Journal of Hydrology 538 (2016) 82–95

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

A worldwide evaluation of basin-scale evapotranspiration estimates against the water balance method

Wenbin Liu [a], Lei Wang [b,c,*], Jing Zhou [b], Yanzhong Li [a], Fubao Sun [a], Guobin Fu [d], Xiuping Li [b], Yan-Fang Sang [a]

[a] Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China
[b] Key Laboratory of Tibetan Environmental Changes and Land Surface Processes, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China
[c] CAS Center for Excellence in Tibetan Plateau Earth Sciences, Beijing 100101, China
[d] CSIRO Land and Water, Private Bag 5, Wembley, Western Australia 6913, Australia

ARTICLE INFO

Article history:
Received 31 December 2015
Received in revised form 20 March 2016
Accepted 1 April 2016
Available online 8 April 2016
This manuscript was handled by
Konstantine P. Georgakakos, Editor-in-Chief

Keywords:
Global ET products
Water balance method
Köppen climate classification
Global Land Data Assimilation System
GRACE

SUMMARY

Evapotranspiration (ET) plays a critical role in linking the water and energy cycles but is difficult to estimate at regional and basin scales. In this study, we present a worldwide evaluation of nine ET products (three diagnostic products, three land surface model (LSM) simulations and three reanalysis-based products) against reference ET (ET_{wb}) calculated using the water balance method corrected for the water storage change at an annual time scale over the period 1983–2006 for 35 global river basins. The results indicated that there was no significant intra-category discrepancy in the annual ET estimates for the 35 basins calculated using the different products in 35 basins, but some products performed better than others, such as the Global Land surface Evaporation estimated using the Amsterdam Methodology (GLEAM_E) in the diagnostic products, ET obtained from the Global Land Data Assimilation System version 1 (GLDAS 1) with the Community Land Model scheme (GCLM_E) in LSM simulations, and ET from the National Aeronautics and Space Administration (NASA) Modern Era Retrospective-analysis for Research and Applications reanalysis dataset (MERRA_E) in the reanalysis-based products. Almost all ET products (except MERRA_E) reasonably estimated the annual means (especially in the dry basins) but systematically underestimated the inter-annual variability (except for MERRA_E, GCLM_E and ET simulation from the GLDAS 1 with the MOSAIC scheme – GMOS_E) and could not adequately estimate the trends (e.g. GCLM_E and MERRA_E) of ET_{wb} (especially in the energy-limited wet basins). The uncertainties in nine ET products may be primarily attributed to the discrepancies in the forcing datasets and model structural limitations. The enhancements of global forcing data (meteorological data, solar radiation, soil moisture stress and water storage changes) and model physics (reasonable consideration of the water and energy balance and vegetation processes such as canopy interception loss) will undoubtedly improve the estimation of global ET in the future.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Land evapotranspiration (ET) is an essential component in global water, energy, and carbon cycles, and provides a link between the atmosphere and the Earth's surface (Betts et al., 1996; Jiménez et al., 2011; Tang et al., 2014; Zhang et al., 2012, 2015). It is also an important indicator of hydrologic and heat variations under a changing climate and anthropogenic interference (Brutsaert and Parlange, 1998; Ohmura and Wild, 2002; Wang and Dickinson, 2012). Accurate quantification of ET is thus critical for understanding the hydro-climatologic processes and the interactions of the Earth system (Rodell and Famiglietti, 2002). However, the estimation of large-scale ET from ground-based measurements alone remains challenging due to the sparse network of point observations and the high spatial heterogeneity and temporal variability of ET (Xu and Singh, 2005; Xue et al., 2013). To address this limitation, a number of global ET products have been derived in recent years, including remote sensing-based products (Su, 2002; Mu

* Corresponding author at: Key laboratory of Tibetan Environmental Changes and Land Surface Processes, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, No. 16 Lincui Road, Chaoyang District, Beijing 100101, China.
E-mail addresses: liuwb@igsnrr.ac.cn (W. Liu), wanglei@itpcas.ac.cn (L. Wang), zhoujing@itpcas.ac.cn (J. Zhou), liyz.14b@igsnrr.ac.cn (Y. Li), Sunfb@igsnrr.ac.cn (F. Sun), Guobin.Fu@csiro.au (G. Fu), lixiuping@itpcas.ac.cn (X. Li), sangyf@igsnrr.ac.cn (Y.-F. Sang).

http://dx.doi.org/10.1016/j.jhydrol.2016.04.006
0022-1694/© 2016 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

et al., 2007, 2011; Zhang et al., 2009, 2010; Miralles et al., 2011b; Yang et al., 2013), reanalysis outputs (Simmons et al., 2006; Onogi et al., 2007), land surface model (LSM) simulations (Rodell et al., 2004a; Dirmeyer et al., 2006) and the estimates based on empirical upscaling of in situ observations (Jung et al., 2009).The available ET products have great potential for facilitating estimations of hydrological and energy components and their intrinsic hydro-climatic variability (Roderick and Farquhar, 2011). However, large-scale evaluation among different ET products, which is a prerequisite for their use in global and regional hydrological and energy budget studies, is constrained due to the lack of reference observations (Xu and Chen, 2005).

The global network FLUXNET enables scientists to assess terrestrial ET at different time scales across numerous sites of diverse vegetation types (Running, 1998; Wang and Dickinson, 2012; Tang et al., 2014). However, eddy-covariance (EC) ET measurements need to be treated with caution with respect to regional ET evaluations due to their relatively short period and sparse spatial coverage (particularly in the Southern Hemisphere and the tropics) as well as the lack of energy balance closure observed at some EC sites. An alternative approach is to compare ET products with the reference ET (ET_{wb}) calculated from the terrestrial water budget (observed precipitation P minus the sum of runoff Q and terrestrial water storage change ΔS at the basin scale) for closed basins (Swenson and Wahr, 2006; Sheffield et al., 2009). During the past two decades, a number of studies have focused on ET evaluation in the conterminous United States (Rodell et al., 2004a; Velpuri et al., 2013; Han et al., 2015), West Africa (Andam-Akorful et al., 2014), Tibetan Plateau (Xue et al., 2013; Li et al., 2014), and at the global scale (Ramillien et al., 2006; Rodell et al., 2011; Mueller et al., 2011; Zhang et al., 2010, 2012; Zeng et al., 2014) using the annual or monthly water balance. At the annual scale, for example, Zhang et al. (2010) evaluated the multiyear (1983–2006) averaged satellite-based global ET product (ZHANG_E) against the ET_{wb} derived from observed discharge and gauge-based precipitation (GPCC) and found ZHANG_E was generally in agreement with ET_{wb} in most global basins. Mueller et al. (2011) showed that the intra-category spreads were similar in seven global rivers when comparing the multiyear means (1989–1995) of existing ET products from observation-based datasets, reanalysis-based products, LSMs and Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4) simulations. Moreover, Zhang et al. (2012) found the decadal trends (1983–2006) in ET_{wb} did not match those in evaporation estimated from three simple diagnostic models, especially in 110 global wet basins. In the previous studies, the water storage changes were often neglected in ET_{wb} calculations based on annual-scale water balance. However, the water balance may not always close at an annual scale when ΔS is assumed unchanged in many global river basins due to the effect of snow thaw-melt (Dai et al., 2009; Lettenmaier and Milly, 2009) and anthropogenic impacts such as water diversion, reservoir regulation, agricultural irrigation (Condon and Maxwell, 2014). Therefore, a more comprehensive global reevaluation using the ET_{wb} estimates (considering the inter-annual variability of ΔS) as benchmark values is still imperative (Wan et al., 2015).

With the launch of the Gravity Recovery and Climate Experiment (GRACE) satellites in March 2002, the terrestrial water mass variations (which contribute significantly to the observed water storage change) could be reasonably inferred over sufficiently large regions (Wahr et al., 2004; Landerer and Swenson, 2012). Moreover, the influences of natural processes (e.g. glaciers, snow and frozen soil moisture) and anthropogenic interferences such as reservoir operations and inter-basin water transfers could also be reflected in GRACE-retrieved total water storage anomalies (TWSA). However, the temporal coverage of GRACE data is relatively short (2002 onward) for the validation of historical ET products. Recently, several attempts have been made to extend the TWSA or basin-scale ET_{wb} (considering the water storage change) series to the period before 2002 using empirical or statistical methods (Zeng et al., 2014; Li et al., 2014; Long et al., 2014). Although some uncertainties still exist in these methods (Sun, 2013), the calculated $ET_{wb}$is expected to be more accurate by considering the inter-annual variability of ΔS for the evaluations of global ET products. The objectives of this study are to (1) evaluate nine ET products including three diagnostic products, three LSM simulations and three reanalysis-based products against the reference ET (ET_{wb}) calculated from the bias-corrected water balance method (considering the ΔS) at the annual scale for 35 global river basins, and (2) discuss the potential influences of the forcing data on the different ET products. The paper is organized as follows: data collection and the methodology used in this study are described in Section 2. In Section 3, the evaluation results of the nine ET products are presented for wet and dry basins located in different climate zones. The potential impacts of the forcing data on the ET products are also discussed in this section. In the final section, we summarize the results and provide several recommendations for the improvements of the ET products.

2. Data and methods

2.1. Global ET products

Nine published global ET products (three diagnostic products, three LSMs simulations and three reanalysis-based products) were evaluated against ET_{wb} in this study (Table 1). The diagnostic products include (1) ZHANG_E (1983–2006) derived from the Numerical Terradynamic Simulation Group (http://www.ntsg.umt.edu/project/et), which was calculated using the modified Penman–Monteith method driven by MODIS data, meteorological observations and satellite-based vegetation parameters (Zhang et al., 2010); (2) JUNG_E (1981–2011), which integrated the point-wise ET observations at FLUXNET sites with geospatial information retrieved from the remote sensing and surface meteorological observations in a machine-learning algorithm (Jung et al., 2010) (https://www.bgc-jena.mpg.de/geodb/projects/Home.phs); and (3) GLEAM_E (Global Land surface Evaporation: the Amsterdam Methodology), which estimated three sources of land evaporation separately through different land surface types, namely, (1) bare soil, (2) short vegetation, and (3) vegetation with a tall canopy, using a set of algorithms (e.g. the Priestley-Taylor approach was applied for the calculation of potential evaporation) (Miralles et al., 2011a, 2011b). Moreover, the ice and snow sublimation in the pixels covered with snow was also estimated based on a separate routine (Miralles et al., 2011a). Three LSM simulations: GNOAH_E (from 1948 until present) was obtained from the Global Land Data Assimilation System version 2 (GLDAS-2) with the Catchment Noah scheme, and GCLM_E and GMOS_E, obtained from GLDAS-1with the Community Land Model and the MOSAIC schemes, respectively (Rodell et al., 2004b) (http://disc.sci.gsfc.nasa.gov/hydrology/data-holdings). Additionally, the reanalysis-based products included (1) JRA55_E (1958 onward) from the recently released Japanese 55-year reanalysis (JRA55) product (Kobayashi et al., 2015) (http://jra.kishou.go.jp/JRA-55/index_en.html); (2) ERAI_E (1979 onward) from the ERA-Interim global atmospheric reanalysis dataset (Berrisford et al., 2011) (http://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/), and (3) MERRA_E (1979 to present, Lucchesi, 2012) from the National Aeronautics and Space Administration (NASA) Modern Era Retrospective-analysis for Research and Applications (MERRA) reanalysis dataset (http://disc.sci.gsfc.nasa.gov/mdisc/).

Table 1
Overview of the ET datasets used in this study. The datasets were grouped into three categories based on Mueller et al. (2013), i.e. the diagnostic products, land surface model (LSM) simulations and the reanalysis-based products.

Category	ET products	Scheme description	Spatial resolution	Temporal resolution	Reference
Diagnostic	*ZHANG_E*	Modified Penman–Monteith	8 km	Monthly	Zhang et al. (2010, 2015)
	JUNG_E	Empirical (Model tree ensemble)	0.5° × 0.5°	Monthly	Jung et al. (2010)
	GLEAM_E	Priestley-Taylor	0.25° × 0.25°	Daily	Miralles et al. (2011a, 2011b)
Reanalysis	*JRA55_E*	SiB	T_L319L60 (~60 km)	3 hourly	Kobayashi et al. (2015)
	MERRA_E	GEOS-5 Catchment LSM	1.0° × 1.0°	Hourly/monthly	Lucchesi (2012)
	ERAI_E	TESSEL	0.5° × 0.5°	3 hourly/daily	Berrisford et al. (2011)
LSM	*GNOAH_E*	Penman–Monteith	1.0° × 1.0°	3 hourly	Rui (2011)
	GCLM_E	Monin–Obukhov similarity theory	1.0° × 1.0°	3 hourly	Rodell et al. (2004b)
	GMOS_E	Penman–Monteith	1.0° × 1.0°	3 hourly	Rodell et al. (2004b)

2.2. Global precipitation, streamflow and GRACE data

Global monthly precipitation used for the water balance calculation was obtained from the most recent Global Precipitation Climatology Centre (GPCC version 6, 1901 through near present) dataset from the Earth System Research Laboratory of the National Oceanic and Atmospheric Administration (NOAA) (http://www.esrl.noaa.gov/psd/data/gridded/data.gpcc.html), which comprises gridded datasets (spatial resolution: 0.5°) interpolated based on quality-controlled data from 67,200 stations around the world (Beck et al., 2005). Moreover, the CPC Unified gauge-based daily precipitation at 0.5° resolution (Chen et al., 2008, 1979–2005, https://climatedataguide.ucar.edu/climate-data/cpc-unified-gauge-based-analysis-global-daily-precipitation) and the Climate Research Unit (CRU) TS v3.23 precipitation at 0.5° resolution (Harris et al., 2014, 1901–2014, https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_3.23/) were also obtained for quantifying the uncertainties in the ET_{wb} calculation induced by different observation-based precipitation products. Monthly streamflow datasets and the drainage basin polygons were mainly collected from the Global Runoff Data Centre (GRDC, http://grdc.bafg.de/), which archives data from almost 9000 gauging stations around the world (Harding et al., 2014). A total of 35 river basins with drainage areas exceeding 40,000 km^2 (to minimize the uncertainties in the GRACE data for relatively smaller basins) and with the best available data during 1983–2006 were adopted in this study (Fig. 1 and Table 2). Moreover, the streamflow series for the Songhua River, Yellow River, Yangtze River and Pearl River were obtained from the National Hydrology Almanac of China, and the total observed system inflows to the Murray and Lower Darling system (Murray–Darling basin) were collected from the Murray–Darling Basin Authority of Australia. The basins were categorized into energy-limited (wet) and water-limited (dry) basins (Fig. 1) using the aridity index ($AI = E_p/P$, where E_p and P are annual basin-averaged potential ET and precipitation, respectively) (Ukkola and Prentice, 2013). The annual E_p and P for the 35 basins during the period 1983–2006 were extracted from the CRU TS v3.23 and GPCC datasets, respectively. Basins with $AI > 1.5$ were classified as "dry" and those with $AI \leqslant 1.5$ were classified as "wet" (Zhang et al., 2012).

The GRACE observations are currently the unique dataset for retrieving the terrestrial water storage change and its usefulness has been widely demonstrated in hydrological applications (Rodell and Famiglietti, 2002; Landerer and Swenson, 2012; Long et al., 2014) although its spatial resolution is relatively coarse (around a few hundred kilometers). The three latest global terrestrial water storage anomaly (TWSA and RL05) and water storage change datasets (all available on the GRACE Tellus website http://grace.jpl.nasa.gov/) are processed at the Center for Space Research at the University of Texas (CSR), the Jet Propulsion Laboratory (JPL) and the GeoForschungsZentrum (GFZ). To minimize the uncertainties of the ΔS estimates, the GRACE products from the three processing centers were initially averaged and the global ΔS series were then extracted for the period 2003–2013. Moreover, a glacial isostatic adjustment correction and a destriping filter were also applied to the data to reduce errors (Wahr et al., 2004).

2.3. Temperature, precipitation and radiation forcing datasets

To investigate the influences of the forcing data on the ET products, the temperature, precipitation and radiation forcing data were directly extracted from the inputs or outputs of the three land surface models and three reanalysis datasets (Table 3). For the diagnostic products, the temperature and radiation forcing datasets of ZHANG_E were downloaded from the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR, 1948 onward, 2.5° × 2.5°, http://www.cdc.noaa.gov/cdc/reanalysis/) (Kistler et al., 2001) and NASA World Climate Research Program/Global Energy and Water-Cycle Experiment (WCRP/GEWEX) Surface Radiation Budget (SRB) Release 3.0 dataset (1.0° × 1.0°, 1983–2007) (Sheffield et al., 2006). The temperature and precipitation forcing data of JUNG_E were obtained from the gauge-based CUR-PIK dataset (0.5° × 0.5°, 1901–2002, Mitchell and Jones, 2005) and GPCC. Moreover, the temperature, precipitation and radiation forcing data of GLEAM_E were extracted from the ISCCP + AIRS dataset (Rossow and Dueñas, 2004), CPC Unified precipitation and ERA Interim, respectively.

All gridded datasets (including ET, precipitation products, GRACE data and other forcing datasets) were uniformly aggregated to a common spatial (0.5° × 0.5°) and temporal resolution (monthly) in this study, to make inter-comparison possible. The annual evaluation of ET products was limited to 1983–2006 while the comparisons of forcing temperature, precipitation and radiation were limited to 1983–2002, 1983–2005 and 1983–2006, respectively, in order to ensure the evaluations were conducted over a comparable period for all corresponding datasets with various time lengths. The datasets were finally extracted for the 35 river basins using the basin boundaries obtained from the GRDC.

2.4. Methods

The hydrological water balance method was applied to estimate the reference ET (ET_{wb}) at the basin scale:

$$ET_{wb} = P - R - \Delta S, \quad (1)$$

where P and R are the basin-wide totals of precipitation (mm) and the net stream flow at the basin outlet (mm), respectively, ΔS is the change in terrestrial water storage including the surface, subsurface, and ground water changes (mm) at a monthly or annual scale. The method is simple and sound in theory, and warrants accurate ET estimates provided the other water components can be reasonably measured (Wan et al., 2015).

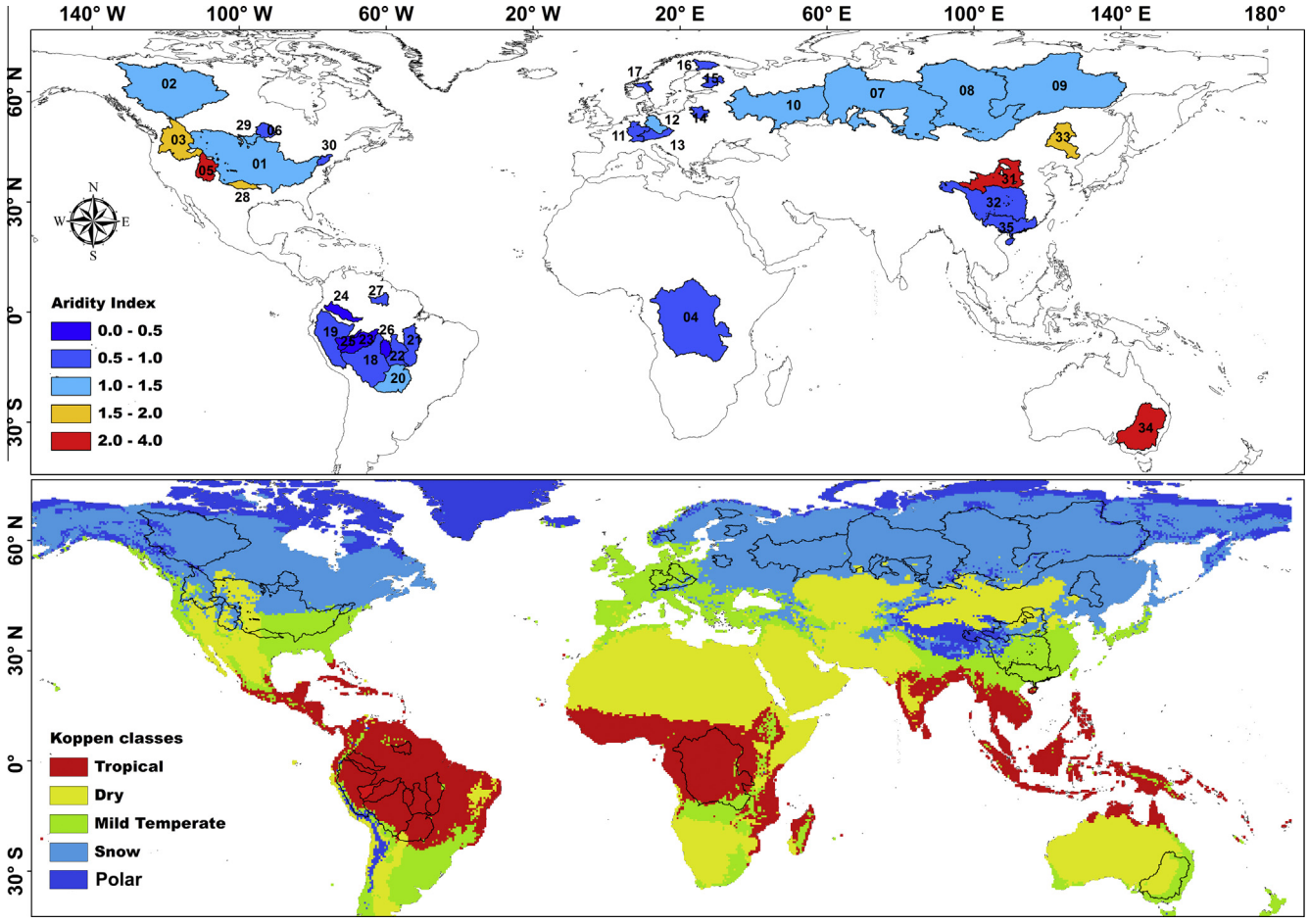

Fig. 1. Global distribution of the 35 river basins used in this study. The aridity index (*AI*) and the Köppen climate classes for each basin are shown in the upper and lower panel, respectively.

Table 2
Descriptions for the 35 global river basins used in this study.

No.	River name	Station	Drainage area (km^2)	No.	River name	Station	Drainage area (km^2)
01	Mississippi	Vicksburg, MS	2,964,255	19	Amazonas	Sao Paulo De Olivenca	990,781
02	Mackenzie	Arctic Red River	1,660,000	20	Rio Paraguay	Porto Murtinho (Fb/dnos)	474,500
03	Columbia	The Dalles, OR	613,830	21	Xingu	Altamira	446,203
04	Congo	Kinshasa	3,475,000	22	Rio Tapajos	Fortaleza	363,000
05	Colorado (Pacific Ocean)	Lees Ferry, Ariz	289,562	23	Rio Purus	Aruma-Jusante	359,853
06	Winnipeg	Slave Falls	126,000	24	Japura	Vila Bittencourt	197,136
07	Ob	Salekhard	2,949,998	25	Rio_Jurua	Gaviao	162,000
08	Yenisey	Igarka	2,440,000	26	Rio_Aripuana	Prainha Velha	131,000
09	Lena	Kyusyur (Kusur)	2,430,000	27	Rio_Branco	Caracarai	124,980
10	Volga	Volgograd Power Plant	1,360,000	28	Red_River (South)	Index	124,398
11	Rhine	Lobith	160,800	29	Red_River (of the North)	Grand Forks, N.D.	77,959
12	Elbe	Neu-Darchau	131,950	30	Susquehanna	Harrisburg, PA.	62,419
13	Danube	Bratislava	131,331	31	Yellow River	Huayuankou	730,036
14	Nemunas-Neman	Smalininkai	81,200	32	Yangtze River	Hankou	1,488,036
15	Vuoski	Tainionkoski	61,061	33	Songhua River	Harbin	391,000
16	Kemijoki	Isolaara	50,686	34	Murray-Darling	Goolwa	1,059,003
17	Glama	Langnes	40,540	35	Pearl River	Wuzhou	327,006
18	Rio Madeira	Manicore	150,000				

At the annual scale, ΔS is usually deemed negligible in many studies when one calculates the water balance ET (observed P minus Q at the basin scale) in unregulated basins, which assumes no water is gained or lost via deep groundwater or through inter-basin transfers, and precipitation (evaporation) is the only source (loss) of water in the basin (Hobbins et al., 2001; Zhang et al., 2012). However, the annual water balance may not close when ΔS is neglected under changing climate and anthropogenic interferences such as water diversion, reservoir regulation, and agricultural irrigation. In this study, we defined $P - R$ as the biased ET (ET_{biased}) relative to the reference ET (ET_{wb}).The ET_{biased} can be corrected based on ET_{wb} measured during the same period at the monthly scale through a two-step bias correction method (BCM, Li et al., 2014). Firstly, at each basin, the monthly ET_{wb} and

Table 3
List of the temperature, precipitation and radiation forcing datasets in different ET products.

ET products	Temperature	Precipitation	Radiation
Zhang_E	NCEP/NCAR	–	WCRP GEWEX SRB v3.0
Jung_E	CRU-PIK	GPCC	–
GLEAM_E	ISCCP + AIRS	CPC-Unified	ERA Interim
GNOAH_E	Princeton forcing		
GCLM_E	GLDAS-1 forcing (e.g. ADAS, GEOS,GDAS, ECMWF)		
GMOS_E	GLDAS-1 forcing (e.g. ADAS, GEOS,GDAS, ECMWF)		
JRA55_E	JRA forcing (e.g. ERA-40, JMA dataset, SYNOP and GMS, MTSAT data)		
ERAI_E	WATCH forcing data		
MERRA_E	MERRA-Land forcing data		

ET_{biased} series over the period 2003–2012 were fitted separately using a gamma distribution (Thom, 1958). This has been shown as an effective method for modeling the probability distribution of evapotranspiration (Bouraoui et al., 1999).

$$f(x|\alpha,\beta)=x^{\alpha-1}\frac{1}{\beta^{\alpha}\Gamma(\alpha)}e^{-\frac{x}{\beta}};\quad x\geqslant 0;\quad \alpha,\beta>0 \tag{2}$$

Here, α and β are the shape and scale parameters, respectively, of gamma distribution. The value of the monthly ET_{biased} series (2003–2012) can be corrected based on the inverse function F^{-1} of the gamma cumulative distribution function (CDF) F for the reference ET by matching the cumulative probabilities between two CDFs (Fig. 2). This procedure can be mathematically expressed as,

$$ET_{\text{corrected}}(m)=F^{-1}(F(ET_{\text{biased}}(m)|\alpha_{\text{biased}},\beta_{\text{biased}})|\alpha_{\text{reference}},\beta_{\text{reference}}) \tag{3}$$

$ET_{\text{corrected}}(m)$ and $ET_{\text{biased}}(m)$ represent the monthly corrected and biased ET, respectively. α_{biased}, β_{biased} and $\alpha_{\text{reference}}$, $\beta_{\text{reference}}$ are the parameters of the two gamma distributions for the monthly ET_{biased} and ET_{wb}, respectively. The second step in the bias correction is to eliminate the annual bias using the ratio of the annual ET_{wb} to the annual $ET_{\text{corrected}}$ obtained in the first step using the following equation:

$$ET_{\text{final}}(m)=\frac{ET_{\text{biased}}(a)}{ET_{\text{corrected}}(a)}\times ET_{\text{corrected}}(m) \tag{4}$$

where $ET_{\text{final}}(m)$ is the final monthly ET after bias correction, $ET_{\text{corrected}}(a)$ and $ET_{\text{corrected}}(m)$ represent the annual and monthly corrected ET, respectively, obtained from the first step, and $ET_{\text{biased}}(a)$ is the annual biased ET ($P-R$). This method was easily applied to correct the monthly ET_{biased} series for each basin during the period 1983–2006 when the GRACE data were not available. The $ET_{\text{final}}(m)$ series were calculated for all selected basins over the period 1983–2006 and were summed to the annual scale as the reference ET (ET_{wb}) for the global ET evaluation.

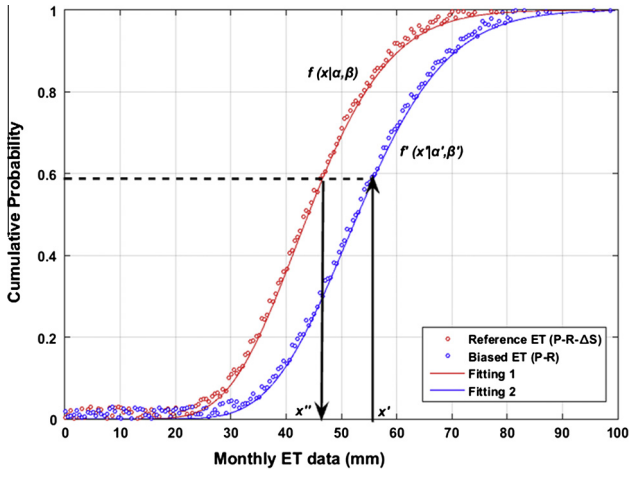

Fig. 2. Schematic diagram for the first step of the bias correction. The monthly reference ET ($P-R-\Delta S$), biased ET ($P-R$) and their fittings are exhibited in red and blue colors, respectively. x' and x'' showed the monthly ET values before and after the bias correction.

The Köppen climate classification system (Köppen, 1936), developed using an empirical relationship between climate and natural vegetation and which provides an efficient means for describing the climatic conditions defined only by temperature and precipitation, was adopted for grouping the basins into different climatic zones in this study. This classification system was recently updated by Chen and Chen (2013) using the newly global gridded monthly temperature and precipitation datasets and the classification results are freely available on the following website: http://hanschen.org/koppen/#home. In order to simplify the number of climates, only four major climate types, namely, tropical climates, mild temperate climates, dry climates and snowy climates, were used in this study. The map of the Köppen classification (long-term average climates from 1901 to 2010) was downloaded and overlapped on the selected global basins (Fig. 1). The climate type for a certain basin was then determined if one type was dominant in the basin (approximately more than half of the basin area was covered by this climate type). In addition, three evaluation criteria: root mean square error (RMSE), Pearson correlation coefficient (CORR) and the Taylor skill score (TaylorS) (Taylor, 2001) were used to evaluate the ET products against the reference ET (ET_{wb}). They are defined as:

$$\text{RMSE}=\sqrt{\sum_{i=1}^{n}(A_i-B_i)^2/N} \tag{5}$$

$$\text{CORR}=\frac{\sum_{i=1}^{N}(A_i-\overline{A})(\text{B}_i-\overline{B})}{\sqrt{\sum_{i=1}^{N}(A_i-\overline{A})^2}\sqrt{\sum_{i=1}^{N}(B_i-\overline{B})^2}} \tag{6}$$

$$\text{TaylorS}=\frac{4\times(1+\text{CORR})}{(\hat{\sigma}+1/\hat{\sigma})^2\times(1+\text{CORR}_0)} \tag{7}$$

where N indicates the total number of years, and A_i and B_i represent the ET product and ET_{wb}, respectively. Moreover, CORR_0 represents the maximum theoretical correlation ($\text{CORR}_0=1.0$ in this study), and $\hat{\sigma}$ is the standard deviation (σ_e) of a certain ET product normalized by the standard deviation (σ_r) of the ET_{wb} (σ_e/σ_r). The higher the Taylor skill score is (~1), the better the ET product performed (Taylor, 2001).

3. Results and discussion

3.1. Estimation of annual ET_{wb} during 1983–2006 using bias-corrected $P-R$

We first show the 10-year (2003–2012) means of the basin-averaged water storage change (ΔS) in the 35 global basins (Fig. 3). The multi-year mean values are close to 0 in almost all basins, which confirms the assumption that the ΔS can be neglected over a long period. However, the standard deviations of annual ΔS in 22 out of 35 basins exceed 50 mm/yr. This means that the information contained in the inter-annual variability of ΔS, caused by both climate change and anthropogenic interferences, can also be neglected when the annual ET is calculated using the annual P minus Q at the basin scale. By virtue of the two-step bias correction method, the annual changes in ΔS could be considered indirectly and thus the ET_{wb} estimations are expected to be

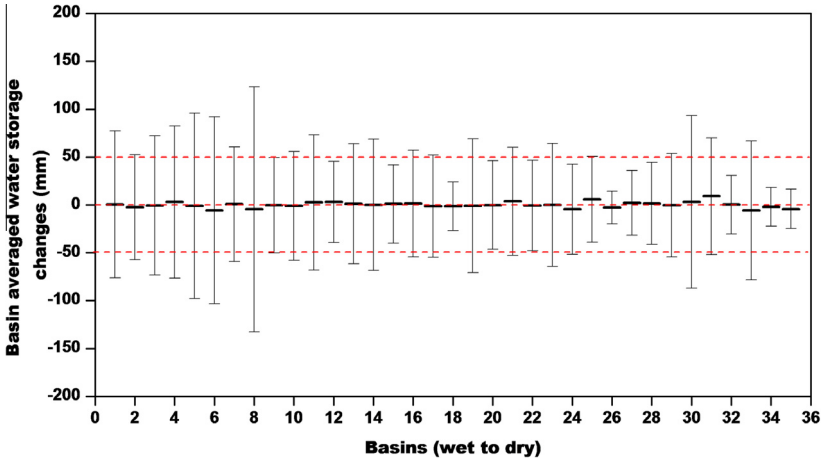

Fig. 3. Mean annual water storage changes (ΔS, bold line) over the period 2003–2012 for the 35 global basins. The error bar showed the standard deviation of annual ΔS for each basin.

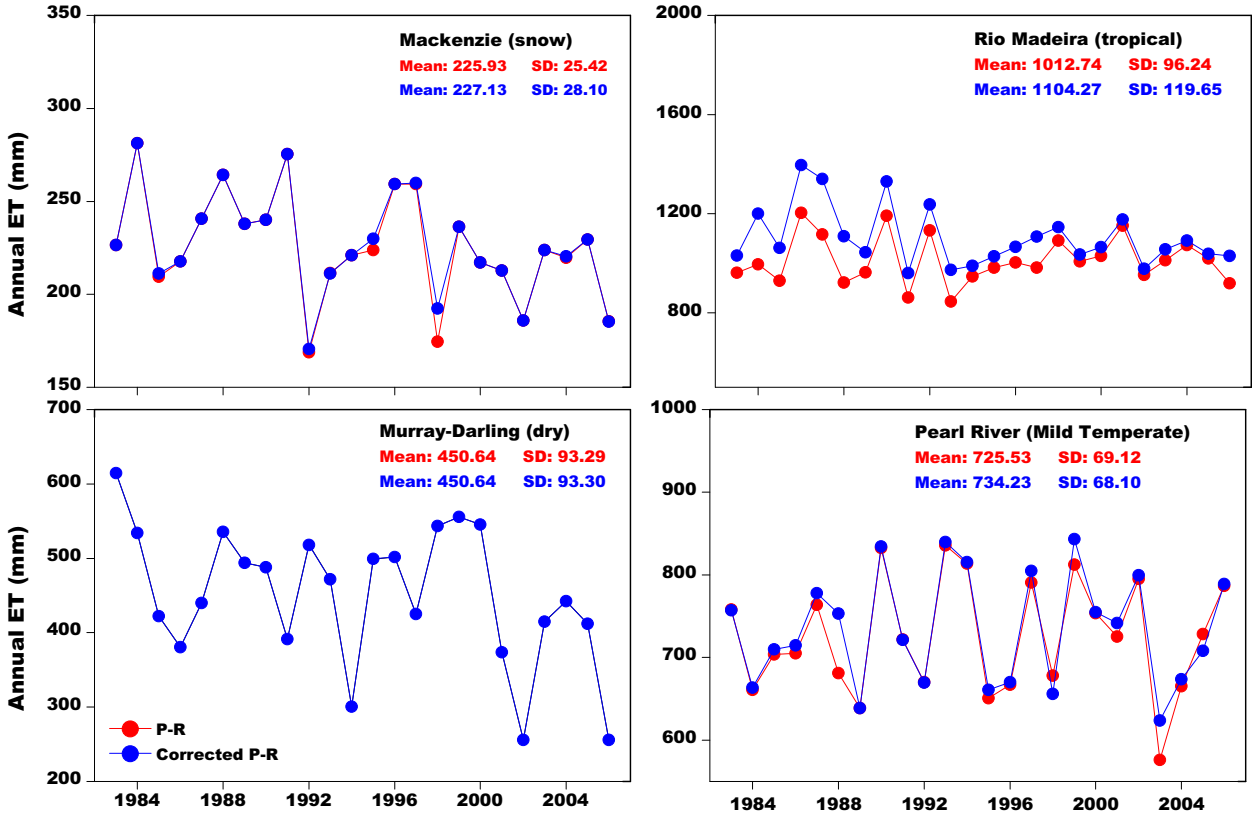

Fig. 4. Comparison of the original $P-R$ (ET_{biased}) and the bias-corrected $P-R$ (ET_{wb}) for the four basins located in different Köppen climates. The multiyear mean and standard deviation (SDs) are also shown for each basin.

improved (Zeng and Cai, 2016). In Fig. 4, the original $P-R$ (ET_{biased}) and bias-correct $P-R$ (ET_{wb}) in four basins located in different Köppen climate zones are compared. Relatively higher biases were corrected in Rio Madeira compared with Mackenzie River, Pearl River and Murray-Darling River. Similar results were obtained for basins in each climate zone (not shown). For example, relatively larger ET_{biased} errors were corrected in the tropical rivers (e.g. Amazonas, Xingu and Rio_Jurua), while in the rivers in dry climates (e.g. Colorado River, Columbia River and Yellow River), the annual ET_{biased} series remained almost uncorrected.

3.2. Evaluation of ET products in 35 global river basins

After calculating the annual reference ET (ET_{wb}) through the bias-corrected $P-R$ for all 35 basins, the ET products (three diagnostic products, three LSMs simulations and three reanalysis-based products) were then be evaluated against ET_{wb} during the period 1983–2006. The results are first shown based on the Taylor skill scores (TaylorS) in Fig. 5. The performances exhibit large overall spread between the different products and basins, for example, the TaylorS ranges from 0.2 to 1 for the Murray-Darling River

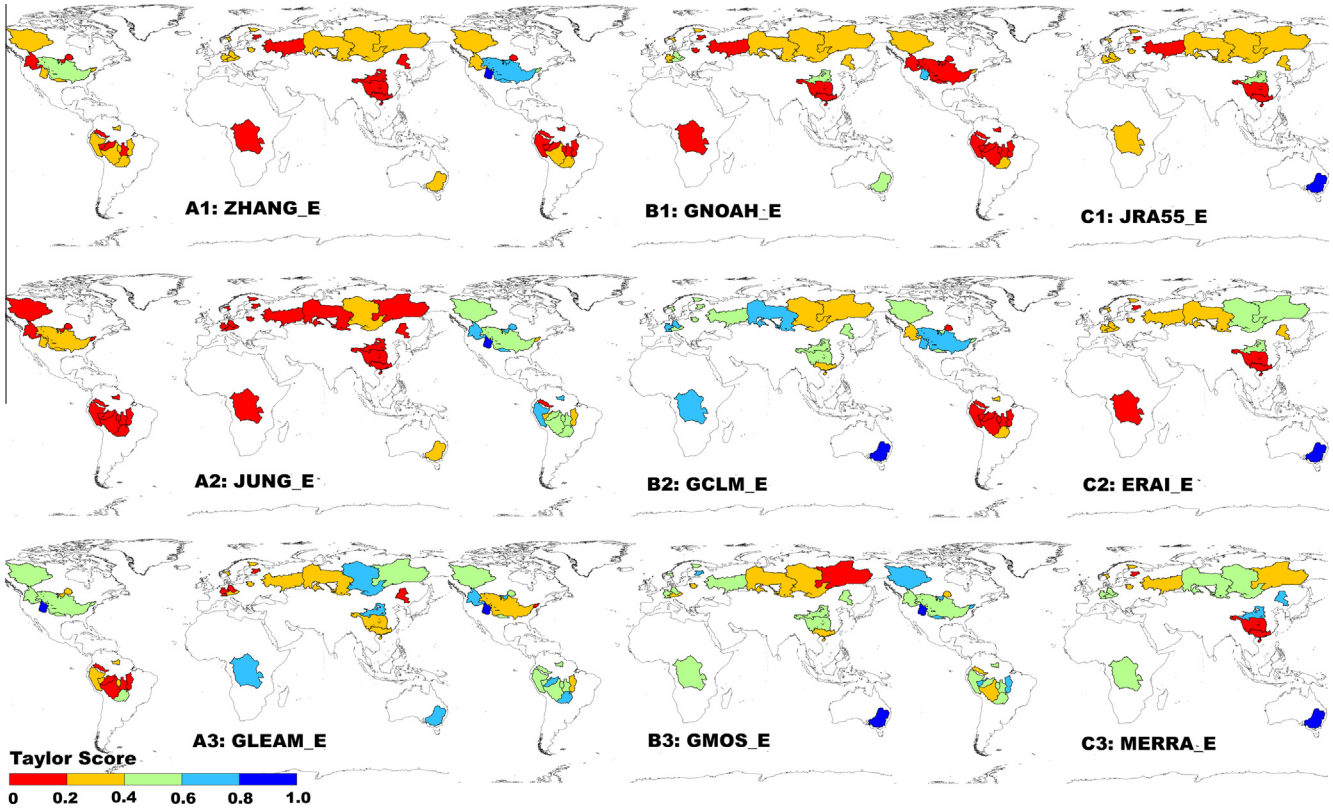

Fig. 5. The Taylor skill scores for different ET products in the 35 global river basins. The letter before the name of certain ET product shows the ET category (A-Diagnostic, B-LSMs and C-Reanalysis).

and from 0 to 0.8 for Congo River, which reveals considerable uncertainties between them (Dirmeyer et al., 2006). Similar to the results from Mueller et al. (2011) and Jiménez et al. (2011), based on a global analysis using satellite-based products, reanalysis, offline models and IPCC AR4 simulations, no obvious discrepancies were found between the three dataset categories. However, some individual ET products were found to perform relatively better than others in different categories and river basins (Fig. 5). For example, GLEAM_E showed relatively higher multi-basin mean for TaylorS (0.35 ± 0.23) in the diagnostic products compared with ZHANG_E (0.22 ± 0.10) and JUNG_E (0.08 ± 0.09) relative to the ET_{wb} during the period 1983–2006. Moreover, GCLM_E (0.52 ± 0.17) performed better than GNOAH_E (0.30 ± 0.22) and GMOS_E (0.46 ± 0.17) in the three LSM simulations while in the reanalysis-based products, the mean TaylorS in MERRA_E (0.49 ± 0.19) was higher than that in ERAI_E (0.30 ± 0.22) and JRA55_E (0.21 ± 0.18) in most global river basins such as the Congo River, Murray-Darling River and Ob River.

Figs. 6 and 7 display the annual means and the standard deviations (SDs) of the nine ET products against ET_{wb} during the period 1983–2006 for the wet and dry basins, respectively. Overall, almost all ET products reasonably estimated the annual means of ET_{wb} except for MERRA_E with relatively higher RMSEs. The results are consistent with some previous comparison studies (Zhang et al., 2010; Mueller et al., 2011; Jiménez et al., 2011). Moreover, the performances of the ET products with respect to the annual mean values of ET_{wb} were better (especially in JUNG_E), with higher multi-basin averaged CORR and lower RMSE in the dry rather than the wet basins (Figs. 6 and 8). However, all ET products underestimated the SDs except for GCLM_E, GMOS_E and MERRA_E, especially in the energy-limited wet basins. Comparing the three categories, the diagnostic products obviously underestimated the SDs of ET_{wb} in all basins, especially for JUNG_E and ZHANG_E. The inter-annual variability of ET is strongly controlled by the water and energy balance components such as precipitation, solar radiation, air temperature, wind speed, water storage changes as well as land surface variables (Ukkola and Prentice, 2013). The underestimation of inter-annual variations in ZHANG_E and JUNG_E may to some extent be attributed to the neglect of precipitation (which explains 66–88% of ET variability in wet basins compared with 96% in dry basins, Ukkola and Prentice, 2013) and solar radiation (e.g. the impact of the abrupt decrease in solar radiation from 1991–1992 due to the eruption of the Pinatubo volcano on the inter-annual variability of ET cannot be reflected) in their estimations (Mueller et al., 2011; Zhang et al., 2012; Zeng et al., 2014). Moreover, Zeng and Cai (2016) indicated that ET estimation without ΔS could not capture the ET variance at all for 25 arid basins and would underestimate ET inter-annual variance in humid climates. This may be another reason for the underestimation of the SDs in the three diagnostic products. The performances of the LSM simulations and reanalysis-based products in estimating the SDs of ET related mainly to whether the water and energy balances were reasonable considered, together with model structural limitations as well as the accuracy of the forcing datasets. In the reanalysis-based products, MERRA_E performed relatively well although it had substantial problems in describing the long-term change in ET (Su et al., 2015). The relatively better estimates (and further the TaylorS) of the inter-annual variability in GCLM_E and GMOS_E may be an illusion because of the obviously abnormal small values during the period 1996–1997 (Fig. 10; similar results were found in Mueller et al., 2013), which originated from changes in the GLDAS 1 forcing data (Matt Rodell, personal communication).

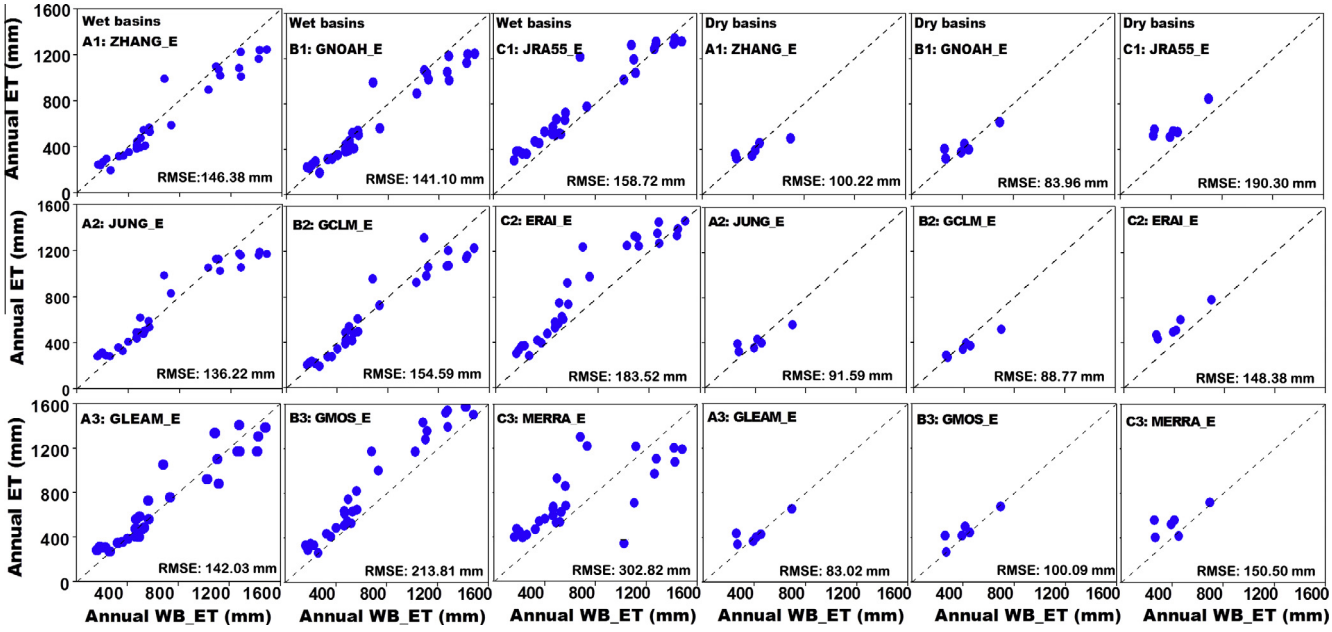

Fig. 6. Comparison of mean annual ET estimated from different products and those from ET_{wb} for the wet and dry basins. The multi-basin averaged RMSE is also shown for each ET product.

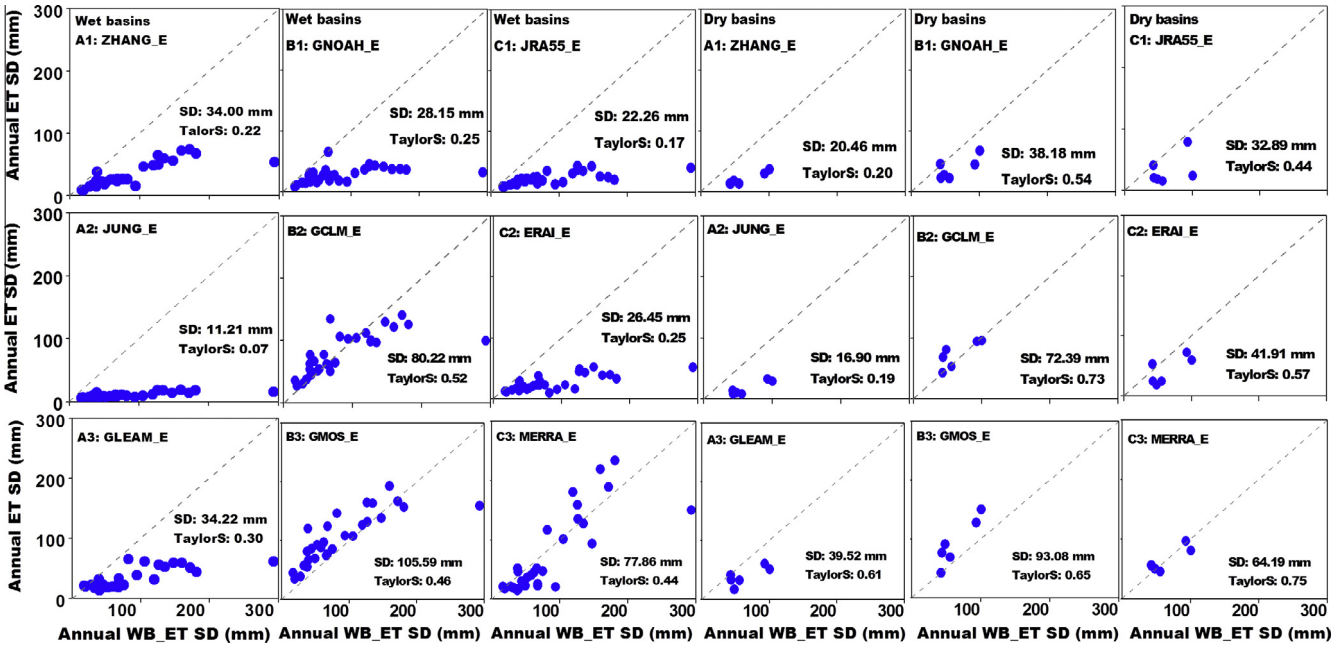

Fig. 7. Comparison of mean annual standard deviations (SDs) estimated from different ET products and those from ET_{wb} for the wet and dry basins. The multi-basin mean SD and Taylor skill score (TaylorS) are also exhibited for each ET product.

The linear trends of annual ET from the different products and basins were also evaluated against those from ET_{wb} in the wet and dry basins during the period 1983–2006 (Fig. 9). Overall, the ET products did not adequately explain the trends in annual ET_{wb} in the wet basins (especially GCLM_E and MERRA_E), which provides significant challenges for the diagnostic products, LSM simulations and the reanalysis-based products under the wet conditions. The performances of the ET products were relatively better in the dry basins compared with wet basins (especially for JRA55_E and ERAI_E) with respect to estimating the trends of annual ET_{wb}. The results are in line with the estimates given by Zhang et al. (2012) based on ET calculated from three satellite-based energy balance methods.

3.3. Evaluation of ET products in river basins located in different climate zones

The ET products were also compared in basins located in different Köppen climatic zones (Fig. 10).The RMSEs were generally higher in the tropical basins for all ET products (Mean AI = 0.62, e.g. Amazonas) relative to the dry (Mean AI = 2.85), snowy (Mean AI = 1.11) and mild temperate (Mean AI = 1.06) climates. This can be attributed to the sparseness of in-situ measurements (Wang and Dickinson, 2012) and the complexity of ET variation in tropical basins (e.g. the irregular seasonal cycles of ET under complicated vegetation–atmosphere interactions and the transitions between water and energy limitations in different seasons) (Hasler and

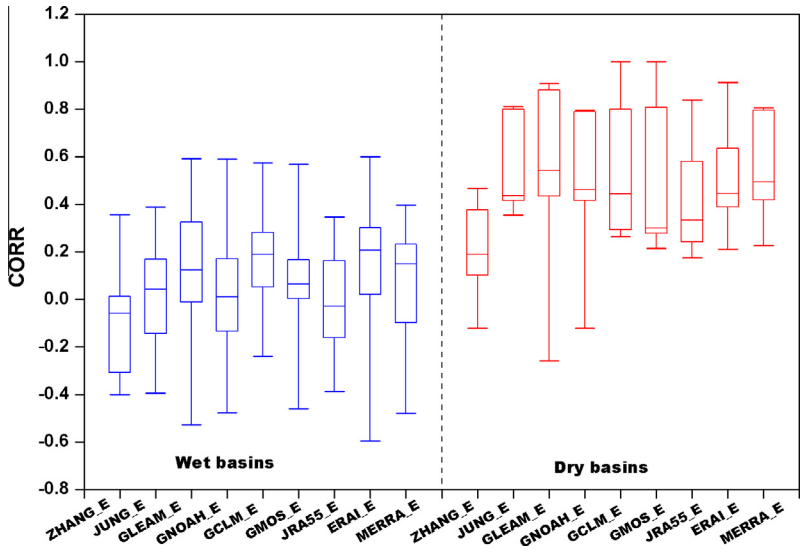

Fig. 8. Box plots for the correlation coefficients between different ET products and ET_{wb} for the wet and dry basins.

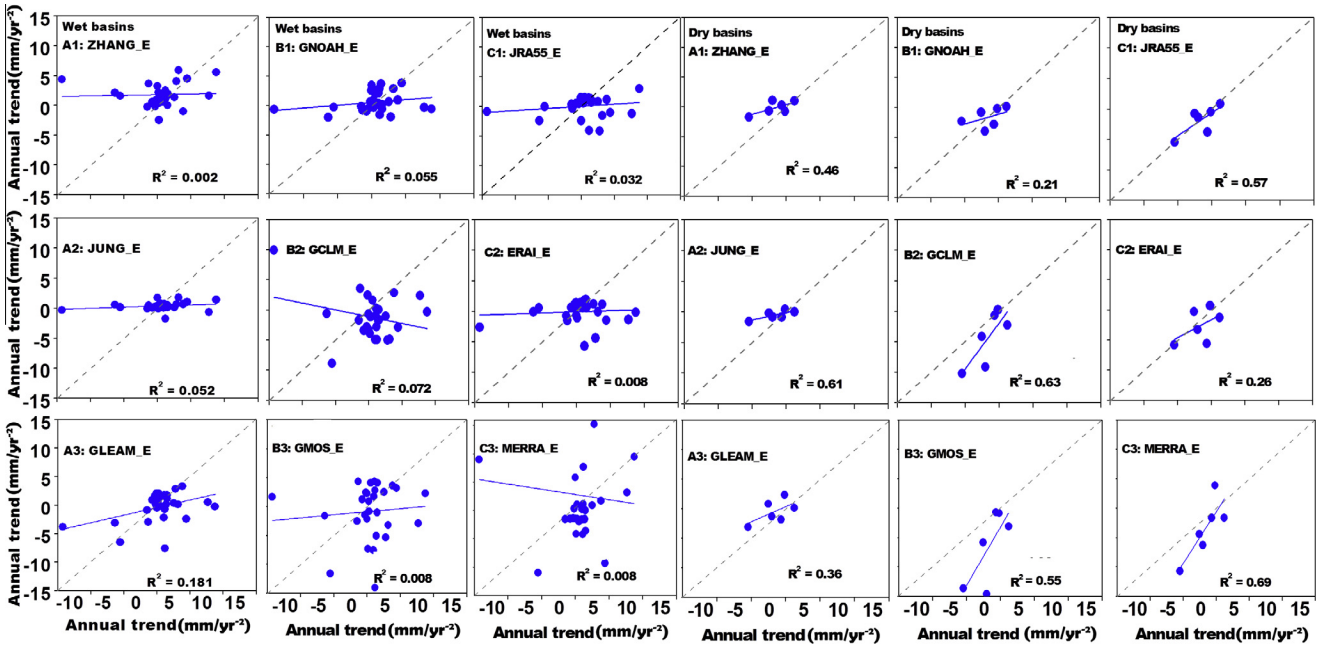

Fig. 9. Comparison of the annual trends in different ET products and those in ET_{wb} for the wet and dry basins.

Avissar, 2007; da Rocha et al., 2009; Rodell et al., 2011; Sahoo et al., 2011). The RMSEs of the annual ET were relatively larger (mainly overestimations) in the reanalysis-based products (especially for MERRA_E in the tropical and snowy basins), although many discrepancies remain among the individual ET products. In the LSM simulations, GMOS_E overestimated the annual mean ET in the dry, mild temperate and snowy basins, while GCLM_E underestimated the mean ET in all climate zones (data not shown). The diagnostic products performed relatively better than the other categories in the dry and mild temperate basins but underestimated the mean ET in the tropical basins. In term of the standard deviation, underestimations were found in GNOAH_E, JRA55_E, ERAI_E and the three diagnostic ET products in all four climate zones. Both GCLM_E and GMOS_E overestimated the SDs in the dry, mild temperate and snowy basins (Fig. 10). In addition, MERRA_E underestimated the SDs in the mild temperate and snowy basins but overestimated SDs in the tropical and dry basins. For example, the inter-annual variability of ET were reasonably estimated in the Murray-Darling River (dry climate) but were underestimated by most ET products in the Mackenzie River, Rio Madeira and Pearl River (Fig. 11).

3.4. Potential influences of the forcing data and other factors on the ET products

Annual ET estimations from different products might be influenced by various sources, for example, the input meteorological data, land cover, as well as model physics (Xue et al., 2013). Here, we mainly discuss the impacts of different meteorological forcing data on the ET products used in the 35 global river basins (Fig. 12) in terms of precipitation, temperature, and net radiation flux. Three datasets, namely, CRU-PIK temperature, GPCC

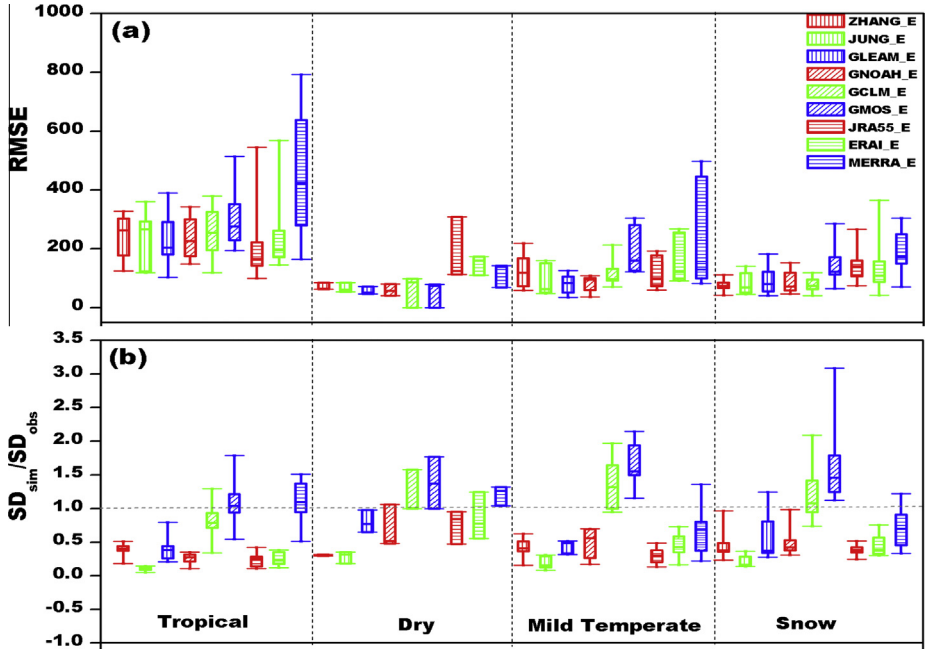

Fig. 10. Comparison of the nine ET products with respect to simulating the RMSEs and SDs for basins located in different Köppen climates.

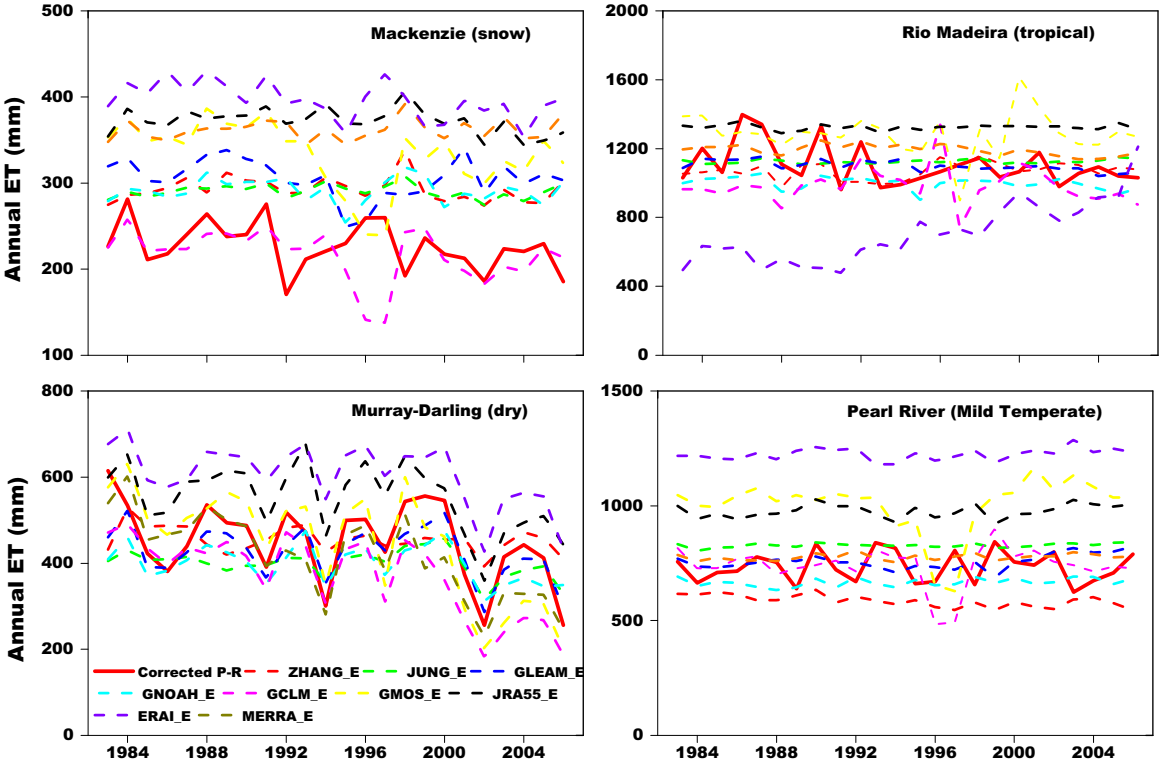

Fig. 11. Annual performances of the nine ET products for four basins located in different Köppen climate zones.

precipitation and WCRP/GEWEX SRB 3.0 radiation, which are the commonly adopted datasets for global evaluations of remote sensing retrievals and LSM simulations (Dee et al., 2011; Xia et al., 2015), were used in this study as baseline data to make the comparison of multiple forcing datasets feasible. The temperature data in Princeton forcing (GNOAH_E), GLDAS-1 forcing (GCLM_E and GMOS_E), JRA forcing (JRA55_E), WATCH forcing (ERAI_E) and MERRA-Land forcing data (MERRA_E) closely approximated those in CRU-PIK (JUNG_E) (Fig. 12 and Table 3).The NCEP/NCAR temperature (ZHANG_E) was slightly negatively biased (especially in warm regions) while ISCCP + AIRS temperature (GLEAM_E) was positively biased (especially in cold regions) compared to CRU-PIK. Relative to GPCC, the precipitation inputs for GLEAM_E (CPC Unified), the three LSM simulations and the three reanalysis-based products (precipitation in JRA55_E and ERAI_E was overestimated compared with GPCC precipitation in the dry basins) were

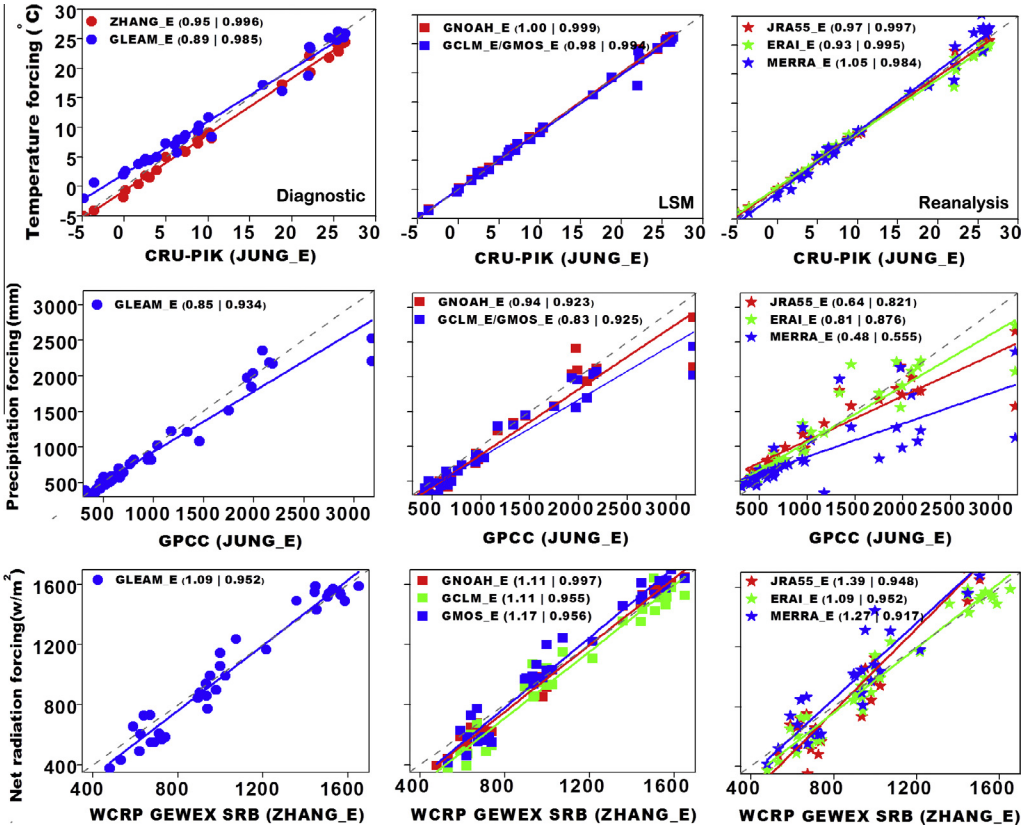

Fig. 12. Comparison of the basin-averaged forcing data (annual mean temperature, annual precipitation and net radiation) for different ET products. The linear slope and R^2 for each product are also exhibited after the name of ET product in the legend.

all underestimated, especially for basins under wet conditions. The net radiation forcing in the bias-corrected WCRP/GEWEX SRB (Sheffield et al., 2006) in GLDAS (GCLM_E and GMOS_E), Princeton forcing (GNOAH_E) and WATCH forcing (GLEAM_E and ERAI_E) was close to that in WCRP/GEWEX SRB 3.0, except for slightly underestimations in some basins. Additionally, some overestimations of net radiation were found in JRA forcing (JRA55_E) and MERRA-Land forcing (MERRA_E) compared to WCRP/GEWEX SRB net radiation.

The relatively better performances of JUNG_E and GNOAH_E with respect to estimating the annual mean ET were closely related to their higher quality forcing datasets (Fig. 12). In the tropical basins, the overestimated net radiation could be the main reason for the overestimation of *JRA55_E* (although the precipitation forcing was underestimated) while the underestimation of temperature (NCEP/NCAR) may have partly contributed to the negative-biases in annual ZHANG_E with respect to ET_{wb}. Moreover, the slight overestimations of mean annual ET in JRA55_E and ERAI_E may be attributed to the positively biased precipitation and net radiation in most basins under the dry, mild temperate and snow climates (Xue et al., 2013; Li et al., 2014). The uncertainties in precipitation forcing could also influence the inter-annual variability and trend estimates of different ET products (Ukkola and Prentice, 2013). The remotely sensed radiation datasets, which are a key inputs in most ET products (except for JUNG_E), also have large differences (especially in energy-limited basins) and may hinder the attribution of ET (e.g. global and regional ET trends over the past several decades) to solar radiation changes (Zhang et al., 2012; Ukkola and Prentice, 2013). Besides the meteorological inputs, other forcing data such as the limited EC data (JUNG_E) and vegetation parameters (Normalized Difference Vegetation Index, NDVI and the Fraction of Photosynthetically Active Radiation, FPAR) may also influence the ET means, inter-annual variability and trends estimated using different products. For example, the Model Tree Ensemble (MTE) method in JUNG_E was trained using only data from EC sites with an average duration of 2 years (Jung et al., 2010; Zeng et al., 2014), and it thus significantly underestimated the standard deviations (and further the TaylorS) relative to other ET products. The energy imbalance of some EC measurements as training samples in JUNG_E and ZHANG_E is another source of uncertainty for ET estimation (Jung et al., 2010; Zhang et al., 2010). In addition, the Advanced Very High Resolution Radiometer (AVHRR) Global Inventory Modeling and Mapping Studies (GIMMS) NDVI applied for constructing the harmonized long-term global FPAR records in JUNG_E and for determining the bio-specific canopy conductance in ZHANG, is also linked to corresponding uncertainties in ET estimations (Jung et al., 2010; Ferguson et al., 2010). For instance, the spatial resolution (8 km) of AVHRR GIMMS NDVI is much larger than the tower measurement footprint (typically 1 km in size) which may not adequately capture the sub-grid scale vegetation signals in areas of heterogeneous land cover and complex topography (Zhang et al., 2010).

Model structural and physical limitations (i.e., how the important components were described/linked in the water and energy balance, such as water storage changes, soil moisture stress, and radiations) may also be responsible for the discrepancies (means, inter-annual variability and trends) in different ET products. The simple diagnostic products can provide ET means that are similar to estimates from more complex LSM simulations and reanalysis-based datasets (Mueller et al., 2011), although the water budget is not considered. However, neglecting the water storage changes (in diagnostic ET products) and solar radiation (in JUNG_E) would deteriorate the estimated inter-annual variability of ET (Jung et al., 2010; Zeng and Cai, 2016). Another great challenge

in satellite-based ET over dry areas is the parameterization of soil moisture stress, which has been considered in Mu et al. (2011) through the use of air humidity as a proxy for soil wetness, but this has not been included in ZHANG_E. Moreover, vegetation processes, such as foliage cover (Miralles et al., 2010), stomatal conductance variations with varied atmospheric CO_2 concentrations (Betts et al., 2007), as well as forest rainfall interception loss, are also found to influence the ET estimations (Ukkola and Prentice, 2013). Higher evaporation rates from wet canopies would be induced, given the conditions of sufficient water supply and low aerodynamic resistance unrestrained by stomatal control (Miralles et al., 2011b). In the diagnostic products, the process of canopy interception loss is considered in GLEAM_E while not included in the algorithm of ZHANG_E, although the process is particularly important for the tropical basins (Zeng et al., 2014). The comparisons and evaluations regarding the influences of model physics on ET estimation between different land surface models and reanalysis systems are difficult and currently beyond the scope of this study.

3.5. Uncertainties

The main source of uncertainties in calculating ET_{wb} from the water balance method is the quality of the precipitation data. Errors in precipitation can be very large in regions (e.g. tropical regions) where the network of rain gauges is sparse (Oki and Kanae, 2006). We calculated the bias-corrected P minus Q based on three observation-based precipitation datasets (Fig. 13), namely GPCC, CPC Unified and CRU precipitation, and found there are particular uncertainties in different basins. The uncertainties in ET_{wb} estimates due to different precipitation inputs are obviously lower than those among different ET products except for the basins under tropical and wet conditions such as Ria_Jurua and Japura. Moreover, the quality of the GPCC precipitation over those sparsely gauged areas (e.g. tropical regions) may vary with a change in rain gauge density, and precipitation in snowy regions may be underestimated due to the effects of snow under-catch (Kauffeldt et al., 2013). The uncertainties in precipitation input may thus challenge the ET estimations and evaluations when using the water balance method in tropical and snowy basins (Fekete et al., 2004; Jiménez et al., 2011). It is found that the multiyear averaged ET_{wb} calculated from GPCC was relatively closer to the ET products than that calculated from CPC Unified and CRU precipitation (Fig. 13). Further, GPCC is currently considered a homogeneous observation-based precipitation dataset for the global analysis of precipitation variability and trends as well as the verification of satellite-based products, LSM simulations and reanalysis-based products (Becker et al., 2013), and has widely been applied in related water balance studies (Ramillien et al., 2006; Jung et al., 2010; Sahoo et al., 2011; Pan et al., 2012; Zhang et al., 2010; Ukkola and Prentice, 2013; Mueller et al., 2013). Moreover, the streamflow data are measured directly and are thus considered the most reliable input. However, the observed streamflow in global large rivers may be affected due to certain human activities such as inter-basin water diversion, reservoir regulation, and agricultural irrigation. By including the GRACE water storage changes in the water balance calculation at the annual scale, the influences of human activities on streamflow can be reasonably considered. A two-step bias correction method, based on probability distribution mapping, has been adopted to close the annual water budget by empirically considering the impacts of water storage changes. This method may also inherit some uncertainties, for example, generally only the systematic errors induced by the inter-annual variability of water storage changes can be corrected. However, the correction was assumed not only effective to remove the biases of P minus Q (especially in the tropical basins), but also gave us more confidence for further analysis of the temporal variations and trends of reference ET (ET_{wb}) under the changing climate and anthropogenic interferences (Li et al., 2014).

4. Summary and recommendations

This study presents a worldwide evaluation of nine ET products (three diagnostic products, three LSM simulations and three reanalysis-based products) in 35 global river basins during the period from 1983–2006. The study was conducted by comparing these products against the annual reference ET (ET_{wb}) calculated using a bias-corrected water balance method, which incorporates observed hydrological cycle components as inputs. The results show that overall, there was no obvious intra-category differences in the TaylorS of annual ET estimated from the nine products between different basins, but some individual products performed relatively better than others such as GLEAM_E in the diagnostic products, GCLM_E in the LSM simulations, and MERRA_E in the reanalysis-based products. Almost all ET products (except for MERRA_E) reasonably estimated the annual means (especially in the dry basins) but underestimated the inter-annual variability (except for GCLM_E, GMOS_E and MERRA_E) and could not adequately explain the trends (e.g. GCLM_E and MERRA_E) of ET_{wb} (especially in the energy-limited wet basins). Although many discrepancies remain among the nine ET products, the RMSEs of the ET estimates were generally larger in the tropical basins due to the sparse observations and complicated ET variations in tropical regions and in the reanalysis-based products (especially for MERRA_E in the tropical

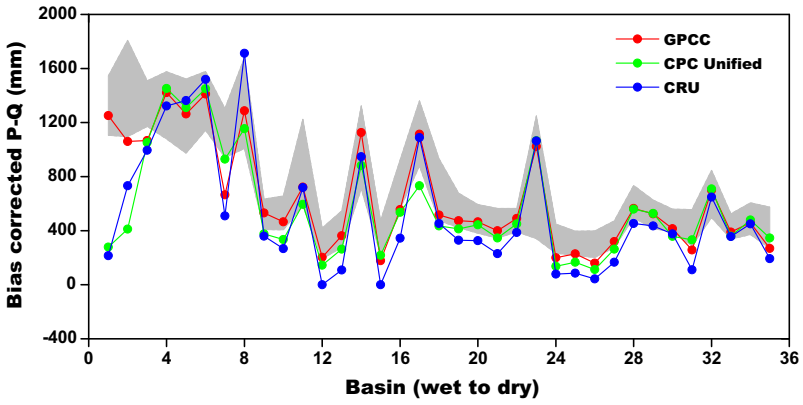

Fig. 13. Uncertainty of the bias-corrected $P - Q$ forced by different observation-based precipitation data (GPCC, CPC Unified and CRU ts3.23) from wet to dry basins. The gray shade shows the uncertainty (minimum–maximum) among different ET products.

and snowy basins) compared with the other climate zones and ET products. The different performances (means, inter-annual variability and trends) among the ET products in the 35 basins may to some extent be attributed to the discrepancies in the forcing datasets (e.g. temperature, precipitation, radiation, flux data and NDVI) . Moreover, the well-performing standard deviation estimates in GCLM_E and GMOS_E may be illusive due to the changes in the GLDAS 1 forcing dataset (Mueller et al., 2013).

Regarding the superiorities and shortcomings among the different ET products, it is thus recommended that the optimal product should be adopted based on the requirements of specific applications (e.g. trend analysis and inter-annual variability) for any given basin. Improvements are needed in global forcing datasets, e.g. with respect to meteorological forcing, streamflow, solar radiation, soil moisture stress, water storage changes and NDVI data, to enhance the accuracy of ET estimations (Zhang et al., 2012). Additionally, "ground truth" observations should be collected to constrain the model estimates of ET and to validate the performances of different ET products using direct (eddy covariance) or indirect (water balance) methods (Mueller et al., 2011). Moreover, reasonable considerations of the components of water and energy balance (e.g. precipitation, soil moisture stress, and ΔS) and the vegetation processes (forest rainfall interception loss and stomatal conductance variations) in the model structures will undoubtedly improve the estimation of global ET in the future (Miralles et al., 2011b; Zhang et al., 2012; Zeng and Cai, 2016).

Acknowledgments

This study was supported by the National Program on Key Basic Research Project of China (2013CBA01800), National Natural Science Foundation of China (41401037 and 41322001), the "Strategic Priority Research Program" of the Chinese Academy of Sciences (XDB03030000), the National Key Technologies R&D Program of China (2013BAB05B00), the Hundred Talents Programs of Chinese Academy of Sciences (Dr. Lei Wang and Dr. Fubao Sun), the Initial Founding of Scientific Research (Y5V50019YE) and the program for the "Bingwei" Excellent Talents from the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences. We are grateful to the International Precipitation Working Group (IPWG) for collecting the available precipitation datasets and the NASA MEaSUREs Program (Sean Swenson) for providing the GRACE land data processing algorithms. We thank the GRDC (Bundesanstalt fuer Gewaesserkunde, 56068 Koblenz, Germany) and Murray–Darling Basin Authority of Australia for providing the streamflow data, and we thank Dr. Diego Miralles and Dr. Yongqiang Zhang for providing the latest released ET datasets. We also wish to thank the editors and three anonymous reviewers for their invaluable comments and constructive suggestions used to improve the quality of the manuscript.

References

Andam-Akorful, S.A., Ferreira, V.G., Awange, J.L., Forootan, E., He, X.F., 2014. Multi-model and multi-sensor estimations of evapotranspiration over the Volta Basin, West Africa. Int. J. Climatol. 35 (10), 3132–3145.

Beck, C., Grieser, J., Rudolf, B., 2005. A new monthly precipitation climatology for the global land areas for the period 1951 to 2000. In: Klimastatusbericht 2004. Dtsch. Wetterdienst, Offenbach, Germany, pp. 181–190.

Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Schamm, K., Schneider, U., Ziese, M., 2013. A description of the global land-surface precipitation products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901-present. Earth Syst. Sci. Data 5, 71–99.

Berrisford, P, Lee, D., Poli, P., Brugge, R., Fielding, K., Fuentes, M., Kållberg, P., Kobayashi, S., Uppala, S., Simmons, A., 2011. The ERA-interim archive. ERA Reports Series No. 1 Version 2.0. Available from: <http://old.ecmwf.int/publications/library/ecpublications/_pdf/era/era_report_series/RS_1_v2.pdf>.

Betts, A., Ball, J., Beljaars, A., Miller, M., Viterbo, P.A., 1996. The land surface-atmosphere interaction: a review based on observational and global modeling perspectives. J. Geophys. Res. 101, 7209–7226.

Betts, R.A., Boucher, O., Collins, M., Cox, P.M., Falloon, P.D., Gedney, N., Hemming, D. L., Huntingford, C., Jones, C.D., Sexton, D.M.H., Webb, M.J., 2007. Projected increase in continental runoff due to plant responses to increasing carbon dioxide. Nature 448, 1037–1041.

Bouraoui, F., Vachaud, G., Li, L.Z.X., LeTreut, H., Chen, T., 1999. Evaluation of the impact of climate changes on water storage and groundwater recharge at the watershed scale. Clim. Dyn. 15 (2), 153–161.

Brutsaert, W., Parlange, M.B., 1998. Hydrologic cycle explains the evaporation paradox. Nature 396 (6706), 30.

Chen, D.L., Chen, W.T., 2013. Using the Köppen classification to quantify climate variation and change: an example for 1901–2010. Environ. Develop. 6, 69–79.

Chen, M.Y., Shi, W., Xie, P.P., Silva, V.B.S., Kousky, V.E., Higgins, R.W., Janowiak, J.E., 2008. Assessing objective techniques for gauge-based analyses of global daily precipitation. J. Geophys. Res. 113, D04110.

Condon, L.E., Maxwell, R.M., 2014. Feedbacks between managed irrigation and water availability: diagnosing temporal and spatial patterns using an integrated hydrologic model. Water Resour. Res. 50 (3), 2600–2616.

da Rocha, H.R., Manzi, A.O., Cabral, O.M., Miller, S.D., Goulden, M.L., Saleska, S.R., -Coupe, N.R., Wofsy, S.C., Borma, L.S., Artaxo, P., Vourlitis, G., Nogueira, J.S., Kruijt, B., Freitas, H.C., von Randow, C., Aguiar, R.G., Maia, J.F., 2009. Patterns of water and heat flux across a biome gradient from tropical forest to savanna in Brazil. J. Geophys. Res. 114, G00B12. http://dx.doi.org/10.1029/2007JG000640.

Dai, A., Qian, T.T., Trenberth, K.E., Milliman, J.D., 2009. Changes in continental freshwater discharge from 1948 to 2004. J. Climate 22, 2773–2792.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., HÓlm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., Vitart, F., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Quart. J. Roy. Meteorol. Soc. 137, 553–597.

Dirmeyer, P.A., Gao, X., Zhao, M., Guo, Z., Oki, T., Hanasaki, N., 2006. GSWP-2: Multimodel analysis and implications for our perception of the land surface. Bull. Am. Meteorol. Soc. 87, 1397–1831.

Fekete, B.M., Vörösmarty, C.J., Roads, J.O., Willmott, C.J., 2004. Uncertainties in precipitation and their impacts on runoff estimates. J. Climate 17, 294–302.

Ferguson, C.R., Sheffield, J., Wood, E.F., Gao, H., 2010. Quantifying uncertainty in a remote sensing-based estimate of evapotranspiration over continental USA. Int. J. Remote Sens. 31 (14), 3821–3865.

Han, E.J., Crow, W.T., Hain, C.R., Anderson, M.C., 2015. On the use of a water balance to evaluate interannual terrestrial ET variability. J. Hydrometeorol. 16, 1102–1108.

Harding, R.J., Weedon, G.P., van Lanen, H.A.J., Clark, D.B., 2014. The future for global water assessment. J. Hydrol. 518, 186–192.

Harris, I., Jones, P.D., Osborn, T.J., Lister, D.H., 2014. Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset. Int. J. Climatol. 34 (3), 623–642.

Hasler, N., Avissar, R., 2007. What controls evapotranspiration in the Amazon basin? J. Hydrometeorol. 8, 380–395.

Hobbins, M.T., Ramírez, J.A., Brown, T.C., 2001. The complementary relationship in estimation of regional evapotranspiration: an enhanced advection-aridity model. Water Resour. Res. 37 (5), 1389–1403.

Jiménez, C., Prigent, C., Mueller, B., Seneviratne, S.I., McCabe, M.F., Wood, E.F., Rossow, W.B., Balsamo, G., Betts, A.K., Dirmeyer, P.A., Fisher, J.B., Jung, M., Kanamitsu, M., Reichle, R.H., Reichstein, M., Rodell, M., Sheffield, J., Tu, K., Wang, K., 2011. Global intercomparison of 12 land surface heat flux estimates. J. Geophys. Res. 116, D02102. http://dx.doi.org/10.1029/2010JD014545.

Jung, M., Reichstein, M., Bondeau, A., 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.

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S.I., Sheffield, J., Goulden, M.L., Bonan, G., Cescatti, A., Chen, J., de Jeu, R., Dolman, A.J., Eugster, W., Gerten, D., Gianelle, D., Gobron, N., Heinke, J., Kimball, J., Law, B.E., Montagnani, L., Mu, Q., Mueller, B., Oleson, K., Papale, D., Richardson, A.D., Roupsard, O., Running, S., Tomelleri, E., Viovy, N., Weber, U., Williams, C., Wood, E., Zaehle, S., Zhang, K., 2010. Recent decline in the global land evapotranspiration trend due to limited moisture supply. Nature 467, 951–954.

Kauffeldt, A., Halldin, S., Rodhe, A., Xu, C.-Y., Westerberg, I.K., 2013. Disinformative data inlarge-scale hydrological modeling. Hydrol. Earth Syst. Sci. 17, 2845–2857.

Kistler, R., Kalnay, E., Collins, E., Saha, S., White, G., Woollen, J., Chellian, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R., Fiorino, M., 2001. The NCEP-NCAR 50-Year reanalysis: monthly means CD-ROM and documentation. Bull. Am. Meteor. Soc. 82, 247–268.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., kamahori, H., kobayashi, C., Endo, H., miyaoka, K., Takahashi, K., 2015. The JRA-55 Reanalysis: General specifications and basic characteristics. J. Meteor. Soc. Jpn. 93 (1), 5–58. http://dx.doi.org/10.2151/jmsj.2015-001.

Köppen, W., 1936. Das geographische system der klimate. In: Köppen, W., Geiger, R. (Eds.), Handbuch der Klimato - logie. Gebrüder Borntraeger, Berlin, pp. 1–44.

Landerer, F.W., Swenson, S.C., 2012. Accuracy of scaled GRACE terrestrial water storage estimates. Water Resour. Res. 48, W04531.

Lettenmaier, D.P., Milly, P.C.D., 2009. Land waters and sea level. Nature Geosci. 2 (7), 452–454.
Li, X.P., Wang, L., Chen, D.L., Yang, K., Wang, A.H., 2014. Seasonal evapotranspiration changes (1983–2006) of four large basins on the Tibetan Plateau. J. Geophys. Res. 119 (23), 13079–13095.
Long, D., Shen, Y.J., Sun, A., Hong, Y., Longuevergne, L., Yang, Y.T., Li, B., Chen, L., 2014. Drought and flood monitoring for a large karst plateau in Southwest China using extended GRACE data. Remote Sen. Environ. 155, 145–160.
Lucchesi, R., 2012. File specification for MERRA products. GMAO Office Note No. 1 (version 2.3), 82 pp. Available from <http://gmao.gsfc.nasa.gov/pubs/office_notes>.
Miralles, D.G., De Jeu, R.A.M., Gash, J.H., Holmes, T.R.H., Dolman, A.J., 2011a. Magnitude and variability of land evaporation and its components at the global scale. Hydrol. Earth Syst. Sci. 15, 967–981.
Miralles, D.G., Gash, J.H., Holmes, T.R.H., de Jeu, R.A.M., Dolman, A.J., 2010. Global canopy interception from satellite observations. J. Geophys. Res. 115, D16122.
Miralles, D.G., Holmes, T.R.H., De Jeu, R.A.M., Gash, J.H., Meesters, A.G.C.A., Dolman, A.J., 2011b. Global land-surface evaporation estimated from satellite-based observations. Hydrol. Earth Syst. Sci. 15, 453–469.
Mitchell, T.D., Jones, P.D., 2005. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. Int. J. Climatol. 25, 659–712.
Mu, Q., Heinsch, F.A., Zhao, M., Running, S.W., 2007. Development of a global evapotranspriation algorithm based on MODIS and global meteorology data. Remote Sens. Environ. 111 (4), 519–536.
Mu, Q., Zhao, M., Running, S.W., 2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. Remote Sens. Environ. 115 (8), 1781–1800.
Mueller, B., Seneviratne, S.I., Jimenez, C., Corti, T., Hirschi, M., Balsamo, G., Ciais, P., Dirmeyer, P., Fisher, J.B., Guo, Z., Jung, M., Maignan, F., McCabe, M.F., Reichle, R., Reichstein, M., Rodell, M., Sheffield, J., teuling, A.J., Wang, K., Wood, E.F., Zhang, Y., 2011. Evaluation of global observations-based evapotranspriation datasets and IPCC AR4 simulations. Geophys. Res. Lett. 38, L06402. http://dx.doi.org/10.1029/2010GL046230.
Mueller, B., Hirschi, M., Jimenez, C., Ciais, P., Dirmeyer, P.A., Dolman, A.J., Fisher, J.B., Jung, M., Ludwig, F., Maignan, F., Miralles, D.G., McCabe, M.F., Reichstein, M., Sheffield, J., Wang, K., Wood, E.F., Zhang, Y., Seneviratne, S.I., 2013. Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis. Hydrol. Earth Syst. Sci. 13, 3707–3720.
Ohmura, A., Wild, M., 2002. Is the hydrological cycle accelerating? Science 298, 1345–1356.
Oki, T., Kanae, S., 2006. Global hydrological cycles and world water resources. Science 313, 1068–1072.
Onogi, K., Tsutsui, J., Koide, H., Sakamoto, M., Kobayashi, S., Hatsushika, H., Matsumoto, T., Yamazaki, N., Kamahori, H., Takahashi, K., Kadokura, S., Wada, K., Kato, K., Oyama, R., Ose, T., Mannoji, N., Taira, R., 2007. The JRA-25 reanalysis. J. Meteorol. Soc. Jpn. Ser. II 85 (3), 369–432.
Pan, M., Sahoo, A.K., Troy, T.J., Vinukollu, R.K., Sheffield, J., Wood, E.F., 2012. Multisource estimation of long-term terrestrial water budget for major global river basins. J. Climate 25, 3191–3206.
Ramillien, G., Frappart, F., Guntner, A., Ngo-Duc, T., Cazenave, A., Laval, K., 2006. Time variations of the regional evapotranspiration rate from Gravity Recovery and Climate Experiments (GRACE) satellite gravimetry. Water Resour. Res. 42 (10), W10403. http://dx.doi.org/10.1029/2005WR004331.
Rodell, M., Famiglietti, J.S., 2002. The potential for satellite-based monitoring of groundwater storage changes using GRACE: the Hight Plains aquifer, central U. S. J. Hydrol. 263, 245–256.
Rodell, M., Famiglietti, J.S., Chen, J., Seneviratne, S.I., Viterbo, P., Holl, S., Wilson, C.R., 2004a. Basin scale estimates of evapotranspiration using GRACE and other observations. Geophys. Res. Lett. 31, L20504.
Rodell, M., Houser, P.R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J.K., Walker, P., Lohmann, D., Toll, D., 2004b. The global land data assimilation system. Bull. Am. Meteorol. Soc. 85, 381–394.
Rodell, M., McWilliams, E.B., Famiglietti, J.S., Beaudoing, H.K., Nigro, J., 2011. Estimating evapotranspiration using an observation based terrestrial water budget. Hydrol. Process. 25, 4082–4092.
Roderick, M.L., Farquhar, G.D., 2011. A simple framework for relating variations in runoff to variations in climatic conditions and catchment properties. Water Resour. Res. 47 (12), W00G07.
Rossow, W.B., Dueñas, E.N., 2004. The International Satellite Cloud Climatology Project (ISCCP) Web Site: an online resource for research. Bull. Am. Meteorol. Soc. 85, 167–172.
Rui, H., 2011. README Document for Global Land Data Assimilation System Version 2 (GLDAS-2) Products, GES DISC.
Running, S., 1998. A blueprint for improved global change monitoring of the terrestrial biosphere. Earth Obs. 10, 8–12.
Sahoo, A.K., Pan, M., Troy, T.J., Vinukollu, R.K., Sheffield, J., Wood, E.F., 2011. Reconciling the global terrestrial water budget using satellite remote sensing. Remote Sens. Environ. 115, 1850–1865.
Sheffield, J., Ferguson, C.R., Troy, T.J., Wood, E.F., McCabe, M.F., 2009. Closing the terrestrial water budget from satellite remote sensing. Geophys. Res. Lett. 36, L07403. http://dx.doi.org/10.1029/2009GL037338.
Sheffield, J., Goteti, G., Wood, E.F., 2006. Development of a 50-year high-resolution global dataset of meteorological forcing for land surface modeling. J. Climate 19 (13), 3088–3111.
Simmons, A., Uppala, S., Dee, D., Kubayashi, S., 2006. ERA-Interim: New ECMWF reanalysis products from 1989 onwards. ECMWF Newsl. 110, 25–35.
Su, T., Feng, T.C., Feng, G.L., 2015. Evaporation variability under climate warming in five reanalyses and its association with pan evaporation over China. J. Geophys. Res. Atmos. 120, 8080–8098.
Su, Z., 2002. The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. Hydrol. Earth Syst. Sci. 6 (1), 85–99.
Sun, A.Y., 2013. Predicting groundwater level changes using GRACE data. Water Resour. Res. 49, 5900–5912.
Swenson, S.C., Wahr, J., 2006. Estimating large-scale precipitation minus evapotranspiration from GRACE satellite gravity measurements. J. Hydrometeor. 7 (2), L252–L270.
Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res. 106 (D7), 7183–7192.
Tang, X.G., Li, H.P., Desai, A.R., Nagy, Z., Luo, J.H., Kolb, T.E., Olioso, A., Xu, X.B., Yao, L., Kutsch, W., Pilegaard, K., Köstner, B., Ammann, C., 2014. How is water-use efficiency of terrestrial ecosystems distributed and changing on Earth? Scientific Report 4, 7483.
Thom, H.C.S., 1958. A note on the gamma distribution. Mon. Weather Rev. 86 (4), 117–122.
Ukkola, A.M., Prentice, I.C., 2013. A worldwide analysis of trends in water-balance evapotranspiration. Hydrol. Earth Syst. Sci. 17, 4177–4187.
Velpuri, N.M., Senay, G.B., Singh, R.L., Bohms, S., Verdin, J.P., 2013. A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous Unites States: using point and grided FLUEXNET and water balance ET. Remote Sens. Environ. 139, 35–49.
Wahr, J., Swenson, S., Velicogna, I., Zlotnicki, V., 2004. Time-variable gravity from GRACE: first results. Geophys. Res. Lett. 31, L11501.
Wan, Z.M., Zhang, K., Xue, X.W., Hong, Z., Hong, Y., Gourley, J.J., 2015. Water balance based actual evapotranspiration reconstruction from ground and satellite observations over the Conterminous United States. Water Resour. Res. http://dx.doi.org/10.1002/2015WR017311.
Wang, K.C., Dickinson, R.E., 2012. A review of global terrestrial evapotranspiration: observation, modeling, climatology, and climatic variability. Rev. Geophys. 50, RG2005. http://dx.doi.org/10.1029/2011RG000373.
Xia, Y.L., Cosgrove, B.A., Mitchell, K.E., Peters-Lidard, C.D., Ek, M.B., Kumar, S., Mocko, D., Wei, H.L., 2015. Basin-scale assessment of the land surface energy budget in the NCEP Operational and Research NLDAS-2 stystem. J. Geophys. Res. Atmos. http://dx.doi.org/10.1002/2015JD023889.
Xu, C., Chen, D., 2005. Comparison of seven models for estimation of evapotranspiration and groundwater recharge using lysimeter measurement data in Germany. Hydrol. Process. 19, 3717–3734. http://dx.doi.org/10.1002/hyp.5853.
Xu, C.Y., Singh, V.P., 2005. Evaluation of three complementary relationship evapotranspiration model by water balance approach to estimate actual regional evapotranspiration in different climatic regions. J. Hydrol. 308 (1–4), 105–121.
Xue, B.L., Wang, L., Li, X.P., Yang, K., Chen, D.L., Sun, L.T., 2013. Evaluation of evapotranspriation estimates for two river basins on the Tibetan Plateau by a water balance method. J. Hydrol. 492, 290–297.
Yang, Y.T., Long, D., Shang, S.H., 2013. Remote estimation of terrestrial evapotranspiration without using meteorological data. Geophys. Res. Lett. 40, 3026–3030.
Zeng, R.J., Cai, X.M., 2016. Climatic and terrestrial storage control on evapotranspiration temporal variability: analysis of river basins around the world. Geophys. Res. Lett. 43, 185–195.
Zeng, Z.Z., Wang, T., Zhou, F., Ciais, P., Mao, J.F., Shi, X.Y., Piao, S.L., 2014. A worldwide analysis of spatiotemporal changes in water balance-based evapotranspiration from 1982–2009. J. Geophys. Res. Atmos. 119, 1186–1202.
Zhang, K., Kimbal, J.S., Mu, Q., Jones, L.A., Goetz, S.J., Running, S.W., 2009. Satellite based analysis of northern ET trends and associated changes in the regional water balance from 1983 to 2005. J. Hydrol. 379, 92–110.
Zhang, K., Kimball, J.S., Nemani, R.R., Running, S.W., 2010. A continuous satellite-derived global record of land surface evapotranspiration from 1983 to 2006. Water Resour. Res. 46 (9), W09522.
Zhang, K., Kimball, J.S., Nemani, R.R., Running, S.W., Hong, Y., Gourley, J.J., Yu, Z.B., 2015. Vegetation greening and climate change promote multidecadal rises of global land evapotranspiration. Scientific Report 5, 15956.
Zhang, Y., Leuning, R., Chiew, F.H.S., Wang, E., Zhang, L., Liu, C.M., Sun, F.B., Peel, M. C., Shen, Y.J., Jung, M., 2012. Decadal trends in evaporation from global energy and water balances. J. Hydrometeor. 13, 379–391.