Evapotranspiration Reconstruction Based on Land Surface Models and Observed Water Budget Components While Considering Irrigation^①

MEIXIA LV

Key Laboratory of Regional Climate-Environment for Temperate East Asia, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

ZHUGUO MA

Key Laboratory of Regional Climate-Environment for Temperate East Asia, Institute of Atmospheric Physics, Chinese Academy of Sciences, and University of Chinese Academy of Sciences, Beijing, China

LIANG CHEN

Key Laboratory of Regional Climate-Environment for Temperate East Asia, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

SHAOMING PENG

Yellow River Engineering Consulting Co. Ltd., Zhengzhou, China

(Manuscript received 18 April 2019, in final form 4 August 2019)

ABSTRACT

The accurate estimation of evapotranspiration (ET) is essential for understanding the land surface–atmosphere interaction; however, current ET products have large uncertainties, and irrigation effects on ET are not well represented. In this study, the monthly ET was reconstructed (ET_{recon}) from GLDAS land surface models (LSMs) over the Yellow River basin of China, which was achieved by using observation-based precipitation, naturalized streamflow, and downscaled consumed irrigation water from the census annual data via an irrigation scheme. The results showed that the monthly ET_{recon} series were generally improved relative to the original LSM-based ET, with improvements in the correlation coefficient, Nash–Sutcliffe efficiency, mean absolute error, and root-mean-square error by 0.6%–1.8%, 1.2%–14.6%, 1.3%–21.0%, and 2.1%–20.4%, respectively. The ET_{recon} results were also superior to the collected ET synthesis products in terms of statistics, with generally higher peak values occurring in ET_{recon} . Regarding the annual time scale, the ET_{recon} values were close to the water balance ET values, which have been widely used as benchmark data. The interannual variability in ET_{recon} was good overall and was associated with the LSM precipitation variability and partitioning of precipitation into ET and runoff. The reconstruction method can provide an alternative ET estimate for other river basins. This study will also be valuable for studies and applications in climate change evaluation, drought assessment, and water resources management.

1. Introduction

Evapotranspiration (ET) is a major component of the global water cycle, which links the water and energy cycles at all scales. Accurate ET estimates are critical

for a better understanding of the interaction between climate and hydrology. However, ET remains particularly difficult to measure and simulate, especially at large spatial scales (Rodell et al. 2011; Mao and Wang 2017) due to land surface heterogeneity and multiple controlling factors (Liu et al. 2013). Although direct ET measurements are available via eddy covariance systems and Bowen ratio energy balance systems, these measurements usually have limited durations and spatial coverage (Mao and Wang 2017). Sparse measurements cannot represent the spatial heterogeneity of ET at large

^① Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JHM-D-19-0090.s1>.

Corresponding author: Zhuguo Ma, mazg@tea.ac.cn

scales (Gao et al. 2010), and there is currently no methodology to directly observe the intra-annual ET over a large region (Castle et al. 2016). Importantly, the currently available ET products have large uncertainties (Wang and Dickinson 2012; Mueller et al. 2013; Lei et al. 2015; Liu et al. 2016b; Dong and Dai 2017; Mao and Wang 2017). ET has been reported as one of the most uncertain variables that may induce failures in the water balance estimates of terrestrial water storage (TWS) change (Syed et al. 2009; Long et al. 2015). For the water budget, in terms of precipitation, ET, runoff, and TWS change (TWSC), Zhang et al. (2018) reported that the attribution of water budget nonclosure errors to ET was the largest among the water budget terms, both across the globe (45.4%) and in the Yellow River basin (51%). Therefore, the uncertainty in ET needs to be further reduced.

Notably, irrigation effects have not been well represented in the currently available ET estimates (Gao et al. 2007; Gowda et al. 2007; Liu et al. 2013; Lei et al. 2015; Castle et al. 2016; Lv et al. 2017). For instance, numerical model estimates of ET have tended to underestimate ET in managed river basins because they respond to natural forcing alone and typically do not take human water use impacts into account (Rodell et al. 2011; Castle et al. 2016; Pan et al. 2017). Furthermore, satellite remote sensing products generally tend to underestimate ET (Gao et al. 2010; Jia et al. 2012; Lei et al. 2015; Wan et al. 2015). Lei et al. (2015) found that remotely sensed ET cannot correctly capture the trend in simulated ET with the observed irrigation over the Haihe River basin of China, in which the Moderate Resolution Imaging Spectroradiometer (MODIS) ET (Mu et al. 2011), ET from the University of Montana (Zhang et al. 2010), and model tree ensembles-based ET (Jung et al. 2011) were employed. Mueller et al. (2013) and Zhang et al. (2018) proposed synthesized monthly ET products from multiple sources [e.g., in situ observation, land surface model (LSM), reanalysis, and satellite remote sensing], but the irrigation effects on ET were not explicitly mentioned in either study. Nevertheless, irrigation accounts for ~70% of the global freshwater withdrawals and ~90% of consumptive water uses (Siebert et al. 2010) and can lead to a decrease in the streamflow and an increase in ET (Haddeland et al. 2006). Castle et al. (2016) reported that 38% of the total satellite-based observed ET signal, which was regarded as a value that accounted for human impacts in their study and was calculated from MODIS ET (Tang et al. 2009) and ET derived from the Gravity Recovery and Climate Experiment (GRACE), was due to irrigation in July over the Colorado River basin. An average proportion of 12% in the annual water balance ET was

evoked by water consumption in a number of dry basins worldwide (Ukkola and Prentice 2013).

Among multiple ET products, LSM ET has been reported to have a relatively low uncertainty (Mueller et al. 2013; Long et al. 2014; Wang et al. 2015). For instance, Long et al. (2014) found that the uncertainty in the North American Land Data Assimilation System (NLDAS) ET estimates, which were basically estimated by LSMs, was the lowest compared with the uncertainties in the ET results derived from remote sensing and GRACE satellites. Mao and Wang (2017) demonstrated that the offline land reanalysis ET values were closest to the water balance ET for China compared with estimates derived from the Penman–Monteith model and atmospheric reanalyses. However, most LSMs do not account for human interventions (e.g., irrigation) (Rodell et al. 2011; Leng et al. 2015; Long et al. 2015; Scanlon et al. 2018). Admittedly, there are some studies that have incorporated irrigation and groundwater pumping into LSMs (e.g., Haddeland et al. 2006; Leng et al. 2014; Lawston et al. 2015; Pokhrel et al. 2015), but the actual irrigation practices are difficult to accurately represent and assessments of the realism of irrigation schemes are difficult due to limited observations (Lei et al. 2015; Lawston et al. 2017). Nie et al. (2018), for instance, failed to replicate the reported magnitudes of groundwater irrigation in the U.S. High Plains aquifer (~49% relative error), and they emphasized the need to better simulate the inefficiency of current, widely employed irrigation practices. Lawston et al. (2017) reported that the individual actions of farmers were not captured by the assessed irrigation algorithm, and the amounts of irrigation water applied at the two studied irrigated fields were generally overestimated. Generally, most LSM ET products need to be corrected for river basins with intensive irrigation. Furthermore, to reduce the uncertainty in ET estimates, the use of more “ground truth” datasets to constrain LSM ET is necessary (Mueller et al. 2011; Liu et al. 2016b). Among the water budget terms, in situ measurements with small uncertainties exist for precipitation and runoff. Therefore, it is worth studying how to constrain LSM-simulated ET using high-quality precipitation, runoff, and irrigation water use data.

This study is an extension of our earlier work, reported by Lv et al. (2017), in which the annual ET was reconstructed using widely used Global Land Data Assimilation System (GLDAS) 1.0 LSMs, observation-based precipitation, naturalized streamflow without human intervention, and irrigation water. This study focuses on a finer time scale of the ET reconstruction (i.e., monthly), from six LSM runs in GLDAS 1.0, 2.0,

FIG. 1. Location and land cover in 2010 from the Moderate Resolution Imaging Spectroradiometer (MODIS) for the Yellow River basin of China.

and 2.1. The primary objectives of this study are to 1) correct the monthly LSM-based ET using observation-based precipitation, naturalized streamflow, and census annual consumed irrigation water; 2) propose an offline irrigation scheme to downscale the census annual irrigation amount to individual months; and 3) verify the reconstructed monthly ET using the available multidataset synthesis products and GRACE-derived water balance ET.

2. Study area and data

a. Study area

In this study, the Yellow River basin in northern China (hereafter, YRB; Fig. 1) was chosen as the study region due to the availability of data; however, the method could theoretically be applied to other basins. The YRB has attracted substantial attention due to its crucial water supply function for Northwest and North China. The mean annual precipitation was $\sim 457 \text{ mm yr}^{-1}$ from 1961 to 2013, and ET consumes $\sim 80\%$ of the basin total water input (Liu and Yang 2010; Pei et al. 2017). The YRB is a heavily managed river basin with a considerable number of dams, reservoir pump sites, and groundwater wells. The basin is influenced by intensive agricultural irrigation (Lei et al. 2013; Lv et al. 2017), and the consumed irrigation water was reported to account for $\sim 70\%$ of the total water consumption in the basin according to the Bulletin of Water Resources (Yellow River Conservancy

Commission 2018). Therefore, this basin is an ideal test-bed for ET reconstruction from LSMs using high-quality precipitation, streamflow, and irrigation data. The study period was specified to 1998–2010 due to the data availability of the naturalized streamflow and irrigation water. The actual irrigation information for the Zuncun irrigation district and the entire basin was collected.

b. Precipitation, streamflow, and irrigation water

Two sets of precipitation data were collected in this study (Table 1). One was the precipitation dataset measured in situ, which was provided by the China Meteorological Administration (CMA, available at <http://data.cma.cn/en>). It is a 0.5° monthly gridded dataset from 1961 to present, which was derived from ~ 2470 gauge stations throughout China (hereafter, CMA- 0.5°). The second dataset was the Multi-Source Weighted-Ensemble Precipitation (MSWEP; Beck et al. 2017; available at <http://www.gloh2o.org>). It is a global gridded precipitation dataset with a 0.25° spatial and 3-hourly temporal resolution, obtained by merging gauge, satellite, and reanalysis data. The MSWEP data had been validated using four state-of-the-art gauge-adjusted precipitation datasets by Beck et al. (2017), and it performed fairly well in the YRB modeling (Liu et al. 2016a). To reduce the uncertainty, the average of these two precipitation datasets was used (hereafter, observation-based precipitation). The comparison of these two precipitation datasets is shown in Fig. S1 in the online supplemental material,

TABLE 1. Summary of the employed datasets in this study.

Variable	Source	Spatial resolution	Spatial extent	Temporal resolution	Time span	Reference
<i>P</i>	1. CMA 2. MSWEP	0.5° 0.25°	China Global	Monthly 3-hourly	1961–present 1979–2015	http://data.cma.cn/en Beck et al. (2017)
<i>R</i>	1. Gauged runoff 2. Naturalized runoff	Basin integrated Basin integrated	Yellow River basin Yellow River basin	Monthly Monthly	1998–2010 1998–2010	Yellow River Conservancy Commission Yellow River Conservancy Commission
TWSA	1. Spherical harmonic solutions 2. JPL mascon solution	1° 0.5°	Global Global	Monthly Monthly	Apr 2002–Jan 2017 Apr 2002–Aug 2016	Landerer and Swenson (2012) Watkins et al. (2015)
ET	1. GLDAS 1.0 (Noah 2.7, CLM 2.0, VIC 4.0.4, and Mosaic) 2. GLDAS 2.0 (Noah 3.3) 3. GLDAS 2.1 (Noah 3.3) 4. Jung 5. CDR 6. LandFlux-EVAL 7. Water balance ET ($ET = P - R - TWSC$)	1° 0.25° 0.25° 0.5° 0.5° 1° Basin integrated	Global Global Global Global Global Global Yellow River basin	Monthly Monthly Monthly Monthly Monthly Monthly Monthly	1979–present 1948–2010 2000–present 1982–2011 1984–2010 1989–2005 2003–10	Rodell et al. (2004b) Rui (2015) Rui et al. (2018) Rui et al. (2018) Jung et al. (2010) Zhang et al. (2018) Mueller et al. (2013) Based on the average <i>P</i> between CMA-0.5° and MSWEP, gauged runoff, and GRACE-derived TWSC
Irrigation	1. Census consumed irrigation water amount 2. Measured seasonal irrigation amount 3. Statistical estimate of mean monthly irrigated water proportions	Basin integrated Irrigation district integrated Basin integrated	Yellow River basin Zuncun irrigation district Yellow River basin	Annual Spring, summer, and winter Mean monthly	1998–present 1998–2006 1950–80	Yellow River Conservancy Commission Zuncun irrigation district administration Ruan and Fang (1996)

which shows that the root-mean-square difference (RMSD) was $4.83 \text{ mm month}^{-1}$ in the basin-averaged precipitation between the MSWEP and CMA-0.5° datasets for 1998–2015 over the YRB.

The monthly naturalized and observed streamflow (hereafter, R_{nature} and R_{ob} , respectively) for 1998–2010 and the census annual consumed irrigation water for 1998 to present were collected for the YRB. Both the R_{nature} (regarded as the flow without any human interventions) and R_{ob} data were provided by the Yellow River Conservancy Commission, which is responsible for the quality control and release of the water data of the basin. The difference between the naturalized streamflow and observed streamflow is due to human interventions, such as reservoir regulation, water withdrawal from the river channel, and water transported into and out of the basin (Fu et al. 2007). The river commission has developed comparatively mature technology for establishing the naturalized streamflow value from the observed flow, which was realized by adding back the water regulated by reservoirs and consumed by irrigation, industry, and domestic uses (Li et al. 2001; Fu et al. 2004). The naturalized streamflow is by far the most reliable natural value and has been widely applied in hydrological analysis and water resource management (Fu et al. 2007; Tang et al. 2013; Yuan et al. 2016). The annual irrigation water amount of the entire basin was collected from the Bulletin of Water Resources (Yellow River Conservancy Commission 2018). As stated in the Bulletin of Water Resources, the consumed irrigation water is the volume finally consumed as evapotranspiration regardless of the sources of irrigated water (i.e., surface water or groundwater). More information on consumed irrigation water can be found in the study reported by Lv et al. (2017). Furthermore, the measured seasonal irrigation water amounts of the Zuncun district in the spring, summer, and winter were collected from the Zuncun irrigation district administration, which were available from 1998 to 2006. The reported statistical estimates of the mean monthly irrigated water proportions from 1950 to 1980 over the entire YRB were collected from the study of Ruan and Fang (1996).

c. GRACE TWS anomaly

The monthly GRACE-derived TWS anomaly (TWSA) values, which represent terrestrial water storage changes related to climate variability and human impacts, were collected to estimate the water balance ET (available at <http://grace.jpl.nasa.gov/>). Three TWSA spherical harmonic solutions are available from the Center for Space Research (CSR), the German Research Center for Geoscience (GFZ), and the Jet Propulsion Laboratory (JPL). These three TWSA datasets are provided at 1° grids along with the associated leakage and

measurement errors (Landerer and Swenson 2012). In addition, the more recent GRACE mass concentration (mascon) solution (Watkins et al. 2015) from the JPL (hereafter, JPL mascon) was also collected, which is provided on 0.5° global grids. For example, Rodell et al. (2018) used the JPL mascon solution for the trend analysis and mapping of TWSA worldwide. GRACE can resolve the monthly TWSA with sufficient accuracy over regions from $\sim 200\,000 \text{ km}^2$ at low latitudes to $90\,000 \text{ km}^2$ near the poles (Rodell et al. 2018). In our study, the area of the YRB is $795\,044 \text{ km}^2$, hence GRACE TWSA was expected to have a sufficient accuracy over this basin. Finally, the average of the four TWSA datasets was used in this study to minimize the uncertainties associated with data processing.

d. Data from the GLDAS land surface models

The ET results of six LSM runs from the widely used GLDAS 1.0, 2.0, and 2.1 (Rodell et al. 2004b) were used as the initial values for ET reconstruction (Table 1). GLDAS is a global offline terrestrial modeling system that was designed to generate the optimal fields of land surface states and fluxes via the assimilation of ground-based and satellite products. The forcing data were different among the three GLDAS versions. The forcing data sources in GLDAS 1.0 were switched several times among the 1979–93, 1994–99, 2000, and 2001–present periods (Rui 2015). GLDAS 2.0 was forced entirely with the Princeton meteorological forcing data, and GLDAS 2.1 was forced with a combination of model and observation-based forcing datasets (Rui et al. 2018). There are four LSMs in GLDAS 1.0 with a 1° spatial resolution, that is, the Community Land Model (CLM 2.0), Variable Infiltration Capacity (VIC 4.0.4), Noah 2.7, and Mosaic. The Noah 3.3 model was included in GLDAS 2.0 and 2.1 at a 0.25° resolution. Surface and subsurface runoff were simulated by these LSMs without the consideration of groundwater and water resources management (e.g., irrigation and reservoir regulation). The Noah 0.25° soil moisture values from GLDAS 2.0 and 2.1 were used to estimate the irrigation water requirements (see below). The performances of the GLDAS 2.0 and 2.1 soil moisture results in the YRB have been validated by Lv et al. (2018) and Lv et al. (2019).

3. Methods

a. Downscaling of the census annual irrigation water to individual months

Since only the annual consumed irrigation water is available in the Bulletin of Water Resources, the census

annual values had to be downscaled to individual months to reconstruct the monthly ET. To achieve the downscaling, two offline soil-moisture-dependent irrigation schemes were proposed in this study. The first irrigation scheme shown in Eq. (1) was proposed according to irrigation and drainage engineering (Guo 2007). Given that there are very few paddy lands in the study basin, only nonpaddy lands were considered in this irrigation scheme. The “Global Map of Irrigation Areas” (Siebert et al. 2013), which was downloaded from the Food and Agriculture Organization (FAO) of the United Nations at <http://www.fao.org/nr/water/aquastat/irrigationmap/index10.stm>, was applied in this scheme. This dataset shows irrigation area information around the year 2005 at a 5-min spatial resolution. According to the statistical results in Ruan and Fang (1996), the irrigation months were from March to November in the area above the Longmen station and from January to December below the Longmen station in the YRB. The irrigated time was designed accordingly when using the scheme to estimate the irrigation water at a monthly time scale without accounting for different crop types. As shown in Eq. (2), another irrigation scheme was proposed that was based on crop-specific monthly growing areas (Portmann et al. 2010; available at http://www.uni-frankfurt.de/45218031/Data_download_center_for_MIRCA2000). The following four major crops were taken into account in the second scheme: wheat, maize, soybeans, and rice. According to Guo (2007), only the water in one certain soil layer, which is usually called the target irrigated soil layer, can be used by crops. The target irrigated soil layer, which is usually determined based on practical experiments, was determined to be 0–100 cm in the irrigation schemes of our study. First, the depth of 0–100 cm was generally considered as the root zone (Cai et al. 2014; Baldwin et al. 2017), which has been applied in existing irrigation models (e.g., Hanasaki et al. 2010; Marcella and Eltahir 2014; Zhang et al. 2017). Second, the Noah soil moisture values from GLDAS 2.0/2.1, with soil depths of 0–10, 10–40, 40–100, and 100–200 cm, were employed to estimate the irrigation requirements. The irrigation water amounts were estimated grid-by-grid based on the GLDAS soil moisture, although our study focused on the values integrated over the YRB:

$$W_{\text{irrig},n}(t) = [\omega \times \theta_{\text{FC}} - \theta_n(t)] \times A \times a, \quad (1)$$

$$W_{\text{irrig},n}(t) = [\omega \times \theta_{\text{FC}} - \theta_n(t)] \times \beta(t), \quad (2)$$

where $W_{\text{irrig},n}(t)$ is the simulated irrigation requirement for month t of year n , θ_{FC} is the field capacity (Scholes and Colstoun 2011), $\theta_n(t)$ is the soil moisture content, A

is the area equipped for irrigation, a is the percentage of the actually irrigated area in the area equipped for irrigation, and $\beta(t)$ is the crop-specific monthly growing area. Both A and a were included in the aforementioned Global Map of Irrigation Areas dataset. The $\omega \times \theta_{\text{FC}}$ is the target soil moisture content in the irrigated soil layer. The field capacity values are usually used to determine the irrigation threshold (Hanasaki et al. 2008; Pokhrel et al. 2015; Zhang et al. 2017). For instance, Hanasaki et al. (2010) maintained the soil moisture content of the upper 100 cm at field capacities of 100% and 75% for rice and other crops in the irrigated cropland, respectively. Döll et al. (2014) reported that groundwater depletion and TWS trends were best modeled under the assumption that farmers irrigated at 70% of optimal water demand at the global scale. In our study, ω was set at 100% for rice (Hanasaki et al. 2010; Pokhrel et al. 2015) and 80% for other crops (which was optimized among 70%, 75%, and 80%).

The proposed irrigation schemes were evaluated with respect to the census annual irrigation amounts and reported mean monthly irrigation proportions of the entire YRB, as well as the seasonal irrigation proportions in the Zuncun irrigation district. Based on the simulated monthly irrigation proportion, the census annual total consumed irrigation water could be downscaled into individual months using Eqs. (3) and (4):

$$Q_{\text{irrig},n}(t) = W_n \times r_n(t), \quad (3)$$

$$r_n(t) = \frac{W_{\text{irrig},n}(t)}{\sum_{t=1}^{12} W_{\text{irrig},n}(t)}, \quad (4)$$

where $Q_{\text{irrig},n}(t)$ is the downscaled irrigation water for month t of year n , W_n is the census annual irrigation amount, and $r_n(t)$ is the simulated monthly irrigation proportion. Specifically, $r_n(t)$ is the ratio of the simulated monthly irrigation water to the annual total value.

b. Reconstruction of actual evapotranspiration

In our previous study, Lv et al. (2017), the reconstructed annual ET was obtained from four GLDAS 1.0 LSMs from the perspective of the water balance. The present study is an extension of this previous work, which focused on the monthly scale. It has been found that both trends and variabilities in ET show strong control by precipitation (Ukkola and Prentice 2013), and both the GLDAS precipitation and runoff estimates had biases over the YRB (Lv et al. 2017). Therefore, the GLDAS LSM ET should be corrected for the following aspects (similar to the annual ET reconstruction): the biases in precipitation and runoff, and

the impact of irrigation on ET. The collected high-quality observation-based precipitation, naturalized streamflow, and downscaled consumed irrigation water were employed in the reconstruction procedure introduced below.

First, for the precipitation bias in GLDAS LSMs, the simulated runoff and evapotranspiration were corrected as follows:

$$R_2(t) = R_1(t) + \alpha(t) \times [P_2(t) - P_1(t)], \quad (5)$$

$$ET_2(t) = ET_1(t) + \beta(t) \times [P_2(t) - P_1(t)], \quad (6)$$

where t represents each month, R_1 and ET_1 are the original LSM runoff and evapotranspiration, R_2 and ET_2 are the corrected runoff and evapotranspiration, P_1 is the original LSM precipitation, and P_2 is the observation-based precipitation. Parameter $\alpha(t) = R_1(t)/P_1(t)$ is the runoff coefficient representing the underlying surface characteristics for runoff generation, and $\beta(t) = ET_1(t)/P_1(t)$ is the evapotranspiration ratio. The concept of the runoff coefficient dates back to the early twentieth century, and has since been widely used in hydrological research and applications (Sherman 1932; Sriwongsitanon and Taesombat 2011). The obtained R_2 and ET_2 are considered the LSM results after the bias-correction of the precipitation. Then ET_2 was further corrected for the runoff bias. Generally, an overestimation in the runoff simulation corresponds to an underestimation in ET, and vice versa (Lv et al. 2017). Pei et al. (2017), for instance, found a significant negative correlation between the ET and runoff.

Second, for the runoff bias in GLDAS LSMs, ET_2 was further corrected as shown below:

$$ET_{\text{nature}}(t) = ET_2(t) + [R_2(t) - R_{\text{nature}}(t)] \times \frac{1}{\gamma(t) + 1}, \quad (7)$$

$$\gamma(t) = \left| \frac{1 - \beta(t) - \alpha(t)}{\beta(t)} \right|, \quad (8)$$

where $ET_{\text{nature}}(t)$ represents the bias-corrected natural evapotranspiration, and $\gamma(t)$ is used to partition the runoff bias to ET. At the monthly scale, the systematic bias in the runoff cannot be directly added to ET_2 as at the annual scale in Lv et al. (2017). In terms of the water balance equation of $P = ET + R + \text{TWSC}$, the bias in the simulated runoff was partitioned to ET and TWSC according to their relative magnitudes [i.e., $\gamma(t) = |\text{TWSC}(t)/ET(t)|$; see supplemental material for details].

Finally, the LSM-based ET_{nature} must be corrected by adding the consumed irrigation water to take the

irrigation effect into account, as shown in Eq. (9). ET_{recon} is considered as the final reconstructed evapotranspiration:

$$ET_{\text{recon}}(t) = ET_{\text{nature}}(t) + Q_{\text{irrig}}(t), \quad (9)$$

where Q_{irrig} is the downscaled monthly consumed irrigation water. The reconstructed LSM-based ET_{recon} results were obtained in this study from the Noah, CLM, VIC, and Mosaic LSMs in GLDAS 1.0, 2.0, and 2.1.

c. Verification of the reconstructed ET

Currently, it remains difficult to obtain in situ measured ET for large regions. The ET_{recon} results can be evaluated using the water balance ET (calculated from observed precipitation, runoff, and terrestrial water storage change) and the available multidataset ET synthesis products. In this study, the following four datasets were used: 1) the water balance ET computed as $ET = P - R - \text{TWSC}$ using observation-based precipitation, the gauged runoff, and GRACE-derived TWSC; 2) the ET from Jung et al. (2009) and Jung et al. (2010) (hereafter, Jung ET); 3) the LandFlux-EVAL multidataset synthesis product (Jiménez et al. 2011; Mueller et al. 2011; Mueller et al. 2013); and 4) the global terrestrial water cycle Climate Data Record (CDR) from Princeton University (Zhang et al. 2018) (Table 1). In this study, the ET verification was conducted against the water balance ET estimate that has been widely used as benchmark data to verify other ET datasets (Mao and Wang 2017; Zhang et al. 2018). The water balance ET integrated over a river basin is generally regarded as “the most firmly observationally based estimator of ET” (Ukkola and Prentice 2013).

The water balance ET was estimated by using the gauged runoff and GRACE-derived TWSC that were both already influenced by water management. The uncertainties in the water balance ET were estimated, which will be discussed in section 6. Note that the observation-based precipitation was used in both the ET reconstruction and water balance ET estimation because this precipitation is almost the most reliable data nowadays, which is not an issue in the study. First, we always prefer to use the most accurate data in analyses, whether it is precipitation bias correcting or water balance ET estimating. Furthermore, the ET reconstruction relied not only on the precipitation amount but also on the LSM partitioning of precipitation into ET and runoff. Second, the water balance ET based on another precipitation dataset, which included 57 CMA daily meteorological stations across the YRB (Lv et al. 2018), was also estimated to examine the possible influence of using the observation-based precipitation (i.e., aforementioned P_2). The results

showed that these two precipitation series led to a very small difference in the estimated water balance ET values, with the RMSD being $4.25 \text{ mm month}^{-1}$ (Fig. S2). The Jung ET, a 0.5° monthly global dataset for 1982–2011, was compiled using the global network of eddy covariance towers (FLUXNET, www.fluxdata.org) data, remote sensing, and meteorological observations in a machine-learning algorithm. The LandFlux-EVAL project (<http://www.iac.ethz.ch/url/research/LandFlux-EVAL>) is aimed at evaluating and intercomparing the currently available land ET datasets and providing global merged benchmark synthesis products. The CDR ET (0.5° monthly for 1984–2010) was developed by merging multiple data sources for precipitation, runoff, ET, and TWSC with enforced water budget closure (available at http://stream.princeton.edu:8080/opendap/MEaSURES/WC_MULTISOURCES_WB_050/). In our study, four statistics, that is, the Pearson correlation coefficient (r), Nash–Sutcliffe efficiency (NSE), mean absolute error (MAE), and root-mean-square error (RMSE), were calculated relative to the water balance ET:

$$\text{NSE} = 1.0 - \frac{\sum_{t=1}^T [\text{ET}(t) - \text{ET}_{\text{WB}}(t)]^2}{\sum_{t=1}^T [\text{ET}_{\text{WB}}(t) - \overline{\text{ET}_{\text{WB}}}]^2}, \quad (10)$$

where $\text{ET}(t)$ is the evapotranspiration, including the reconstructed/original ET from GLDAS and the collected ET products, $\text{ET}_{\text{WB}}(t)$ is the water balance ET, and t is the time.

4. Verification of the irrigation schemes

The proposed irrigation schemes were evaluated with respect to the census annual irrigation water amount and the reported seasonal and mean monthly irrigated water proportions. The results showed that the performances of the irrigation water simulation were similar between these two schemes, but the correlation coefficients under the first scheme based on Eq. (1) were slightly higher (Fig. 2 and Fig. S3). The performance of the first irrigation scheme is therefore discussed below in detail.

Regarding the first scheme shown in Eq. (1), the simulated annual total irrigation water, based on the GLDAS 2.0 soil moisture, over the entire YRB generally compared well with the reported estimates from the Bulletin of Water Resources for 1998–2010 (Fig. 2a). There is no systematic bias in the irrigation amount simulation with a mean annual absolute relative error of 11.0%, and the correlation coefficient was 0.748, which

FIG. 2. Verification of the proposed irrigation scheme shown in Eq. (1) based on the GLDAS 2.0 soil moisture. (a) Annual irrigation water amount over the entire basin from 1998 to 2010 ($10^8 \text{ m}^3 \text{ yr}^{-1}$), (b) mean monthly irrigated water proportions from 1990 to 2010 and from 1950 to 1980 (%), and (c) seasonal irrigated water proportions for each year (i.e., spring, summer, and winter) in the Zuncun district from 1998 to 2006 due to data availability (%).

was significant at the 99% level of the Student's t test. In the study by Haddeland et al. (2006), the absolute relative errors of the simulated annual irrigation water requirements were 3.1%–187.5% for the six regions in the Colorado and Mekong River basins, and the error was 11.1% for the Colorado River basin with a dry climate. The annual irrigation amounts were generally

overestimated in the Hebei Province of the Hai River basin by Lei et al. (2015). In our study, the estimated annual irrigation water based on the soil moisture deficit was also affected by the monthly distribution of the annual precipitation, which may partly explain why the precipitation for 2000 was less than that for 2002 while the census irrigation amount for 2000 was not higher than that for 2002. Irrigation in practice is complicated and flexible, so it is difficult to come up with a universal parameterization scheme (Lei et al. 2015; Castle et al. 2016). The realistic irrigation process is difficult to capture accurately using a simple scheme based only on the soil moisture deficit. Another important influence factor for the irrigation simulation is the irrigation management decision, which is flexible and difficult to account for in parameterizations (Wada et al. 2014; Lawston et al. 2017). Lei et al. (2015) found that the irrigation scheme incorporated in CLM 4 (<http://www.cesm.ucar.edu/models/cesm1.0/clm/CLMcropANDirrigTechDescriptions.pdf>) cannot capture the trend in the annual irrigation amount caused by a nonlinear change in agricultural water management. However, it can generally reproduce the inter-annual variation of the annual irrigation when irrigation water management is stable. In our study, the census annual irrigation amounts are relatively stable; hence, the impact of irrigation water management is relatively small. The annual irrigation water simulation of our study exhibited an overall good performance with an acceptable error.

The particular interest of our study is the monthly variability of irrigation water. Given that the reported mean monthly irrigation proportions were only available for 1950–80 and that the ET reconstruction period was 1998–2010, the simulated mean monthly proportion values in 1950–80 and 1990–2010 (to be close to the study period of 1998–2010 and the length of 1950–80) were examined for the verification of the irrigation simulation based on the GLDAS 2.0 soil moisture. The simulated mean monthly irrigation proportions for 1950–80 and 1990–2010 matched well with the reported estimates for 1950–80 over the entire YRB (Fig. 2b). The correlation coefficients were 0.935 and 0.915, respectively, which were both significant at the 99% level. Note that the wintertime irrigation proportions were overestimated in the simulations, which was partly related to the different time periods between the simulations and the collected data of the actually irrigated area (around the year 2005). As shown in Fig. 2c, the simulated irrigation proportions for the spring, summer, and winter seasons showed a largely good agreement with the observations in the Zuncun irrigation district, with a correlation coefficient of 0.817 for 1998–2006. The relatively large deviations in

the simulated seasonal proportions in 2001 and 2002 (Fig. 2c) may be related to biases in the LSM soil moisture. The simulated three-season irrigation proportions of the Zuncun irrigation district were not as good as the mean monthly proportions over the entire YRB, which is likely due to the small area of the Zuncun district and the different statistical time (i.e., one is a time series with three values in each year and the other is about the long-term average).

As mentioned earlier, both the GLDAS 2.0 and 2.1 soil moisture values had been validated by Lv et al. (2018) and Lv et al. (2019) in the YRB, and these two datasets were employed to examine the sensitivity of the irrigation simulation to the applied soil moisture values, especially for the monthly irrigation proportion that we focused on. The performance of the irrigation simulation based on the GLDAS 2.1 soil moisture beginning in 2000 was similar to that from the GLDAS 2.0 soil moisture (Fig. 3). Both the irrigation simulations from the GLDAS 2.0 and 2.1 soil moistures were satisfactory with acceptable errors, and the mismatch of the simulated annual irrigation amounts against the census values mainly came from the used LSM soil moisture and the simplistic representation of irrigation. The difference in the estimated irrigation water amounts from GLDAS 2.0 and 2.1 was due to the difference in the soil moisture of 0–100 cm, which was related to the precipitation inputs (An et al. 2017). The downscaled monthly irrigation amounts under these two soil moisture datasets are shown in Fig. 4, which shows a generally good match between these two downscaled series. The correlation coefficient was 0.927 (p value < 0.0001) and the mean monthly absolute difference was $0.413 \text{ mm month}^{-1}$ from 2000 to 2010. That is, these two soil moisture datasets had little effect on the downscaled monthly irrigation water that was ultimately used in the ET reconstruction, although there is a relatively large difference in the annual precipitation between GLDAS 2.0 and 2.1. Note that, since the simulated monthly irrigation proportions rather than the irrigation amounts were used in our study, the impacts of uncertainties in the soil moisture and precipitation were accordingly attenuated. In summary, the first irrigation scheme based on Eq. (1) generally exhibited a satisfactory performance in simulating the total annual irrigation water amounts and monthly proportions; hence, the scheme was used to downscale the census annual consumed irrigation water to individual months.

5. Reconstruction of the actual evapotranspiration

The monthly basin-averaged ET values were reconstructed from six LSM runs in GLDAS 1.0, 2.0, and

FIG. 3. As in Fig. 2, but based on the GLDAS 2.1 soil moisture. (a) Annual irrigation water amount over the entire basin from 2000 to 2015 ($10^8 \text{ m}^3 \text{ yr}^{-1}$), (b) mean monthly irrigated water proportions from 2000 to 2010 (%), and (c) seasonal irrigated water proportions for each year (i.e., spring, summer, and winter) in the Zuncun district from 2000 to 2006 (%).

2.1, together with the observation-based precipitation, naturalized streamflow, and downscaled irrigation water amounts. These three GLDAS versions are different in the precipitation forcing datasets and model versions. The degree of consistency among the reconstructed ET from the above six LSM runs could be used to examine the robustness of the reconstruction method. The study time periods were 1998–2010 and 2000–10 for the reconstructed ET based on GLDAS

1.0/2.0 and 2.1, respectively. Given the relatively long time span, the reconstructed ET values using the downscaled irrigation water based on the GLDAS 2.0 soil moisture were mainly analyzed in this paper. The reconstructed results based on the GLDAS 2.1 soil moisture can be found in the supplemental material.

a. Reconstructed ET from GLDAS 1.0

The reconstructed annual ET results from each of the GLDAS 1.0 LSMs are shown in Fig. 5, which shows that the ET_{recon} time series were close to one another, even though the original ET series were scattered. Particularly, the original CLM ET was much lower than the other three LSM ET values (Fig. 5a), which corresponded to a high overestimation of the runoff (Fig. S4a). The GLDAS 1.0 CLM was also found to underestimate the annual mean ET in all climate zones by Liu et al. (2016b). In our study, the magnitudes of the original GLDAS 1.0 LSMs ET and the three collected ET estimates were generally lower than those of the water balance ET (i.e., $\text{ET} = P - R - \text{TWSC}$), and the underestimations could be partially attributed to the ignored influences of water management (Pan et al. 2017). For instance, Castle et al. (2016) mainly attributed the underestimates of LSM ET ($21.0 \pm 12.3 \text{ km}^3 \text{ yr}^{-1}$) relative to the satellite-based ET (GRACE-derived and MODIS-based) to human-induced inputs ($22.0 \pm 5.5 \text{ km}^3 \text{ yr}^{-1}$), especially from irrigation that is not represented in the LSM ET. Zhang et al. (2018) verified the CDR ET using the observed $P - R$ and FLUXNET tower data, and attributed the disagreements between CDR ET and $P - R$ to the effects of water management on their estimated runoff over some basins. Moreover, it was reported that the eddy covariance technique, which is applied by FLUXNET, substantially underestimates λE (latent heat flux) (Wang and Dickinson 2012). More importantly, the reconstructed annual ET values were similar to the water balance ET (Fig. 5b), validating the robustness of the proposed reconstruction method. After reconstruction, the r and MAE indices were all improved for the four LSM ET_{recon} values compared with the original ET values. For example, r for the CLM ET was improved from 0.876 to 0.882, and the MAE was reduced from 90.12 to 11.90 mm yr^{-1} from 2003 to 2010.

In the year 2000, the ET_{recon} annual values derived from the four GLDAS 1.0 LSMs were excessively low, mainly because of the invalid inherent relationships of $\text{ET} \sim P$ in 2000 of the GLDAS 1.0 outputs rather than the reconstruction method. First, from the perspective of different GLDAS versions, the annual values of P and ET were respectively 499.24 and 397.67 mm under the GLDAS 1.0 Noah LSM, with an ET/P of 79.7%

FIG. 4. Comparison of the downscaled monthly irrigation water amounts (mm month^{-1}) based on the soil moisture from GLDAS 2.0 and 2.1 over the entire Yellow River basin.

(Fig. 6). However, they were 409.74 and 361.41 mm ($\text{ET}/P = 88.2\%$) for the GLDAS 2.0 Noah, and 423.71 and 397.01 mm ($\text{ET}/P = 93.7\%$) under the GLDAS 2.1 Noah. The original GLDAS 2.1 ET values are closest to the water balance estimates from 2003 to 2010. Second, from the perspective of the GLDAS 1.0 original outputs alone, the peak value in the precipitation of 2000 (Fig. 6a) was accompanied by a relatively low ET magnitude compared with the adjacent four years, especially for the CLM and VIC models (Fig. 5a), which is unreasonable. The GLDAS 1.0 forcing data switched several times, that is, the forcing data source in 2000 was different from those from 1994 to 1999 and from 2001 to present (Rui 2015). The abrupt shifts in the forcing data in 2000 may be largely responsible for the unreasonable water budget distribution. As indicated by Wang et al. (2016), the GLDAS 1.0 temperature had an abrupt shift and was greatly underestimated in 2000, and the snowfall had an obvious step change in 2000 with the value approximately doubled in China. In our study, similar phenomena in the temperature and snowfall were also found over the YRB (Fig. S5). Precipitation has been reported to, on average, account for 94%–95% of the ET interannual variability in dry basins worldwide, including the YRB (Ukkola and Prentice 2013). The interannual variation of the GLDAS 1.0 precipitation agreed well with the observation-based precipitation, excluding the year 2000 during which the GLDAS precipitation was also highly overestimated over China (Wang et al. 2016). Accordingly, the reconstructed ET became even lower due to the correction of the high overestimation in the precipitation ($\sim 24\%$, Fig. 6a and Fig. S6a). The phenomenon of a particularly low ET_{recon} value did not exist for the reconstructed results from GLDAS 2.0 and 2.1, as will be discussed later. In summary, the

rationality of the interannual variability in ET_{recon} also depended on the valid relationship of $\text{ET} \sim P$ in the LSMs, as ET results from a water–energy balance in which precipitation is the most important water supply.

Regarding the monthly time scale, the mean standard deviation of the four ET_{recon} series was $3.21 \text{ mm month}^{-1}$ over the YRB, which accounted for 8.5% of the mean monthly ET_{recon} . The peak values of the average LSMs ET_{recon} were generally higher than those of the Jung, CDR, and LandFlux-EVAL estimates, but they matched relatively well with those of the water balance ET (Fig. 7). All the statistics for the monthly ET_{recon} were improved compared with the original LSM ET and the three collected ET products, except for r for CDR (Fig. 8). Relative to the original LSM-based ET, r , the NSE, MAE, and RMSE for ET_{recon} were improved by 0.6%–1.8%, 1.2%–14.6%, 6.4%–21.0%, and 2.1%–20.4%, respectively. Importantly, according to the matched-pair Student's t test, the monthly ET series before and after reconstruction were significantly different at the 99% level. The indices in Fig. 8 were also estimated from the monthly ET series after removing the seasonal cycle to examine the interannual variability that is distributed to individual months. After reconstruction, the deseasonalized monthly GLDAS ET compared better with the deseasonalized water balance ET with all the indices being improved, especially for r and the NSE (which increased by 0.236–0.305 and 0.128–0.174, respectively) (Fig. 9). Furthermore, the performance of the CDR ET was better than those of Jung ET and LandFlux-EVAL for the monthly variability. In addition, the ET_{recon} results reconstructed using downscaled irrigation water based on the GLDAS 2.1 soil moisture were similar to those results that

FIG. 5. Comparisons of the evapotranspiration (mm yr^{-1}) among the (a) original and (b) reconstructed results from GLDAS 1.0, the three collected products (LandFlux-EVAL, CDR, and Jung), and the water balance estimate ($P - R - \text{TWSC}$).

were based on the GLDAS 2.0 soil moisture (Fig. S7).

b. Reconstructed ET from GLDAS 2.0

As shown in Fig. 10a, the GLDAS 2.0 ET_{recon} annual magnitudes were closer to the water balance ET than the original LSM results, with a reduction in the MAE from 54.10 to 24.86 mm yr^{-1} . The annual ET_{recon} did not properly capture the interannual variations of the water balance ET in 2003–10 because of the relatively high overestimations in 2005 and 2009. At first glance, the interannual variability of the original LSM ET compared relatively well with that of the water balance ET, but this is merely due to the year 2005. There was a relatively large difference in the precipitation interannual variations between the GLDAS 2.0 and the observation-based dataset, especially for 2001 and 2005 (Fig. 6b); thus, the original ET interannual variability was not retained after the correction of the precipitation for the annual series. Generally, the performances of LSM simulations and reanalysis-based products in

estimating the ET interannual variability are primarily related to three aspects: 1) whether the water and energy balances were considered reasonably; 2) model structural limitations; and 3) the accuracy of the forcing data (Liu et al. 2016b). In our study, the obtained ET_{recon} interannual variability relies on the variability of the GLDAS precipitation and the inherent partitioning of precipitation into ET and runoff in the LSM. Furthermore, we note that the Jung ET has nearly the smallest variability in the YRB. Liu et al. (2016b) also found that the Jung ET underestimated the interannual variability of the water balance ET in 35 global river basins. They partly attributed the underestimation to the neglect of the solar radiation in the ET estimation, together with the training of the Model Tree Ensemble method in Jung ET, using only 2-yr data from eddy-covariance sites on average. On the other hand, the r value in the monthly ET_{recon} after removing the seasonal cycle was obviously improved after reconstruction (Fig. 9), which is not exactly identical to the conclusion in Fig. 10a. This is because the deseasonalized time series reflects the monthly distribution of the interannual variability.

Similarly, as shown in Fig. 8, all the statistics for the monthly GLDAS 2.0 ET_{recon} (with the seasonal cycle) were improved relative to the original Noah ET. The NSE was increased from 0.76 to 0.80, and the RMSE was decreased from 13.43 to 12.27 mm month^{-1} . The ET_{recon} peaks were generally higher than those of the Jung, CDR, and LandFlux-EVAL estimates (Figs. 10b,c), which was consistent with the feature of monthly ET_{recon} from GLDAS 1.0. Furthermore, a new peak value appeared in May of 2001 of the GLDAS 2.0 ET_{recon} , which was due to the correction of the ~ 12 mm underestimation in May and the almost no bias in June in the GLDAS 2.0 precipitation (Fig. S6b). This result once again indicates that precipitation is a key factor in the reconstruction process. The performance of the ET_{recon} obtained using downscaled irrigation water based on the GLDAS 2.1 soil moisture was comparable to that based on the GLDAS 2.0 soil moisture (Fig. S8).

c. Reconstructed ET from GLDAS 2.1

Unlike the results from GLDAS 1.0 and 2.0, both ET_{recon} and original Noah ET annual values from GLDAS 2.1 were close to the water balance estimate (Fig. 11a), but ET_{recon} was relatively better in terms of statistics (Figs. 8 and 9). The reconstruction method improved the small interannual variability in the GLDAS 2.1 annual ET, with r increasing from 0.67 to 0.80. The NSE was improved from 0.39 to 0.51, MAE was reduced from 17.38 to 14.95 mm yr^{-1} , and RMSE was reduced from 18.78 to 16.79 mm yr^{-1} . Regarding the deseasonalized monthly ET_{recon} series, r increased

FIG. 6. (a)–(c) Evaluations of the precipitation in the land surface models from GLDAS 1.0, 2.0, and 2.1, respectively, using observation-based values over the entire Yellow River basin (mm yr^{-1}). (d) The ratios of evapotranspiration to precipitation ($\text{mm mm}^{-1} \text{yr}^{-1}$).

from 0.171 to 0.445 compared with the deseasonalized series before reconstruction. The peak values of monthly ET_{recon} (with the seasonal cycle) were also generally higher than those of the Jung, CDR, and LandFlux-EVAL ET estimates (Figs. 11b,c). The resulting ET_{recon} based on the GLDAS 2.1 soil moisture can be found in Fig. S9.

The results showed that the ET simulation in GLDAS 2.1 had been greatly improved compared with those of GLDAS 1.0 and 2.0, which was closely associated with the precipitation quality (Fig. 6 and Fig. S6). For instance, although GLDAS 2.0 and 2.1 ran the same LSM of Noah 3.3, a systematic deviation existed between the two original annual ET series, for which precipitation was an

FIG. 7. Comparisons of the monthly evapotranspiration (ET; mm month^{-1}) among the average reconstructed ET, the three collected products, and the water balance estimates. The census annual irrigation water was downscaled into monthly values using the soil moisture from GLDAS 2.0 during the reconstruction.

important factor (Fig. 6c). It is noteworthy that the simulated two runoff series from GLDAS 2.0 and 2.1 were both underestimated relative to the naturalized runoff (Figs. S4b,c), although much better results were obtained for the precipitation quality in GLDAS 2.1. This finding was consistent with the conclusion of Wang et al. (2016), that the Noah LSM had a defect of underestimating runoff, even if little bias was present in the forcing data. This finding justified the necessity for the runoff correction in the ET reconstruction process, as biased runoff simulation is an indirect indicator of the ET simulation. In our study, although the original ET datasets were from different GLDAS LSMs, the ET_{recon} results converged to the water balance estimate. The monthly ET series before and after reconstruction were significantly different at the 99% level of the matched-pair Student's t test, except those from GLDAS 2.1. In summary, the above analyses confirmed the robustness of our reconstruction method.

6. Importance of the GLDAS ET evaluation/correction and uncertainties in the water balance ET

a. Evaluation and correction of GLDAS ET

To the authors' knowledge, the monthly ET results from all GLDAS versions have been less evaluated.

Although Wang et al. (2016) examined the forcing data and Noah simulations from GLDAS 1.0 and 2.0 over China, they compared the GLDAS-simulated monthly ET + TWSC with the observed $P - R$ instead of evaluating ET directly. Liu et al. (2016b) conducted a worldwide evaluation of nine ET products, including ET from Noah of GLDAS 2.0 and CLM/Mosaic of GLDAS 1.0, but the evaluation focused on the annual time scale. Pan et al. (2017) noted the underestimates in P and ET and the high biases in the runoff in four GLDAS 1.0 LSMs, and they used the difference between the GRACE-based water balance ET and GLDAS 1.0 ET, without any correction, as the human-induced ET estimate. Therefore, it is very important to identify whether GLDAS can produce the natural hydrological processes accurately, as well as how to correct the water budget components if the answer is no. In our study: 1) the ET results from GLDAS 1.0, 2.0, and 2.1 were evaluated using the water balance estimate and the available multi-dataset synthesis products of ET at both the annual and monthly time scales; 2) the GLDAS precipitation and runoff were also examined with respect to the observations; and 3) the reconstructed natural and actual ET were obtained via bias corrections. The evaluations of the widely used GLDAS products are useful for people with interests in the datasets. Such evaluations in major global river basins are helpful for GLDAS to improve its simulations in the future from the aspects of both the forcing data and the LSM structure.

Compared with the GRACE-based water balance ET, the reconstructed ET of our study is not constrained by the time period or region size. The regional, mean annual ET was usually estimated by using the observed precipitation minus the runoff, with the assumption that $TWSC = 0$. After the launch of the GRACE satellites, monthly TWSA values have become available and monthly water balance ET estimates have become possible. Nevertheless, the monthly water balance ET (i.e., $P - R - TWSC$) can only be obtained over large enough regions starting from 2002, because the spatial resolution of TWSA is coarse and the time span is short starting from 2002. As stated earlier, monthly GRACE-TWSA data with sufficient accuracy are guaranteed over regions from $\sim 200\,000\text{ km}^2$ at low latitudes to $90\,000\text{ km}^2$ near the poles (Rodell et al. 2018). In this paper, only basin-mean reconstructed ET was analyzed due to the lack of gridded observational datasets. First, there is no spatial information in the collected naturalized streamflow because it was compiled using the streamflow observed at one gauge station, which

FIG. 8. Verification of the reconstructed evapotranspiration from GLDAS 1.0, 2.0, and 2.1 against the water balance ET estimates (i.e., $P - R - TWSC$) according to the (a) Pearson correlation coefficient, (b) Nash-Sutcliffe efficiency, (c) mean absolute error (mm month^{-1}), and (d) root-mean-square error (mm month^{-1}) from 2003 to 2010.

represents the entire hydrological regime above that station. Second, only the irrigation water information over the entire basin and a small district were available; thus, the performance of the spatial irrigation water simulation is still unknown. The conditions for spatial ET reconstruction are not yet fully met at the current stage. The gridded reconstructed ET can be realized with the following efforts in the future. The gridded natural streamflow with a high quality can be obtained through well-calibrated hydrological models, which depends on having sufficient collected naturalized streamflow data across the basin. Moreover, monthly irrigation water data at multiple irrigation districts need to be collected to verify the spatial irrigation simulation, although these data are very difficult to obtain.

b. Uncertainties in the water balance ET

The naturalized runoff was used in the reconstruction, while the gauged runoff was employed in the water balance ET estimation. These two runoff datasets were both provided by the Yellow River Conservancy Commission and can be regarded as the most reliable runoff data. The uncertainties in these two runoff datasets

were conservatively assigned to be 10% according to Pan et al. (2017), Pan et al. (2012), Landerer et al. (2010), and Rodell et al. (2004a). Following Lv et al. (2017), the precipitation uncertainty was estimated from the standard deviation ($\pm 2\sigma$) at a confidence level of 95%. The uncertainty in the average GRACE-TWSA (u_{TWSA}) was estimated using the leakage/measurement errors of the spherical harmonic solution and the released uncertainty for the JPL mascon solution. Specifically, according to the error propagation, u_{TWSA} was estimated as $u_{TWSA} = (1/4)\sqrt{3 \times u_1^2 + u_2^2}$, where u_1 and u_2 are the uncertainties in the spherical harmonic and mascon solutions, respectively. The TWSA uncertainties were estimated to be 12.4 and 19.3 mm month^{-1} over the YRB for the spherical harmonic solution and the JPL mascon solution, respectively. Second, given that the standard deviation is widely used in estimating uncertainty (Rodell et al. 2007; Long et al. 2014), u_{TWSA} was also estimated in another way, $u_{TWSA} = \pm 2\sigma$, where σ is the standard deviation of the four TWSA series. Third, the uncertainties in TWSA and P were also estimated using $\pm \sigma$ in a less-conservative estimation (Pan et al. 2017). The uncertainty in TWSC (u_{TWSC}) was computed as $u_{TWSC} = u_{TWSA} \times \sqrt{2}$. Finally, the

FIG. 9. As in Fig. 8, but for evapotranspiration after removing the seasonal cycle (* indicates significance at the 95% level of the Student's *t* test).

uncertainty in the water balance ET was obtained by combining the uncertainty terms of *P*, TWSA, and the runoff by propagation.

The monthly water balance ET and its uncertainties estimated in the above three ways are shown in Fig. 12, which shows that the uncertainties estimated using $\pm 2\sigma$ were the highest, and the four TWSA series became closer with a decreasing trend in the standard deviation. The average TWSA was used to reduce the associated uncertainties, so the uncertainty in the average result was supposed to be less than each TWSA dataset. Additionally, the uncertainty in the water balance estimated ET mainly came from the GRACE-derived TWSC (Fig. 13), in line with the findings of Lv et al. (2017), Pan et al. (2017), and Sheffield et al. (2009). Similarly, Wan et al. (2015) obtained ET from the GRACE-based water balance equation over the conterminous United States, and the RMSD of the multiple ET estimates used to evaluate the derived ET varied from 64.4 to 146.3 mm yr⁻¹, which was comparable to the uncertainty of GRACE-TWSC. Despite the uncertainties in the observed precipitation, runoff and GRACE-derived TWSC, the water balance ET integrated over a river basin remains “the most firmly observationally based estimator of ET” (Ukkola and Prentice 2013). Importantly, the GRACE Follow-On

mission will afford a small increase in the spatial resolution and accuracy (Flechtner et al. 2016; Rodell et al. 2018), and the TWSC-based water balance ET will be improved accordingly.

7. Conclusions and summary

An accurate estimation of evapotranspiration is essential for the understanding of land surface–atmosphere interaction, drought assessment, and water resources management. However, in situ measurements of ET for a large region are difficult to obtain, the currently available ET products have large uncertainties, and the irrigation effects are not well represented in the currently available ET products. We consider it potentially feasible to use high-quality precipitation and runoff data to constrain LSM ET. The reconstruction method was therefore proposed from a water balance perspective with consideration of irrigation at a monthly time scale. In this study, the monthly basin-averaged ET time series were reconstructed from six LSM runs in GLDAS 1.0, 2.0, and 2.1 using observation-based precipitation, naturalized streamflow, and census consumed annual irrigation water use. The YRB of China, which is influenced by intense irrigation, was employed as a case study due to data

FIG. 10. Evaluations of the reconstructed (a) annual (mm yr^{-1}) and (b),(c) monthly (mm month^{-1}) evapotranspiration results from GLDAS 2.0 with the downscaled monthly irrigation water amount based on the GLDAS 2.0 soil moisture.

availability. The method can also be applied to other basins.

To obtain the monthly ET_{recon} , the available census annual irrigation amount was first downscaled into individual months through monthly proportions simulated by a proposed soil-moisture-based offline irrigation scheme. Before the downscaling, this irrigation scheme was validated with respect to the census annual irrigated water amount and the reported seasonal and mean monthly irrigation proportions. Specifically, the mean annual absolute relative error in the simulated annual irrigation amount was less than 12.0%, and the correlation coefficient was significant at the 99% level of the Student's t test for the simulated mean monthly irrigation proportion. After reconstruction, the monthly ET_{recon} values were generally improved relative to the

FIG. 11. As in Fig. 10, but for evapotranspiration reconstructed from GLDAS 2.1, beginning in 2000.

original LSM-based ET, with improvements of r , the NSE, the MAE, and the RMSE by 0.6%–1.8%, 1.2%–14.6%, 1.3%–21.0%, and 2.1%–20.4%, respectively. Furthermore, the monthly ET_{recon} results were superior to the collected three ET synthesis products in terms of statistics. The peak values of the monthly ET_{recon} were closer to the water balance estimates than those of the three collected synthesis products. Regarding the annual time scale, the values of ET_{recon} were close to the water balance ET, although the original LSM-based ET series were scattered. Overall, the interannual variability in ET_{recon} compared well with that of the water balance estimate. The interannual variability in ET_{recon} was highly associated with the precipitation variability and inherent partitioning of precipitation into ET and runoff in the LSM, as precipitation is the most important water supply.

FIG. 12. Uncertainties (mm month^{-1} ; shaded area) in the water balance estimate of the evapotranspiration (i.e., $ET = P - R - TWSC$), for which the uncertainty in TWSA was estimated in three ways: (a) released uncertainties for spherical harmonic and JPL mascon solutions, (b) two standard deviations ($\pm 2\sigma$) among the four TWSA datasets, and (c) one standard deviation ($\pm \sigma$) among the four TWSA datasets.

Despite the uncertainties in the observed precipitation, runoff, and GRACE-TWSC, the water balance ET integrated over a river basin remains the most reliable estimator of ET. Therefore, the water balance ET was used to evaluate the ET_{recon} values in this study. The downscaled monthly irrigation water amount was a source of uncertainty for ET_{recon} . The simulated mean monthly irrigation proportions were evaluated using census data from 1950 to 1980, which is earlier than the study period in this study. Therefore, the monthly irrigation water data needs to be collected to further evaluate and improve the proposed irrigation schemes. The reconstruction method of our study can provide an alternative estimation of the actual ET, and gridded

FIG. 13. Uncertainties (mm month^{-1}) in each of the water budget terms. The uncertainty in R was 10%, and the TWSC uncertainty was based on the leakage/measurement errors of the spherical harmonic solution and the released uncertainty for the JPL mascon solution. The uncertainties in P and $TWSC'$ were estimated using the ± 2 standard deviations σ . The uncertainties in P'' and $TWSC''$ were estimated based on $\pm \sigma$. The uncertainty in ET was calculated by propagating the errors of P , R , and $TWSC$.

reconstructed ET can also be obtained using this method if spatial high-quality runoff and irrigation data are available.

Acknowledgments. This study is jointly sponsored by the National Natural Science Foundation of China (41530532; 41705072), National Key R&D Program of China (2018YFA0606002), General Financial Grant from the China Postdoctoral Science Foundation (2016M601102), and Jiangsu Collaborative Innovation Center for Climate Change China. GRACE land data are available at <http://grace.jpl.nasa.gov>, supported by the NASA MEaSUREs Program. The data used in this study were acquired as part of the mission of NASA's Earth Science Division and archived and distributed by the Goddard Earth Sciences (GES) Data and Information Services Center (DISC). The used "Global Map of Irrigation Areas" was downloaded from the Food and Agriculture Organization of the United Nations at <http://www.fao.org/nr/water/aquastat/irrigationmap/index10.stm>. We are grateful to the anonymous reviewers for their useful suggestions to improve the paper.

REFERENCES

- An, W., Z. Li, S. Wang, X. Wu, Y. Lu, G. Liu, and B. Fu, 2017: Exploring the effects of the "Grain for Green" program on the differences in soil water in the semi-arid Loess Plateau of China. *Ecol. Eng.*, **107**, 144–151, <https://doi.org/10.1016/j.ecoleng.2017.07.017>.
- Baldwin, D., S. Manfreda, K. Keller, and E. A. H. Smithwick, 2017: Predicting root zone soil moisture with soil properties and satellite near-surface moisture data across the conterminous

- United States. *J. Hydrol.*, **546**, 393–404, <https://doi.org/10.1016/j.jhydrol.2017.01.020>.
- Beck, H. E., A. I. J. M. van Dijk, V. Levizzani, J. Schellekens, D. G. Miralles, B. Martens, and A. de Roo, 2017: MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrol. Earth Syst. Sci.*, **21**, 589–615, <https://doi.org/10.5194/hess-21-589-2017>.
- Cai, X., Z. L. Yang, Y. Xia, M. Huang, H. Wei, L. R. Leung, and M. B. Ek, 2014: Assessment of simulated water balance from Noah, Noah-MP, CLM, and VIC over CONUS using the NLDAS test bed. *J. Geophys. Res. Atmos.*, **119**, 13 751–13 770, <https://doi.org/10.1002/2014JD022113>.
- Castle, S. L., J. T. Reager, B. F. Thomas, A. J. Purdy, M.-H. Lo, J. S. Famiglietti, and Q. Tang, 2016: Remote detection of water management impacts on evapotranspiration in the Colorado River Basin. *Geophys. Res. Lett.*, **43**, 5089–5097, <https://doi.org/10.1002/2016GL068675>.
- Döll, P., H. Müller Schmied, C. Schuh, F. T. Portmann, and A. Eicker, 2014: Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites. *Water Resour. Res.*, **50**, 5698–5720, <https://doi.org/10.1002/2014WR015595>.
- Dong, B., and A. Dai, 2017: The uncertainties and causes of the recent changes in global evapotranspiration from 1982 to 2010. *Climate Dyn.*, **49**, 279–296, <https://doi.org/10.1007/s00382-016-3342-x>.
- Flechtner, F., K.-H. Neumayer, C. Dahle, H. Döbbslaw, E. Fagioli, J.-C. Raimondo, and A. Güntner, 2016: What can be expected from the GRACE-FO laser ranging interferometer for earth science applications? *Surv. Geophys.*, **37**, 453–470, <https://doi.org/10.1007/s10712-015-9338-y>.
- Fu, G., S. Chen, C. Liu, and D. Shepard, 2004: Hydro-climatic trends of the Yellow River basin for the last 50 years. *Climatic Change*, **65**, 149–178, <https://doi.org/10.1023/B:CLIM.0000037491.95395.bb>.
- , S. P. Charles, N. R. Viney, S. Chen, and J. Q. Wu, 2007: Impacts of climate variability on stream-flow in the Yellow River. *Hydrol. Processes*, **21**, 3431–3439, <https://doi.org/10.1002/hyp.6574>.
- Gao, G., D. Chen, C. Xu, and E. Simelton, 2007: Trend of estimated actual evapotranspiration over China during 1960–2002. *J. Geophys. Res.*, **112**, D11120, <https://doi.org/10.1029/2006JD008010>.
- Gao, H., Q. Tang, C. R. Ferguson, E. F. Wood, and D. P. Lettenmaier, 2010: Estimating the water budget of major US river basins via remote sensing. *Int. J. Remote Sens.*, **31**, 3955–3978, <https://doi.org/10.1080/01431161.2010.483488>.
- Gowda, P. H., J. L. Chávez, P. D. Colaizzi, S. R. Evett, T. A. Howell, and J. A. Tolk, 2007: Remote sensing based energy balance algorithms for mapping *ET*: Current status and future challenges. *Trans. ASABE*, **50**, 1639–1644, <https://doi.org/10.13031/2013.23964>.
- Guo, Y., 2007: *Irrigation and Drainage Engineering* (in Chinese). 3rd ed. China WaterPower Press, 29–53.
- Haddeland, I., D. P. Lettenmaier, and T. Skaugen, 2006: Effects of irrigation on the water and energy balances of the Colorado and Mekong river basins. *J. Hydrol.*, **324**, 210–223, <https://doi.org/10.1016/j.jhydrol.2005.09.028>.
- Hanasaki, N., S. Kanae, T. Oki, K. Masuda, K. Motoya, N. Shirakawa, Y. Shen, and K. Tanaka, 2008: An integrated model for the assessment of global water resources – Part 1: Model description and input meteorological forcing. *Hydrol. Earth Syst. Sci.*, **12**, 1007–1025, <https://doi.org/10.5194/hess-12-1007-2008>.
- , T. Inuzuka, S. Kanae, and T. Oki, 2010: An estimation of global virtual water flow and sources of water withdrawal for major crops and livestock products using a global hydrological model. *J. Hydrol.*, **384**, 232–244, <https://doi.org/10.1016/j.jhydrol.2009.09.028>.
- Jia, Z., S. Liu, Z. Xu, Y. Chen, and M. Zhu, 2012: Validation of remotely sensed evapotranspiration over the Hai River Basin, China. *J. Geophys. Res.*, **117**, D13113, <https://doi.org/10.1029/2011JC007124>.
- Jiménez, C., and Coauthors, 2011: Global intercomparison of 12 land surface heat flux estimates. *J. Geophys. Res.*, **116**, D02102, <https://doi.org/10.1029/2010JD014545>.
- Jung, M., M. Reichstein, and A. Bondeau, 2009: Towards global empirical upscaling of FLUXNET eddy covariance observations: Validation of a model tree ensemble approach using a biosphere model. *Biogeosciences*, **6**, 2001–2013, <https://doi.org/10.5194/bg-6-2001-2009>.
- , and Coauthors, 2010: Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature*, **467**, 951–954, <https://doi.org/10.1038/nature09396>.
- , and Coauthors, 2011: Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *J. Geophys. Res.*, **116**, G00J07, <https://doi.org/10.1029/2010JG001566>.
- Landerer, F. W., and S. C. Swenson, 2012: Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resour. Res.*, **48**, W04531, <https://doi.org/10.1029/2011WR011453>.
- , J. O. Dickey, and A. Güntner, 2010: Terrestrial water budget of the Eurasian pan-Arctic from GRACE satellite measurements during 2003–2009. *J. Geophys. Res.*, **115**, D23115, <https://doi.org/10.1029/2010JD014584>.
- Lawston, P. M., J. A. Santanello Jr., B. F. Zaitchik, and M. Rodell, 2015: Impact of irrigation methods on land surface model spinup and initialization of WRF forecasts. *J. Hydrometeorol.*, **16**, 1135–1154, <https://doi.org/10.1175/JHM-D-14-0203.1>.
- , —, T. E. Franz, and M. Rodell, 2017: Assessment of irrigation physics in a land surface modeling framework using non-traditional and human-practice datasets. *Hydrol. Earth Syst. Sci.*, **21**, 2953–2966, <https://doi.org/10.5194/hess-21-2953-2017>.
- Lei, M., Z. Jia, and S. Xiao, 2013: Study of agricultural development scale of the Yellow River Basin (in Chinese). *Yellow River*, **35**, 99–103.
- Lei, H., D. Yang, H. Yang, Z. Yuan, and H. Lv, 2015: Simulated impacts of irrigation on evapotranspiration in a strongly exploited region: A case study of the Haihe River basin, China. *Hydrol. Processes*, **29**, 2704–2719, <https://doi.org/10.1002/hyp.10402>.
- Leng, G., M. Huang, Q. Tang, H. Gao, and L. R. Leung, 2014: Modeling the effects of groundwater-fed irrigation on terrestrial hydrology over the conterminous United States. *J. Hydrometeorol.*, **15**, 957–972, <https://doi.org/10.1175/JHM-D-13-049.1>.
- , —, —, and L. R. Leung, 2015: A modeling study of irrigation effects on global surface water and groundwater resources under a changing climate. *J. Adv. Model. Earth Syst.*, **7**, 1285–1304, <https://doi.org/10.1002/2015MS000437>.
- Li, D., X. Jiang, Y. Wang, and H. Li, 2001: Analysis of calculation of natural runoff in the Yellow River basin (in Chinese). *Yellow River*, **23**, 35–37.

- Liu, Q., and Z. Yang, 2010: Quantitative estimation of the impact of climate change on actual evapotranspiration in the Yellow River Basin, China. *J. Hydrol.*, **395**, 226–234, <https://doi.org/10.1016/j.jhydrol.2010.10.031>.
- , T. R. Mcvicar, Z. Yang, R. J. Donohue, L. Liang, and Y. Yang, 2016a: The hydrological effects of varying vegetation characteristics in a temperate water-limited basin: Development of the dynamic Budyko-Choudhury-Porporato (dBCP) model. *J. Hydrol.*, **543**, 595–611, <https://doi.org/10.1016/j.jhydrol.2016.10.035>.
- Liu, W., L. Wang, J. Zhou, Y. Li, F. Sun, G. Fu, X. Li, and Y. F. Sang, 2016b: A worldwide evaluation of basin-scale evapotranspiration estimates against the water balance method. *J. Hydrol.*, **538**, 82–95, <https://doi.org/10.1016/j.jhydrol.2016.04.006>.
- Liu, Y., and Coauthors, 2013: Evapotranspiration and water yield over China's landmass from 2000 to 2010. *Hydrol. Earth Syst. Sci.*, **17**, 4957–4980, <https://doi.org/10.5194/hess-17-4957-2013>.
- Long, D., L. Longuevergne, and B. R. Scanlon, 2014: Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites. *Water Resour. Res.*, **50**, 1131–1151, <https://doi.org/10.1002/2013WR014581>.
- , and Coauthors, 2015: Deriving scaling factors using a global hydrological model to restore GRACE total water storage changes for China's Yangtze River Basin. *Remote Sens. Environ.*, **168**, 177–193, <https://doi.org/10.1016/j.rse.2015.07.003>.
- Lv, M., Z. Ma, X. Yuan, M. Lv, M. Li, and Z. Zheng, 2017: Water budget closure based on GRACE measurements and reconstructed evapotranspiration using GLDAS and water use data for two large densely-populated mid-latitude basins. *J. Hydrol.*, **547**, 585–599, <https://doi.org/10.1016/j.jhydrol.2017.02.027>.
- , —, and M. Lv, 2018: Effects of climate/land surface changes on streamflow with consideration of precipitation intensity and catchment characteristics in the Yellow River Basin. *J. Geophys. Res. Atmos.*, **123**, 1942–1958, <https://doi.org/10.1002/2017JD027625>.
- , —, M. Li, and Z. Zheng, 2019: Quantitative analysis of terrestrial water storage changes under the Grain for Green program in the Yellow River basin. *J. Geophys. Res. Atmos.*, **124**, 1336–1351, <https://doi.org/10.1029/2018JD029113>.
- Mao, Y., and K. Wang, 2017: Comparison of evapotranspiration estimates based on the surface water balance, modified Penman-Monteith model, and reanalysis datasets for continental China. *J. Geophys. Res. Atmos.*, **122**, 3228–3244, <https://doi.org/10.1002/2016JD026065>.
- Marcella, M. P., and E. A. B. Eltahir, 2014: Introducing an irrigation scheme to a regional climate model: a case study over West Africa. *J. Climate*, **27**, 5708–5723, <https://doi.org/10.1175/JCLI-D-13-00116.1>.
- Mu, Q., M. Zhao, and S. W. Running, 2011: Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sens. Environ.*, **115**, 1781–1800, <https://doi.org/10.1016/j.rse.2011.02.019>.
- Mueller, B., and Coauthors, 2011: Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations. *Geophys. Res. Lett.*, **38**, L06402, <https://doi.org/10.1029/2010GL046230>.
- , and Coauthors, 2013: Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis. *Hydrol. Earth Syst. Sci.*, **17**, 3707–3720, <https://doi.org/10.5194/hess-17-3707-2013>.
- Nie, W., B. F. Zaitchik, M. Rodell, S. V. Kumar, M. C. Anderson, and C. Hain, 2018: Groundwater withdrawals under drought: Reconciling GRACE and land surface models in the United States High Plains Aquifer. *Water Resour. Res.*, **54**, 5282–5299, <https://doi.org/10.1029/2017WR022178>.
- Pan, M., A. K. Sahoo, T. J. Troy, R. K. Vinukollu, J. Sheffield, and E. F. Wood, 2012: Multisource estimation of long-term terrestrial water budget for major global river basins. *J. Climate*, **25**, 3191–3206, <https://doi.org/10.1175/JCLI-D-11-00300.1>.
- Pan, Y., C. Zhang, H. Gong, P. J.-F. Yeh, Y. Shen, Y. Guo, Z. Huang, and X. Li, 2017: Detection of human-induced evapotranspiration using GRACE satellite observations in the Haihe River basin of China. *Geophys. Res. Lett.*, **44**, 190–199, <https://doi.org/10.1002/2016GL071287>.
- Pei, T., X. Wu, X. Li, Y. Zhang, F. Shi, Y. Ma, P. Wang, and C. Zhang, 2017: Seasonal divergence in the sensitivity of evapotranspiration to climate and vegetation growth in the Yellow River Basin, China. *J. Geophys. Res. Biogeosci.*, **122**, 103–118, <https://doi.org/10.1002/2016JG003648>.
- Pokhrel, Y. N., S. Koirala, P. J. F. Yeh, N. Hanasaki, L. Longuevergne, S. Kanae, and T. Oki, 2015: Incorporation of groundwater pumping in a global Land Surface Model with the representation of human impacts. *Water Resour. Res.*, **51**, 78–96, <https://doi.org/10.1002/2014WR015602>.
- Portmann, F. T., S. Siebert, and P. Döll, 2010: MIRCA2000-Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. *Global Biogeochem. Cycles*, **24**, GB1011, <https://doi.org/10.1029/2008GB003435>.
- Rodell, M., J. S. Famiglietti, J. Chen, S. I. Seneviratne, P. Viterbo, S. Holl, and C. R. Wilson, 2004a: Basin scale estimates of evapotranspiration using GRACE and other observations. *Geophys. Res. Lett.*, **31**, L20504, <https://doi.org/10.1029/2004GL020873>.
- , and Coauthors, 2004b: The Global Land Data Assimilation System. *Bull. Amer. Meteor. Soc.*, **85**, 381–394, <https://doi.org/10.1175/BAMS-85-3-381>.
- , J. Chen, H. Kato, J. S. Famiglietti, J. Nigro, and C. R. Wilson, 2007: Estimating groundwater storage changes in the Mississippi River basin (USA) using GRACE. *Hydrogeol. J.*, **15**, 159–166, <https://doi.org/10.1007/s10040-006-0103-7>.
- , E. B. McWilliams, J. S. Famiglietti, H. K. Beaudoin, and J. Nigro, 2011: Estimating evapotranspiration using an observation based terrestrial water budget. *Hydrol. Processes*, **25**, 4082–4092, <https://doi.org/10.1002/hyp.8369>.
- , J. S. Famiglietti, D. N. Wiese, J. T. Reager, H. K. Beaudoin, 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>.
- Ruan, B., and R. Fang, 1996: Characteristics and forecasting methods of irrigation water use in the Yellow River Basin (in Chinese). *Yellow River*, **5**, 45–49.
- Rui, H., 2015: Readme document for global land data assimilation system version 1 (GLDAS-1) products. Goddard Earth Sciences Data and Information Services Center, accessed July 2015, 32 pp., https://hydro1.gesdisc.eosdis.nasa.gov/data/GLDAS_V1/README.GLDAS.pdf.
- , H. Beaudoin, and C. Loeser, 2018: Readme document for NASA GLDAS version 2 data products. Goddard Earth Sciences Data and Information Services Center, accessed August 2018, 22 pp., https://hydro1.gesdisc.eosdis.nasa.gov/data/GLDAS/GLDAS_NOAH025_M.2.0/doc/README_GLDAS2.pdf.

- 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**, E1080–E1089, <https://doi.org/10.1073/pnas.1704665115>.
- Scholes, R. J., E. Brown de Colstoun, F. G. Hall, G. J. Collatz, B. W. Meeson, S. O. Los, and D. R. Landis, 2011: ISLSCP II Global Gridded Soil Characteristics. Oak Ridge National Laboratory Distributed Active Archive Center, accessed May 2017, <https://doi.org/10.3334/ORNLDAAAC/1004>.
- Sheffield, J., C. R. Ferguson, T. J. Troy, E. F. Wood, and M. F. McCabe, 2009: Closing the terrestrial water budget from satellite remote sensing. *Geophys. Res. Lett.*, **36**, L07403, <https://doi.org/10.1029/2009GL037338>.
- Sherman, L., 1932: Streamflow from rainfall by unit hydrograph method. *Eng. News Rec.*, **108**, 501–505.
- Siebert, S., J. Burke, J. M. Faures, K. Frenken, J. Hoogeveen, P. Döll, and F. T. Portmann, 2010: Groundwater use for irrigation—A global inventory. *Hydrol. Earth Syst. Sci.*, **14**, 1863–1880, <https://doi.org/10.5194/hess-14-1863-2010>.
- , V. Henrich, K. Frenken, and J. Burke, 2013: Global map of irrigation areas version 5. Food and Agriculture Organization of the United Nations, <http://www.fao.org/nr/water/aquastat/irrigationmap/index10.stm>.
- Sriwongsitanon, N., and W. Taesombat, 2011: Effects of land cover on runoff coefficient. *J. Hydrol.*, **410**, 226–238, <https://doi.org/10.1016/j.jhydrol.2011.09.021>.
- Syed, T. H., J. S. Famiglietti, and D. P. Chambers, 2009: GRACE-based estimates of terrestrial freshwater discharge from basin to continental scales. *J. Hydrometeorol.*, **10**, 22–40, <https://doi.org/10.1175/2008JHM993.1>.
- Tang, Q., S. Peterson, R. H. Cuenca, Y. Hagimoto, and D. P. Lettenmaier, 2009: Satellite-based near-real-time estimation of irrigated crop water consumption. *J. Geophys. Res.*, **114**, D05114, <https://doi.org/10.1029/2008JD010854>.
- Tang, Y., Q. Tang, F. Tian, Z. Zhang, and G. Liu, 2013: Responses of natural runoff to recent climatic variations in the Yellow River basin, China. *Hydrol. Earth Syst. Sci.*, **17**, 4471–4480, <https://doi.org/10.5194/hess-17-4471-2013>.
- Ukkola, A. M., and I. C. Prentice, 2013: A worldwide analysis of trends in water-balance evapotranspiration. *Hydrol. Earth Syst. Sci.*, **17**, 4177–4187, <https://doi.org/10.5194/hess-17-4177-2013>.
- Wada, Y., D. Wisser, and M. F. P. Bierkens, 2014: Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources. *Earth System Dyn.*, **5**, 15–40, <https://doi.org/10.5194/esd-5-15-2014>.
- Wan, Z., K. Zhang, X. Xue, Z. Hong, Y. Hong, and J. J. Gourley, 2015: Water balance-based actual evapotranspiration reconstruction from ground and satellite observations over the conterminous United States. *Water Resour. Res.*, **51**, 6485–6499, <https://doi.org/10.1002/2015WR017311>.
- Wang, K., and R. E. Dickinson, 2012: A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. *Rev. Geophys.*, **50**, RG2005, <https://doi.org/10.1029/2011RG000373>.
- Wang, S., and Coauthors, 2015: Comparing evapotranspiration from eddy covariance measurements, water budgets, remote sensing, and land surface models over Canada. *J. Hydrometeorol.*, **16**, 1540–1560, <https://doi.org/10.1175/JHM-D-14-0189.1>.
- Wang, W., W. Cui, X. Wang, and X. Chen, 2016: Evaluation of GLDAS-1 and GLDAS-2 forcing data and Noah model simulations over China at the monthly scale. *J. Hydrometeorol.*, **17**, 2815–2833, <https://doi.org/10.1175/JHM-D-15-0191.1>.
- Watkins, M. M., D. N. Wiese, D. N. Yuan, C. Boening, and F. W. Landerer, 2015: Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons. *J. Geophys. Res. Solid Earth*, **120**, 2648–2671, <https://doi.org/10.1002/2014JB011547>.
- Yellow River Conservancy Commission, 2018: Yellow River Water Resources Bulletin. Accessed May 2018, <http://www.yrcc.gov.cn/>.
- Yuan, X., F. Ma, L. Wang, Z. Zheng, Z. Ma, A. Ye, and S. Peng, 2016: An experimental seasonal hydrological forecasting system over the Yellow River basin-Part 1: Understanding the role of initial hydrological conditions. *Hydrol. Earth Syst. Sci.*, **20**, 2437–2451, <https://doi.org/10.5194/hess-20-2437-2016>.
- Zhang, K., J. S. Kimball, R. R. Nemani, and S. W. Running, 2010: A continuous satellite-derived global record of land surface evapotranspiration from 1983 to 2006. *Water Resour. Res.*, **46**, W09522, <https://doi.org/10.1029/2009WR008800>.
- Zhang, X., Z. Xiong, and Q. Tang, 2017: Modeled effects of irrigation on surface climate in the Heihe River Basin, Northwest China. *J. Geophys. Res. Atmos.*, **122**, 7881–7895, <https://doi.org/10.1002/2017JD026732>.
- Zhang, Y., and Coauthors, 2018: A Climate Data Record (CDR) for the global terrestrial water budget: 1984–2010. *Hydrol. Earth Syst. Sci.*, **22**, 241–263, <https://doi.org/10.5194/hess-22-241-2018>.