Y. He, 2018: Assumption-simulation-feedback-adjustment (ASFA) framework for real-time correction of water resources allocation: A case study of Longgang River basin in southern China. *Water Resour. Manage.*, **32**, 3871–3886, https://doi.org/10.1007/s11269-018-2024-3.

Liu, X. J., A. J. Kettner, J. Cheng, and S. B. Dai, 2020: Sediment characteristics of the Yangtze River during major flooding. *J. Hydrol.*, **590**, 125417, https://doi.org/10.1016/j.jhydrol.2020.125417.

Long, D., Y. Shen, A. Sun, Y. Hong, L. Longuevergne, Y. Yang, B. Li, and L. Chen, 2014: Drought and flood monitoring for a large karst plateau in southwest China using extended GRACE data. *Remote Sens. Environ.*, **155**, 145–160, https://doi.org/10.1016/j.rse.2014.08.006.

——, L. Longuevergne, and B. R. Scanlon, 2015: Global analysis of approaches for deriving total water storage changes from GRACE satellites. *Water Resour. Res.*, **51**, 2574–2594, https://doi.org/10.1002/2014WR016853.

Lu, E., 2009: Determining the start, duration, and strength of flood and drought with daily precipitation: Rationale. *Geophys. Res. Lett.*, **36**, L12707, https://doi.org/10.1029/2009GL038817.

——, W. Cai, Z. Jiang, Q. Zhang, C. Zhang, R. W. Higgins, and M. S. Halpert, 2014: The day-to-day monitoring of the 2011 severe drought in China. *Climate Dyn.*, **43**, 1–9, https://doi.org/10.1007/s00382-013-1987-2.

Maimaitijiang, M., V. Sagan, P. Sidike, S. Hartling, F. Esposito, and F. B. Fritschi, 2020: Soybean yield prediction from UAV using multimodal data fusion and deep learning. *Remote Sens. Environ.*, **237**, 111599, https://doi.org/10.1016/j.rse.2019.111599.

Mayer-Gürr, T., S. Behzadpour, M. Ellmer, A. Kvas, B. Klinger, S. Strasser, and N. Zehentner, 2018: ITSG-Grace2018 - Monthly, Daily and Static Gravity Field Solutions from GRACE. GFZ Data Services, accessed 21 January 2022, https://doi.org/10.5880/ICGEM.2018.003.

Miao, Y., and A. Wang, 2020: A daily 0.25 degrees × 0.25 degrees hydrologically based land surface flux dataset for conterminous China, 1961–2017. *J. Hydrol.*, **590**, 125413, https://doi.org/10.1016/j.jhydrol.2020.125413.

Nie, N., W. Zhang, H. Chen, and H. Guo, 2018: A global hydrological drought index dataset based on Gravity Recovery and Climate Experiment (GRACE) data. *Water Resour. Manage.*, **32**, 1275–1290, https://doi.org/10.1007/s11269-017-1869-1.

Otkin, J. A., and Coauthors, 2021: Development of a flash drought intensity index. *Atmosphere*, **12**, 741, https://doi.org/10.3390/atmos12060741.

Pelletier, C., S. Valero, J. Inglada, N. Champion, and G. Dedieu, 2016: Assessing the robustness of random forests to map land cover with high resolution satellite image time series over large areas. *Remote Sens. Environ.*, **187**, 156–168, https://doi.org/10.1016/j.rse.2016.10.010.

Probst, P., M. N. Wright, and A. L. Boulesteix, 2019: Hyperparameters and tuning strategies for random forest. *Wiley Interdiscip. Rev.: Data Min. Knowl. Discovery*, **9**, e1301, https://doi.org/10.1002/widm.1301.

Ramillien, G., J. S. Famiglietti, and J. Wahr, 2008: Detection of continental hydrology and glaciology signals from GRACE: A review. *Surv. Geophys.*, **29**, 361–374, https://doi.org/10.1007/s10712-008-9048-9.

Reager, J. T., and J. S. Famiglietti, 2009: Global terrestrial water storage capacity and flood potential using GRACE. *Geophys. Res. Lett.*, **36**, L23402, https://doi.org/10.1029/2009GL040826.

——, B. F. Thomas, and J. S. Famiglietti, 2014: River basin flood potential inferred using GRACE gravity observations at several months lead time. *Nat. Geosci.*, **7**, 589–593, https://doi.org/10.1038/ngeo2203.

Rodell, M., J. S. Famiglietti, D. N. Wiese, J. T. Reager, H. K. Beaudoing, F. W. Landerer, and M. H. Lo, 2018: Emerging trends in global freshwater availability. *Nature*, **557**, 651–659, https://doi.org/10.1038/s41586-018-0123-1.

Sakumura, C., S. Bettadpur, H. Save, and C. McCullough, 2016: High-frequency terrestrial water storage signal capture via a regularized sliding window mascon product from GRACE. *J. Geophys. Res. Solid Earth*, **121**, 4014–4030, https://doi.org/10.1002/2016JB012843.

Scanlon, B. R., and Coauthors, 2018: Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data. *Proc. Natl. Acad. Sci. USA*, **115**, 201704665, https://doi.org/10.1073/pnas.1704665115.

Schindelegger, M., A. A. Harker, R. M. Ponte, H. Dobslaw, and D. A. Salstein, 2021: Convergence of daily GRACE solutions and models of sub-monthly ocean bottom pressure variability. *J. Geophys. Res. Oceans*, **126**, e2020JC017031, https://doi.org/10.1029/2020JC017031.

Shah, D., and V. Mishra, 2020: Integrated drought index (IDI) for drought monitoring and assessment in India. *Water Resour. Res.*, **56**, e2019WR026284, https://doi.org/10.1029/2019WR026284.

Sharma, D., S. Patnaik, B. Biswal, and J. T. Reager, 2020: Characterization of basin-scale dynamic storage-discharge relationship using daily GRACE based storage anomaly data. *Geosciences*, **10**, 404, https://doi.org/10.3390/geosciences10100404.

Sinha, D., T. H. Syed, and J. T. Reager, 2019: Utilizing combined deviations of precipitation and GRACE-based terrestrial water storage as a metric for drought characterization: A case study over major Indian River basins. *J. Hydrol.*, **572**, 294–307, https://doi.org/10.1016/j.jhydrol.2019.02.053.

Sun, A. Y., B. R. Scanlon, H. Save, and A. Rateb, 2021: Reconstruction of GRACE total water storage through automated machine learning. *Water Resour. Res.*, **57**, e2020WR028666, https://doi.org/10.1029/2020WR028666.

Sun, Z., X. Zhu, Y. Pan, and J. Zhang, 2017: Assessing terrestrial water storage and flood potential using GRACE data in the Yangtze River basin, China. *Remote Sens.*, **9**, 1011, https://doi.org/10.3390/rs9101011.

——, ——, ——, ——, and X. Liu, 2018: Drought evaluation using the GRACE terrestrial water storage deficit over the Yangtze River Basin, China. *Sci. Total Environ.*, **634**, 727–738, https://doi.org/10.1016/j.scitotenv.2018.03.292.

——, D. Long, W. Yang, X. Li, and Y. Pan, 2020: Reconstruction of GRACE data on changes in total water storage over the global land surface and 60 basins. *Water Resour. Res.*, **56**, e2019WR026250, https://doi.org/10.1029/2019WR026250.

Swenson, S., and J. Wahr, 2006: Post-processing removal of correlated errors in GRACE data. *Geophys. Res. Lett.*, **33**, L08402, https://doi.org/10.1029/2005GL025285.

Tapley, B. D., S. Bettadpur, M. Watkins, and C. Reigber, 2004: The gravity recovery and climate experiment: Mission overview and early results. *Geophys. Res. Lett.*, **31**, L09607, https://doi.org/10.1029/2004GL019920.

——, and Coauthors, 2019: Contributions of GRACE to understanding climate change. *Nat. Climate Change*, **9**, 358–369, https://doi.org/10.1038/s41558-019-0456-2.

Thomas, A. C., J. T. Reager, J. S. Famiglietti, and M. Rodell, 2014: A GRACE-based water storage deficit approach for hydrological drought characterization. *Geophys. Res. Lett.*, **41**, 1537–1545, https://doi.org/10.1002/2014GL059323.

Trenberth, K. E., 2011: Changes in precipitation with climate change. *Climate Res.*, **47**, 123–138, https://doi.org/10.3354/cr00953.

Vicente-Serrano, S. M., S. Beguería, and J. I. López-Moreno, 2010a: A multi-scalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. *J. Climate*, **23**, 1696–1718, https://doi.org/10.1175/2009JCLI2909.1.

——, ——, ——, M. Angulo, and A. E. Kenawy, 2010b: A new global 0.5° gridded dataset (1901–2006) of a multi-scalar drought index: Comparison with current drought index datasets based on the Palmer drought severity index. *J. Hydrometeor.*, **11**, 1033–1043, https://doi.org/10.1175/2010JHM1224.1.

Wang, F., Y. Chen, Z. Li, G. Fang, Y. Li, X. Wang, X. Zhang, and P. M. Kayumba, 2021: Developing a long short-term memory (LSTM)-based model for reconstructing terrestrial water storage variations from 1982 to 2016 in the Tarim River basin, Northwest China. *Remote Sens.*, **13**, 889, https://doi.org/10.3390/rs13050889.

Wang, H., J. C. Rogers, and D. K. Munroe, 2015: Commonly used drought indices as indicators of soil moisture in China. *J. Hydrometeor.*, **16**, 1397–1408, https://doi.org/10.1175/JHM-D-14-0076.1.

Wang, J., and Y. Chen, 2021: The applicability of using NARX neural network to forecast GRACE terrestrial water storage anomalies. *Nat. Hazards*, **110**, 1997–2016, https://doi.org/10.1007/s11069-021-05022-y.

Wang, L., Z. Peng, X. Ma, Y. Zheng, and C. Chen, 2021: Multi-scale gravity measurements to characterize 2020 flood events and their spatio-temporal evolution in Yangtze River of China. *J. Hydrol.*, **603**, 127176, https://doi.org/10.1016/j.jhydrol.2021.127176.

Wang, W., P. Wang, and W. Cui, 2015: A comparison of terrestrial water storage data and multiple hydrological data in the Yangtze River basin (in Chinese). *J. Adv. Water Sci.*, **26**, 759–768, https://doi.org/10.14042/j.cnki.32.1309.2015.06.001.

Wang, Z.-Y., and E. J. Plate, 2002: Recent flood disasters in China. *Proc. Inst. Civ. Eng.: Water Marit. Eng.*, **154**, 177–188, https://doi.org/10.1680/wame.2002.154.3.177.

Wei, J., W. Huang, Z. Li, W. Xue, Y. Peng, L. Sun, and M. Cribb, 2019: Estimating 1-km-resolution PM2.5 concentrations across China using the space-time random forest approach. *Remote Sens. Environ.*, **231**, 111221, https://doi.org/10.1016/j.rse.2019.111221.

Wossenyeleh, B. K., K. A. Worku, B. Verbeiren, and M. Huysmans, 2021: Drought propagation and its impact on groundwater hydrology of wetlands: A case study on the Doode Bemde nature reserve (Belgium). *Nat. Hazards Earth Syst. Sci.*, **21**, 39–51, https://doi.org/10.5194/nhess-21-39-2021.

Xia, J., and J. Chen, 2021: A new era of flood control strategies from the perspective of managing the 2020 Yangtze River flood. *Sci. China Earth Sci.*, **64**, 1–9, https://doi.org/10.1007/s11430-020-9699-8.

Yang, P., J. Xia, C. Zhan, and T. Wang, 2018: Reconstruction of terrestrial water storage anomalies in northwest China during 1948-2002 using GRACE and GLDAS products. *Hydrol. Res.*, **49**, 1594–1607, https://doi.org/10.2166/nh.2018.074.

——, ——, X. Luo, L. Meng, S. Zhang, W. Cai, and W. Wang, 2021: Impacts of climate change-related flood events in the Yangtze River Basin based on multi-source data. *Atmos. Res.*, **263**, 105819, https://doi.org/10.1016/j.atmosres.2021.105819.

Yi, H., and L. Wen, 2016: Satellite gravity measurement monitoring terrestrial water storage change and drought in the continental United States. *Sci. Rep.*, **6**, 19909, https://doi.org/10.1038/srep19909.

Yirdaw, S. Z., K. R. Snelgrove, and C. O. Agboma, 2008: GRACE satellite observations of terrestrial moisture changes for drought characterization in the Canadian Prairie. *J. Hydrol.*, **356**, 84–92, https://doi.org/10.1016/j.jhydrol.2008.04.004.

Yuan, X., Z. Ma, M. Pan, and C. Shi, 2015: Microwave remote sensing of short-term droughts during crop growing seasons. *Geophys. Res. Lett.*, **42**, 4394–4401, https://doi.org/10.1002/2015GL064125.

Zhang, D., Q. Zhang, A. D. Werner, and X. Liu, 2016: GRACE-based hydrological drought evaluation of the Yangtze River basin, China. *J. Hydrometeor.*, **17**, 811–828, https://doi.org/10.1175/JHM-D-15-0084.1.

Zhang, D.-F., X. Shi, H. Xu, Q. Jing, X. Pan, T. Liu, H. Wang, and H. Hou, 2020: A GIS-based spatial multi-index model for flood risk assessment in the Yangtze River basin, China. *Environ. Impact Assess. Rev.*, **83**, 106397, https://doi.org/10.1016/j.eiar.2020.106397.

Zhang, H., J. Ding, Y. Wang, D. Zhou, and Q. Zhu, 2021: Investigation about the correlation and propagation among meteorological, agricultural and groundwater droughts over humid and arid/semi-arid basins in China. *J. Hydrol.*, **603**, 127007, https://doi.org/10.1016/j.jhydrol.2021.127007.

Zhao, M., A. Geruo, I. Velicogna, and J. S. Kimball, 2017: Satellite observations of regional drought severity in the continental United States using GRACE-based terrestrial water storage changes. *J. Climate*, **30**, 6297–6308, https://doi.org/10.1175/JCLI-D-16-0458.1.

Zhong, L., L. Hu, and H. Zhou, 2019: Deep learning based multi-temporal crop classification. *Remote Sens. Environ.*, **221**, 430–443, https://doi.org/10.1016/j.rse.2018.11.032.