Evaluating remotely sensed monthly evapotranspiration against water balance estimates at basin scale in the Tibetan Plateau

Wenbin Liu

ABSTRACT

Global evapotranspiration (ET) products, as compensation for eddy-covariance observations, provide useful data sources for understanding terrestrial water-energy budgets at different scales, especially for data-sparse regions. Here, we evaluated three remotely sensed ET products against water balance-based reference ET (ET_{WB}) in 16 river basins across the Tibetan Plateau (TP) on a monthly time scale from 1983 to 2011. The results indicated that ET_GLEAM performed the best overall across the 16 TP river basins in terms of the multi-year average and the interannual variability of monthly ET_{WB}, followed by ET_ZHANG and ET_CSIRO. The multi-year means of monthly ET_{WB} were better estimated overall by the three remotely sensed ET products rather than their interannual variability. However, the performances of the three ET datasets varied among different TP basins based on various evaluation criteria. The seasonal cycle of ET_{WB} was better captured by ET_GLEAM, ET_ZHANG and ET_CSIRO in the Yalong, Yangtze and Salween Basins and the upper Yellow River Basins rather than that in the Yulongkashi, Bayin and Brahmaputra River Basins. Overall, the ET_GLEAM performed relatively better than other datasets. The evaluation results will provide important references for us to select suitable datasets and to apply them in basin-scale water-energy budget studies in data-sparse regions.

Key words | actual evapotranspiration, GRACE, remote sensing, Tibetan Plateau, water balance

INTRODUCTION

Evapotranspiration (ET), which determines the partitioning of available water (precipitation) into surface runoff and groundwater recharge, as well as the partitioning of available energy into sensible and latent heat fluxes, has an essential role in terrestrial energy, water and carbon cycles (Trenberth et al. 2009). During the past few decades, the terrestrial ET was expected to increase with the intensified global hydrological cycles under climate warming (Huntington 2006; Song et al. 2017). Our knowledge about the magnitudes of ET changes and their spatial patterns at relatively larger spatial scales (i.e., basin, regional and continental scales) is still limited due to the sparse ET observation networks and the high spatial-temporal heterogeneity of ET (Xu & Singh 2005).

To compensate for in situ ET observations, a number of global ET products (i.e., remote sensing-based ET, land surface model simulated-ET and reanalysis outputted-ET) have been developed to help build an understanding of the regimes of ET variations at different scales and to facilitate the estimation of hydrological and energy components under the changing environment (Roderick & Farquhar 2011). The gridded water budget products should be carefully evaluated before their use in global and regional hydrological and energy budget studies (Xu & Chen 2005; Ma et al. 2017; Wang et al. 2017). There are two datasets that are always adopted for validating the gridded ET products, eddy-covariance (EC) in situ observations and the reference ET calculated from the water balance. The former enables...
us to evaluate ET products across numerous sites of diverse vegetation categories (Miralles et al. 2011; Zhang et al. 2015, 2016), but the sparse spatial coverage, relatively short period and the lack of energy balance closure in some EC sites limit its application in validating ET products at basin and regional scales. The water balance-based ET is an alternative reference for assessing ET products at the basin scale (Mueller et al. 2011; Rodell et al. 2011). For example, Zeng et al. (2014) developed a water balance-based global ET product and applied it to validate various global ET datasets. Liu et al. (2016a) evaluated nine gridded ET products against the water balance-estimated reference ET in 35 global river basins. Moreover, Li et al. (2017) assessed the hydrological performances of four categories of ET products using the water balance method in the middle Yellow River Basin. These studies usually used water balance-calculated ET from observed runoff, precipitation and satellite-derived (i.e., Gravity Recovery and Climate Experiment (GRACE)) water storage change (ΔS, this term can be neglected at the multi-year scale). However, the coarse resolution of GRACE-derived water storage change always limits the validation of global ET datasets in relatively smaller river basins (for example, the reliability of GRACE data will decrease when the area of interest is smaller than the GRACE footprint, ~4° × 4°) at annual/seasonal/monthly time scales.

The Tibetan Plateau (TP) is the highest plateau in the world, with an average elevation higher than 4,000 metres above sea level (Liu et al. 2018). It is also a vulnerable region under climate warming, with strong interactions among multi-spheres in the earth’s system (Yao et al. 2012; Liu et al. 2016b). Some major Asian rivers (i.e., the Mekong River, Brahmaputra River, Indus River, Yellow River and Yangtze River) originate from the TP, which serves as the ‘Asian water tower’ supporting the livelihoods of hundreds of millions of people in Asian countries (Immerzeel et al. 2010). However, the basic understanding of basin-scale water-energy budgets on the TP is still limited so far, due to the lack of in situ hydrometeorological observations (Ma et al. 2015a, 2015b). Some attempts have been made to evaluate satellite-based water-energy budget components (i.e., ET) to help create an understanding of the basin-scale hydrological regimes over the TP (Li et al. 2014; Liu et al. 2018). For example, Xue et al. (2013) evaluated several global ET products using precipitation minus runoff for two TP river basins (the upper Yellow River and Yangtze River) at an annual scale (they found that the changes of water storage in both basins to be negligible). Moreover, Li et al. (2014) assessed five gridded ETs using the water balance-based reference ET at a monthly scale in four big basins located on the north-eastern and central TP.

In this study, we evaluate three satellite-derived ET products (one has been but others have not been evaluated in Xue et al. (2013) and Li et al. (2014)) against water balance-calculated ET at a monthly time scale during the period of 1983–2011 in 16 TP river basins. The average elevation ranges from 1,650 to 5,015 m, while the drainage area ranges from 2,832 to 191,235 km² among the 16 basins. Some basins used (i.e., the Yulongkashi River and Brahmaputra River) in this study are located on the north-western and southern TP, the evaluation results of which can thus compensate the findings of Xue et al. (2013) and Li et al. (2014). Moreover, the downscaled GRACE-derived terrestrial water storage changes were used to calculate the reference ET, which enables us to evaluate global ET datasets in relatively smaller basins. The paper is organized as follows: the in situ and satellite-observed datasets and related approaches adopted in this study are described in the section ‘Methods and datasets’; the results of multiple satellite-retrieved ET evaluations are presented and discussed in the section ‘Results and discussion’. The uncertainty inherited from this study is also discussed in this section, and is followed by the final section ‘Conclusions’.

METHODS AND DATASETS

Data

In situ and satellite observations

In this study, we use observed monthly precipitation and runoff as well as GRACE-derived changes in terrestrial water storage (Tapley et al. 2004; Landerer & Swenson 2012) to calculate the reference ET (\(ET_{WB}\)) based on a basin-wide water balance approach. The monthly runoff data from 1982 to 2011 for 16 river basins across the TP (Figure 1 and Table 1) were obtained from the National
Hydrology Almanac of China. We used daily gridded precipitation (0.5° × 0.5°) from the Meteorological Science Data Sharing Service Network of the China Meteorological Administration (CMA) (http://cdc.cma.gov.cn/home.do), which interpolated observed precipitation from 2,472 meteorological stations using the ANUSPLIN software. The daily precipitation was summarized to the monthly scale and then clipped and averaged for each TP river basin.

The three latest global GRACE-retrieved terrestrial water storage anomaly and water storage change datasets were used to calculate the basin-wide changes in terrestrial water storage. The datasets were separately processed at the GeoForschungsZentrum (GFZ), the Jet Propulsion Laboratory (JPL) and the Center for Space Research at the University of Texas (CSR). All centres start off with identical GRACE Level-1 observations and further process them to Level-2 and Level-3 products. Different parameter choices and solution strategies (e.g., precise orbit determination, corrections for spacecraft accelerations not related to gravity changes) were explored by the three centres to help to understand the characteristics of the various approaches.

Table 1 | Main characteristics of the selected TP river basins

<table>
<thead>
<tr>
<th>ID</th>
<th>River name</th>
<th>Control station</th>
<th>Station altitude (m)</th>
<th>Drainage area (km²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Yulongkashi</td>
<td>Tongguziluoke</td>
<td>1,650</td>
<td>14,575</td>
</tr>
<tr>
<td>02</td>
<td>Bayin</td>
<td>Zelingou</td>
<td>4,282</td>
<td>5,544</td>
</tr>
<tr>
<td>03</td>
<td>Yellow-s1</td>
<td>Gadatan</td>
<td>3,823</td>
<td>7,893</td>
</tr>
<tr>
<td>04</td>
<td>Yellow-s2</td>
<td>Xining</td>
<td>3,225</td>
<td>9,022</td>
</tr>
<tr>
<td>05</td>
<td>Yellow-s3</td>
<td>Tongren</td>
<td>3,697</td>
<td>2,832</td>
</tr>
<tr>
<td>06</td>
<td>Yellow-s4</td>
<td>Tainhai</td>
<td>2,632</td>
<td>121,972</td>
</tr>
<tr>
<td>07</td>
<td>Yellow-s5</td>
<td>Jimai</td>
<td>4,450</td>
<td>45,015</td>
</tr>
<tr>
<td>08</td>
<td>Yangtze</td>
<td>Jiaoyugiao</td>
<td>2,599</td>
<td>67,514</td>
</tr>
<tr>
<td>09</td>
<td>Yellow-s6</td>
<td>Huangheyan</td>
<td>4,491</td>
<td>20,930</td>
</tr>
<tr>
<td>10</td>
<td>Yangtze</td>
<td>Zhimenda</td>
<td>3,540</td>
<td>137,704</td>
</tr>
<tr>
<td>11</td>
<td>Salween</td>
<td></td>
<td>3,000</td>
<td>72,844</td>
</tr>
<tr>
<td>12</td>
<td>Brahmaputra-s1</td>
<td>Pangduo</td>
<td>5,015</td>
<td>16,459</td>
</tr>
<tr>
<td>13</td>
<td>Brahmaputra-s2</td>
<td>Tangjia</td>
<td>4,982</td>
<td>20,143</td>
</tr>
<tr>
<td>14</td>
<td>Brahmaputra-s3</td>
<td>Gongbujiangda</td>
<td>4,927</td>
<td>6,417</td>
</tr>
<tr>
<td>15</td>
<td>Brahmaputra-s4</td>
<td>Nuxia</td>
<td>2,910</td>
<td>191,235</td>
</tr>
<tr>
<td>16</td>
<td>Brahmaputra-s5</td>
<td>Yangcun</td>
<td>3,600</td>
<td>152,701</td>
</tr>
</tbody>
</table>

The ‘s1-s6’ in the river name column refer to the sub-basins of Yellow River and Brahmaputra River.

Figure 1 | Spatial distribution of the selected river basins on the Tibetan Plateau.
The coarse spatial resolution (1.0° × 1.0°) of GRACE-derived ΔS normally constrains its applications in small- to medium-size catchments (Longuevergne et al. 2010; Long et al. 2015). To minimize the errors and uncertainty of GRACE-derived ΔS in relatively smaller basins, we first corrected the GRACE data based on a scaling factor approach to restore the signal losses stemming from the sampling and post-processing of GRACE data (Swenson et al. 2003). We then downscaled the corrected GRACE data to a 0.25° × 0.25° resolution using a model-based downscaling approach (Wan et al. 2015) by combining both the advantages of terrestrial water storage changes (TWSCs) derived from GRACE and simulated by Variable Infiltration Capacity (VIC) model (Zhang et al. 2014). In the model-based downscaling approach, the 0.25° VIC-simulated TWSC and 1° GRACE-derived TWSC were both aggregated to 4° to match with the reliable footprint of GRACE data. The weight matrix was calculated by matching the GRACE-derived and VIC-simulated TWSCs with the 4° spatial resolution, and it was then used for downscaling the 1° GRACE-derived TWSC to the 0.25° grid. The corrected and downscaled GRACE-ΔS datasets from the three processing centres were averaged, and then clipped and averaged for every TP river basin.

**Remotely sensed ET products**

Three remotely sensed ET products, namely, ET_ZHANG, ET_GLEAM and ET_CSIRO, were evaluated against the basin-scale water balance estimates in this study. ET_ZHANG is estimated through the model-based downsampling approach (Zhang et al. 2010, 2015) driven by MODIS data, satellite-based vegetation parameters and meteorological observations. Monthly ET_ZHANG with a 8-km spatial resolution during the period of 1983–2011 can be freely downloaded from http://luna.ntsg.umt.edu/data/ET_global_monthly/. ET_GLEAM (spatial resolution: 0.25°, time resolution: daily) is provided at https://www.gleam.eu/, which is estimated using a set of algorithms (i.e., the Priestley–Taylor approach was used in the calculation of potential evaporation) in the Global Land Evaporation Amsterdam Model (GLEAM). Different components of evapotranspiration, i.e., bare-soil evaporation, interception loss, transpiration, sublimation and open-water evaporation were estimated separately in ET_GLEAM using reanalysis net radiation and air temperature, satellite and gauged-based soil moisture, precipitation and snow water equivalent (Miralles et al. 2011). Estimations of potential evaporation for land fractions of short canopy, tall canopy and bare soil are converted into actual evaporation through estimates of root-zone soil moisture and a multiplicative evaporative stress factor based on the microwave vegetation optical depth. The Gash analytical model and an adaptation of the Priestley and Taylor equation were used to calculate interception loss and actual evaporation for regions covered by ice and/or snow and water bodies. In this study, the newly released GLEAM v3.1 ET data from 1983 to 2011 were adopted (Martens et al. 2017). Moreover, ET_CSIRO contains actual global monthly evapotranspiration at a 0.5° spatial resolution from 1981 to 2012, which can be downloaded from the link: https://data.csiro.au/dap/landingpage?pid:=csiro:17375&v=2&d=true. ET_CSIRO was estimated through the Penman–Monteith–Leuning (PML) model forced with in situ and satellite observations (Zhang et al. 2016). It should be noted that the evaluation was finally restricted to the period of 1983–2011 due to the availability of all observed and satellite-based data used in this study. All gridded datasets (including precipitation, downscaled GRACE data, ZHANG ET and GLEAM ET) used were interpolated uniformly to a spatial resolution of 0.5° based on a bilinear interpolation to make their inter-comparison possible, and then the averages were extracted for each of the TP basins.

**Methods**

**Water-balanced ET reference (ET\_WB)**

To assess the performances of three remotely sensed ET datasets, we use the monthly ET\_WB as the reference value. The monthly ET\_WB was calculated through the basin-scale water balance approach as follows:

\[
ET_{\text{WB}} = P - Q - \Delta S
\]

Here, \(P\) (mm) and \(Q\) (mm) are basin-wide precipitation and runoff. \(\Delta S\) (mm) is the terrestrial water storage change, which includes the changes in surface water, subsurface
water and groundwater. At the monthly time scale, the term of \( \Delta S \) in Equation (1) cannot be neglected due to the influences of snow cover change and human activities, such as agricultural water withdrawal and reservoir operation.

During the GRACE era from April 2002 to April 2015, the monthly \( \Delta S \) can be estimated directly from GRACE retrievals. The \( ET_{WB} \) can thus be estimated using the observed \( P \), \( Q \) and satellite-derived \( \Delta S \) during the period of 2003–2011. In the non-GRACE era, for example, the period of 1983–2002 in this study, we calculate the monthly \( ET_{WB} \) using a two-step bias-correction procedure (Li et al. 2014). We first define \( P \) minus \( Q \) as the biased ET (\( ET_b \)), compared to the \( ET_{WB} \) (the terrestrial water storage change was considered in its calculation at the basin scale using Equation (1)). The monthly \( ET_b \) (1983–2011) and \( ET_{WB} \) (2003–2011) time series were fitted separately using different gamma distributions, which has been certified as a reasonable probability function for modelling monthly ET series (Li et al. 2014; Liu et al. 2016a). We then bias-corrected the \( ET_b \) series using the inverse function \( F^{-1}(\cdot) \) of the gamma cumulative distribution (CDF, \( F(\cdot) \)) by matching the cumulative probability between the two CDFs (Liu et al. 2016a, 2018; Liu & Sun 2017):

\[
ET_c(m) = F^{-1}(F(ET_b(m)|\alpha_b, \beta_b)|\alpha_{WB}, \beta_{WB})
\]

\[
ET_{wb}(m) = \frac{ET_b(a)}{ET_c(a)} \times ET_c(m)
\]

where \( ET_c(m) \) is the bias-corrected monthly \( ET_b, \alpha_b, \beta_b \) and \( \alpha_{WB}, \beta_{WB} \) are the shape and scale parameters of the gamma distribution for \( ET_b \) and \( ET_{WB} \), respectively. \( ET_c(m) \) can further be corrected to eliminate the annual bias using Equation (3). In Equation (3), \( ET_{wb}(m) \) means the final monthly reference ET after bias correction, while \( ET_b(a) \) and \( ET_c(a) \) represent the annual biased and corrected ET from the first step. The bias-correction procedure was applied to compute the \( ET_{wb}(m) \) series separately for each of the river basins across the TP. We finally regarded the calculated basin-scale \( ET_{wb}(m) \) series as \( ET_{WB} \) (they are almost the same during the GRACE era) during the period of 1983–2011 to evaluate the performances of satellite-derived ET datasets. It should be noted that the variables used (i.e., precipitation, runoff and terrestrial water storage change) in the \( ET_{WB} \) calculation based on the water balance equation are all basin-averaged, with units of mm.

### Evaluation criteria and non-parametric trend detection

Three evaluation criteria (i.e., Spearman’s rank correlation coefficient (\( CORR \)), root-mean-square-error (\( RMSE \)) and Nash–Sutcliffe efficiency coefficient (\( NSE \)) were applied to evaluate the three remotely sensed ET products against the reference ET (\( ET_{WB} \)). The \( RMSE \) and \( NSE \) are defined as:

\[
NSE = 1 - \frac{\sum_{i=1}^{n} (B_i - A_i)^2}{\sum_{i=1}^{n} (B_i - \bar{B})^2}
\]

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{n} (A_i - B_i)^2}{n}}
\]

where, \( A_i \) and \( B_i \) are the satellite-based ET product and \( ET_{WB} \), respectively. \( \bar{A} \) and \( \bar{B} \) are the means of \( A_i \) and \( B_i \) series, respectively. \( N \) indicates the total number of samples (total months in a series). The Spearman’s rank correlation coefficient is defined as the traditional Pearson correlation coefficient between the ranked variables. Here, \( A_i \) and \( B_i \) are first converted to ranks \( rgA_i \) and \( rgB_i \), respectively, and \( CORR \) can be computed using the following popular formula:

\[
CORR = \frac{\sum_{i=1}^{N} d_i / n(n^2 - 1)}
\]

which is the difference between the two ranks, \( rgA_i \) and \( rgB_i \).

The non-parametric Mann–Kendall’s test (MK test) was also used in this study to detect the annual and seasonal trends of ET for TP river basins (Kendall 1975). Before the use of the MK test, a pre-whitening procedure was applied to eliminate the influences of serial correlation for each ET series (Yue et al. 2002). The statistics of the MK test can be given as follows (Liu et al. 2013; Hu et al. 2017):

\[
Z = \begin{cases} 
\frac{S - 1}{\sqrt{Var(S)}} & \text{if } S > 0 \\
0 & \text{if } S = 0 \\
\frac{S + 1}{\sqrt{Var(S)}} & \text{if } S < 0 
\end{cases}
\]

\[
Var(S) = \frac{n(n - 1)(2n + 5) - \sum t(t - 1)(2t + 5)}{18}
\]
\[ S = \sum_{i=1}^{n-1} \sum_{k=i+1}^{n} \text{sgn}(x_k - x_i) \]  

(9)

\[ Sgn(\theta) = \begin{cases} 
1 & \text{if } \theta > 0 \\
0 & \text{if } \theta = 0 \\
-1 & \text{if } \theta < 0 
\end{cases} \]  

(10)

Here, \( x \) is the variable, a sample of \( n \) independent and identically distributed random variables, and \( t \) is the extent of any given tie. The magnitude of a trend, represented by Sen’s slope \(-\beta\) (Hirsch et al. 1982), was also used in this study:

\[ \beta = \text{Median} \left( \frac{x_i - x_j}{|i-j|} \right) \quad 1 < i < j < n \]

where \( \beta \) is the median of all possible combinations of pairs for the entire dataset (Gan 1998). Please refer to Yue et al. (2002) and Liu et al. (2013) for more details about the MK test.

RESULTS AND DISCUSSION

Satellite-based ET validation at monthly scales

We first compare the multi-year average and the interannual variability of three monthly remotely sensed ET products against \( ET_{WB} \) in the 16 TP river basins during the period of 1983–2011 (Figure 2). Overall, ET_GLEAM \( (R^2 = 0.58 \text{ and } R^2 = 0.35 \text{ for monthly mean and standard deviation (SD)), respectively}) \) performed relatively better than ET_ZHANG and ET_CSIRO in estimating both the multi-year average and interannual variability of monthly \( ET_{WB} \) for all basins. Compared with ET_CSIRO, ET_ZHANG \( (R^2 = 0.49 \text{ and } R^2 = 0.21 \text{ for monthly mean and SD}) \) showed better relative performances. The comparison results between ET_GLEAM and ET_ZHANG are consistent with those from 35 global river basins (Liu et al. 2016a). The assessment of ET_GLEAM and ET_ZHANG in the present study is also consistent with that of Liu et al. (2016a).

Although the overall performances of three remotely sensed ET datasets have been shown, they vary among the 16 TP river basins (Figure 3, Table 2). For example, in the Yulongkashi River Basin, ET_ZHANG produced an obvious overestimation, while ET_GLEAM and ET_CSIRO underestimated monthly \( ET_{WB} \). ET_ZHANG had the highest CORR \( (0.81) \), while ET_GLEAM had the highest NSE
(-0.16) and the lowest RMSE (11.53 mm/month) with ET_{WB} compared with the other two ET datasets (Figure 4). In the Bayin River Basin, although almost all satellite-based ET datasets underestimated monthly ET_{WB}, ET_GLEAM performed relatively better (CORR = 0.90, NSE = 0.67, RMSE = 14.61 mm/month). In the six sub-basins of the upper Yellow River, ET_GLEAM performed the best in the Yellow-s1 (NSE = 0.82, RMSE = 12.82 mm/month), Yellow-s2 (NSE = 0.83, RMSE = 12.86 mm/month), Yellow-s4 (NSE = 0.83, RMSE = 11.36 mm/month), Yellow-s5 (NSE = 0.78, RMSE = 13.42 mm/month) and Yellow-s6 (NSE = 0.79, RMSE = 11.69 mm/month) basins in terms of

![Figure 3](http://iwaponline.com/hr/article-pdf/49/6/1977/509899/nh0491977.pdf)

Figure 3 | Validation of three satellite-based monthly ET measurements against water balance estimates (ET_{wb}, mm/month) in 16 TP river basins during the period of 1983–2011.
NSE and RMSE. Considering the CORR, ET_CSIRO performed the best in the Yellow-s1 (0.91), Yellow-s2 (0.94), Yellow-s3 (0.90), Yellow-s4 (0.92) and Yellow-s5 (0.91) basins compared with the other two datasets. However, from the scatter plots of Figure 2, ET_ZHANG seems to be the best fit with the $y = x$ line relative to the other datasets.

The lower monthly ET values have been overestimated, while higher values have been underestimated by three remotely sensed ET products in the Yalong River and Salween River Basins. ET_CSIRO performed the best in the Yalong River Basin, with the highest CORR (0.81) and NSE (0.61) and the lowest RMSE (15.48 mm/month). In the Yalong River Basin, ET_ZHANG had the highest CORR (0.91), while ET_CSIRO had the highest NSE (0.74) and the lowest RMSE (17.26 mm/month) with monthly $ET_{WB}$ compared with other two ET datasets. In the Yangtze River Basin, ET_GLEAM had the highest CORR (0.71) while ET_ZHANG had the highest NSE (0.34) and the lowest RMSE (15.48 mm/month). In terms of CORR, ET_GLEAM performed the best in the Brahmaputra-s1 (0.71), Brahmaputra-s2 (0.69) and Brahmaputra-s5 (0.49) basins, while ET_CSIRO performed the best in the Brahmaputra-s4 (0.75) and Brahmaputra-s5 (0.81) basins. In the Brahmaputra-s1 and Brahmaputra-s2 basins, ET_ZHANG had a relatively higher NSE (0.34 and −0.01) and lower RMSE (15.48 and 17.98 mm/month) with monthly $ET_{WB}$ compared to the other two remotely sensed datasets. ET_GLEAM performed better in terms of NSE and RMSE relative to other datasets in the Brahmaputra-s3 (−0.40 and 23.07 mm/month), Brahmaputra-s4 (−1.43 and 17.98 mm/month) and Brahmaputra-s5 (0.49 and 12.91 mm/month) basins. The satellite-derived ET datasets evaluated in this study performed better overall in all 16 TP river basins except for in the Yulongkashi River Basin and some sub-basins in the Brahmaputra River Basin.

<table>
<thead>
<tr>
<th>River name</th>
<th>ET_ZHANG</th>
<th>ET_CSIRO</th>
<th>ET_GLEAM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CORR</td>
<td>RMSE</td>
<td>CORR</td>
</tr>
<tr>
<td>Yulongkashi</td>
<td>0.65</td>
<td>13.02</td>
<td>0.38</td>
</tr>
<tr>
<td>Bayin</td>
<td>0.86</td>
<td>21.74</td>
<td>0.89</td>
</tr>
<tr>
<td>Yellow-s1</td>
<td>0.89</td>
<td>14.36</td>
<td>0.90</td>
</tr>
<tr>
<td>Yellow-s2</td>
<td>0.91</td>
<td>12.89</td>
<td>0.90</td>
</tr>
<tr>
<td>Yellow-s3</td>
<td>0.85</td>
<td>16.43</td>
<td>0.88</td>
</tr>
<tr>
<td>Yellow-s4</td>
<td>0.89</td>
<td>13.24</td>
<td>0.90</td>
</tr>
<tr>
<td>Yellow-s5</td>
<td>0.87</td>
<td>14.37</td>
<td>0.90</td>
</tr>
<tr>
<td>Yalong</td>
<td>0.87</td>
<td>17.26</td>
<td>0.88</td>
</tr>
<tr>
<td>Yellow-s6</td>
<td>0.85</td>
<td>17.03</td>
<td>0.88</td>
</tr>
<tr>
<td>Yangtze</td>
<td>0.82</td>
<td>13.53</td>
<td>0.87</td>
</tr>
<tr>
<td>Salween</td>
<td>0.74</td>
<td>14.37</td>
<td>0.79</td>
</tr>
<tr>
<td>Brahmaputra-s1</td>
<td>0.63</td>
<td>15.48</td>
<td>0.66</td>
</tr>
<tr>
<td>Brahmaputra-s2</td>
<td>0.63</td>
<td>17.98</td>
<td>0.65</td>
</tr>
<tr>
<td>Brahmaputra-s3</td>
<td>0.42</td>
<td>25.22</td>
<td>0.44</td>
</tr>
<tr>
<td>Brahmaputra-s4</td>
<td>0.68</td>
<td>19.67</td>
<td>0.66</td>
</tr>
<tr>
<td>Brahmaputra-s5</td>
<td>0.74</td>
<td>14.03</td>
<td>0.71</td>
</tr>
</tbody>
</table>

Figure 4 | Comparison of Spearman’s rank correlation coefficient (a) Nash–Sutcliffe efficiency (NSE) (b) and root-mean-square-error (RMSE) (c) of monthly remotely sensed ET products against water balance-based ET in 16 TP basins. The basins are classified into four classes based on the aridity index ($AI = \text{precipitation}/\text{potential evaporation}$), i.e., arid ($0.03 < AI < 0.2$), semi-arid ($0.2 < AI < 0.5$), dry sub-humid ($0.5 < AI < 0.65$) and humid ($AI > 0.65$)
Overall, the performances of the three ET datasets did not show obvious differences among the four basin groups. For example, all ET products performed relatively better in semi-arid and dry sub-humid basins than in arid and humid basins in terms of CORR. However, similar results cannot be concluded in terms of NSE and RMSE.

Seasonal cycles and trends

We also compared the seasonal cycles of ET estimated by three satellite-based datasets against the water balance-based reference ET (Figure 5). In the upper Yellow River, ET_GLEAM estimated the intra-annual variation of ET...
better than the other datasets. Overall, ET_ZHANG overestimated while ET_CSIRO underestimated monthly ET (especially in summer months) in most sub-basins of the upper Yellow River. All satellite-based ET datasets underestimated (overestimated) summer $ET_{WB}$, but overall, ET_GLEAM (ET_ZHANG) performed relatively better than the other datasets in the Yalong, Bayin and Yangtze River Basins (Salween River Basin). ET_ZHANG estimated the seasonal cycle of $ET_{WB}$ better than the other two datasets in the Yulongkashi River Basin. ET_CSIRO even estimated a distinct phase of the seasonal cycle of ET relative to that shown by $ET_{WB}$. Approximately 25.27% and 35.95% of the area in the Yulongkashi River Basin are covered by glacier and snow (Liu et al. 2018), and the different parameterizations of snow/glacier in various ET products (e.g., the adaptation of the Priestley and Taylor equation used in ET_GLEAM) may cause the distinctions of their estimated seasonal cycles of ET in this basin. The overall performances of the three ET products are relatively worse in the Yulongkashi, Bayin and Brahmaputra River Basins compared with those in Yalong, Yangtze, and Salween Basins and the upper Yellow River Basins. All satellite-based ET products overestimated monthly $ET_{WB}$ in most months (especially in summer months) in the five sub-basins of the Brahmaputra River, while ET_GLEAM performed better overall than the other two gridded datasets.

The non-parametric trends in annual and seasonal ET were detected for $ET_{WB}$ and three satellite-derived ET datasets during the period of 1983–2011. Because most trends detected are insignificant at the 0.05 level, we only show the significant trends of annual/seasonal $ET_{WB}$ and compare them with those detected in satellite-derived ET datasets (Table 3). Specifically, both ET_GLEAM and ET_CSIRO simulated the significant increase of autumn $ET_{WB}$ in the Yellow-s3 basin. In the Salween River Basin, the increase of annual $ET_{WB}$ was also detected in ET_GLEAM and ET_CSIRO, but only ET_GLEAM showed a significant trend. Moreover, the increases of winter $ET_{WB}$ were captured by ET_ZHANG and ET_CSIRO in the Brahmaputra-s1 basin and by all three ET datasets in the Brahmaputra-s2 basin, but no significant trends were detected by the satellite-based ET products. In the Brahmaputra-s3 basin, only ET_ZHANG captured the significant decrease in summer $ET_{WB}$. Because of the diverse performances of the three satellite-based ET datasets, one can either use the dataset with the overall better performance (i.e., ET_GLEAM) in the basin-scale water-energy budget studies on the TP, or adopt the dataset with the best performance in the target basin based on the aims of the research (for example, based on whether the mean, interannual variability or trend of ET is the focus).

### Uncertainty

This study provides a useful reference for applying suitable satellite-retrieved ET products in basin-scale water-energy budgets studies on the TP. The various performances of the three ET datasets are mainly attributed to their different forcing datasets (e.g., temperature, radiation, precipitation, NDVI (Xue et al. 2013; Liu et al. 2016a)) and algorithms adopted (whether the water balance/energy balance, the process of canopy interception loss, soil moisture stress and vegetation process are considered or not (Miralles et al. 2011; Zeng et al. 2014)). Knowledge of the potential causes of various performances in these remotely sensed ET datasets is critical for improving the accuracy of these products, but it has been difficult to comprehensively investigate so far. It is thus beyond the scope

### Table 3

Mann–Kendall trends of annual/seasonal remotely sensed and water balance-estimated ET during the period of 1983–2011

<table>
<thead>
<tr>
<th>River name</th>
<th>Time scale</th>
<th>$ET_{WB}$</th>
<th>ET_ZHANG</th>
<th>ET_GLEAM</th>
<th>ET_CSIRO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yellow-s6</td>
<td>Autumn</td>
<td>0.446*</td>
<td>-0.065</td>
<td>0.266*</td>
<td>0.173*</td>
</tr>
<tr>
<td>Salween</td>
<td>Annual</td>
<td>0.180*</td>
<td>-0.109*</td>
<td>0.284*</td>
<td>0.049</td>
</tr>
<tr>
<td>Brahmaputra-s1</td>
<td>Winter</td>
<td>0.215*</td>
<td>0.039</td>
<td>-0.002</td>
<td>0.003</td>
</tr>
<tr>
<td>Brahmaputra-s2</td>
<td>Winter</td>
<td>0.287*</td>
<td>0.051</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>Brahmaputra-s3</td>
<td>Summer</td>
<td>-0.961*</td>
<td>-0.522*</td>
<td>0.380</td>
<td>0.449*</td>
</tr>
</tbody>
</table>

Only significant trends (at the 0.05 level, number labelled with *) in basin-wide $ET_{WB}$ at annual/seasonal scales were selected in the table.
of this study, but it deserves to be further explored in our future studies.

The main uncertainties in this study may be inherited from the calculation of ETWB (Han et al. 2015). First, we used the satellite-derived ΔS together with the observed P and Q in Equation (1) to calculate ETWB for relatively smaller river basins in the GRACE era (2003 afterward). The reliability of GRACE-derived ΔS will decrease when the area of interest is smaller than the GRACE footprint (∼4° × 4°). Therefore, we used the corrected and downscaled (0.25° × 0.25°) GRACE-derived ΔS in this study to enable its applicability in relatively smaller basins. We also compared the finalized GRACE-derived ΔS data from three different processing centres (Figure 6). Although the three seasonal

Figure 6 | Comparison of GRACE-derived terrestrial water storage changes (TWSC) processed at CSR, JPL and GFZ.
patterns of $\Delta S$ (i.e., Pattern I in the Yulongkashi River Basin, Pattern 2 in the Bayin, Yellow-s1, Yellow-s2, Yellow-s3, Yellow-s4 and Yellow-s5 River Basins, as well as Pattern 3 in the Yalong, Salween, Brahmaputra-s1, Brahmaputra-s2, Brahmaputra-s3, Brahmaputra-s4 and Brahmaputra-s5 River Basins) are shown, the values of $\Delta S$ from CSR, JPL and GFZ are close and consistent in time. It revealed the effectiveness and applicability of corrected and downscaled GRACE-derived $\Delta S$ in these small- to medium-size basins. Second, the two-step bias-correction approach, which was used to close the monthly basin-wide water balance by considering the impacts of $\Delta S$ empirically, may also inherit some uncertainties. In this approach, only the systematic biases induced by the variability of $\Delta S$ can be corrected. However, it has been certified effective to correct the biases of $P$ minus $Q$ and obtain the temporal variation (trends) of reference ET under the changing environments (Xue et al. 2013; Liu et al. 2016a, 2018). In addition, the interpolation of gridded datasets with various spatial resolutions and the downsampling procedure used for disaggregating the coarse resolution GRACE data into a finer resolution may also introduce certain uncertainties in this study. However, they are apparently the best choices currently to make the inter-comparison of multiple-resolution datasets possible in medium- to small-sized TP basins.

CONCLUSIONS

In this study, we comprehensively evaluated three remotely sensed ET products (i.e., ET_GLEAM, ET_ZHANG and ET_CSIRO) against $ET_{WB}$ calculated from the water balance in 16 river basins across the TP at a monthly scale during the period of 1983–2011. We found ET_GLEAM performed the best overall against the water balance-calculated $ET_{WB}$ across the 16 TP river basins in terms of the multi-year average and the interannual variability of monthly $ET_{WB}$, followed by ET_ZHANG and ET_CSIRO. The considerations for the processes of canopy interception loss and soil moisture stress included in the algorithm of ET_GLEAM may improve its accuracy in the TP. Compared with the multi-year average, the interannual variability of monthly ET was simulated relatively more poorly for all satellite-based ET products.

When evaluating these satellite-based ET datasets using multiple criteria (i.e., CORR, NSE and RMSE), we found various performances of different datasets among the 16 TP river basins. For example, in terms of NSE and RMSE, ET_GLEAM performed the best in the Yulongkashi, Bayin, Yellow-s1, Yellow-s2, Yellow-s4, Yellow-s5, Yellow-s6, Yangtze, Brahmaputra-s3, Brahmaputra-s4 and Brahmaputra-s5 River Basins, ET_ZHANG performed the best in Yalong, Brahmaputra-s1 and Brahmaputra-s2 River Basins, and ET_CSIRO performed the best in Yellow-s3 and Salween River Basins. The seasonal cycle of $ET_{WB}$ was better captured by the three remotely sensed ET products in the Yalong, Yangtze, and Salween Basins and the upper Yellow River Basins than that in Yulongkashi, Bayin and Brahmaputra River Basins. Overall, ET_GLEAM performed relatively better in simulating the intra-annual variation of $ET_{WB}$. Moreover, the significant increase/decrease trends of annual/seasonal $ET_{WB}$ can also be detected in some ET products and in a few basins (for example, ET_GLEAM and ET_CSIRO in simulating trend of autumn-$ET_{WB}$ in the Yellow-s River Basin). The results obtained will provide important references for us to choose and apply suitable remotely sensed ET products in basin-wide water-energy budgets studies on the TP.

ACKNOWLEDGEMENTS

This study was supported by the National Key Research and Development Program of China (2016YFC0401401 and 2016YFA0602402). The basin-wide water budget series in the TP Rivers used in this study are available from the author upon request (liuwb@igsnrr.ac.cn). We thank the Editor and six anonymous reviewers for their helpful comments.

REFERENCES


Immerzeel, W. W., van Beek, L. P. H. & Bierkens, M. F. P. 2010 Climate change will affect the Asian water towers. Science 328, 1382–1385.


First received 25 December 2017; accepted in revised form 16 June 2018. Available online 11 July 2018.