Evaluation of different evapotranspiration products in the middle Yellow River Basin, China

Yanzhong Li, Kang Liang, Changming Liu, Wenbin Liu and Peng Bai

ABSTRACT

Actual evapotranspiration (ET_a) is a central process in the climate system and a nexus of the water and energy cycles. This study assesses the hydrological performance of the four categories of ET_a products (i.e., land surface models (LSMs), reanalysis, model tree ensemble, and diagnostic models (DMs)) for use in the middle Yellow River Basin (MYRB) using water balance methods. The results show the following. (1) The water storage changes significantly at annual scale and cannot be neglected when calculating the reference ET_a by the water balance methods. (2) ET_a from LSMs, considering the precipitation input, exhibits the best performance in capturing the reference ET_a variation. The MET ET_a (AET_{JUNG}), based on eddy covariance, has fair performance with a small underestimation, followed by the DMs, including MODIS and ZhangKe. Poor performance is found in reanalysis ET_a (JRA55), due to overestimations precipitation and radiation. (3) The reference ET_a showed decreased and then increased trend. ET_a from the LSMs-Noah model captures the trend well, followed by the LSMs-variable infiltration capacity model. Our results are not only meaningful for better understanding ET_a variability in the MYRB, but also significant for improving global ET_a products models' performance in semi-arid and semi-humid regions.

Key words | global actual evapotranspiration products, middle Yellow River Basin, reference actual evapotranspiration, water balance method

Yanzhong Li
Kang Liang (corresponding author)
Changming Liu
Wenbin Liu
Peng Bai
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
E-mail: *liangk@igsnrr.ac.cn*

Yanzhong Li
University of Chinese Academy of Sciences,
Beijing 100049,
China

INTRODUCTION

Actual evapotranspiration (ET_a) is one of the important hydrological processes (Brutsaert 2005; Jung *et al.* 2010). ET_a determines the partitioning of available energy on the land surface into latent and sensible heat flux and can influence regional and global climates (Bonan *et al.* 1992). Global land ET_a returns approximately 60% of annual land precipitation to the atmosphere (Hetherington & Woodward 2003; Koster *et al.* 2004; Oki & Kanae 2006), and the ratio is even larger (>95%) in arid and semi-arid regions (Moiwo & Tao 2015). Acceleration or intensification of the hydrological cycle has occurred with global warming, and ET_a has also been affected (Brutsaert & Parlange 1998; Jung *et al.* 2010). However, the existing observations of ET_a primarily focus on point observations, such as Bowen ratio–energy balances (Shen *et al.* 2002), flux towers (Fisher *et al.* 2008), lysimeters (Xu & Chen 2005; Trajkovic 2010), and large aperture scintillometers (Hemakumara *et al.* 2003), whereas direct observations of ET_a at regional or global scales are still lacking.

Different methods have been used to estimate regional ET_a, such as complementary relationship (Brutsaert 2005; Gao *et al.* 2012; Ma *et al.* 2015), the model tree ensemble (MTE) method (Shutov *et al.* 2006; Jung *et al.* 2010), the remote sensing-based method (Mu *et al.* 2011), land surface models (LSMs) (Lakshmi *et al.* 2011; Yin *et al.* 2013; Cai *et al.* 2014), and the reanalysis model (Nakaegwa 2008; Kobayashi *et al.* 2015). The MTE method, based on eddy covariance, employs a machine-learning algorithm, which is trained by evaporation measurements from the global

doi: 10.2166/nh.2016.120

Flux Network (FLUXNET) database, gridded global meteorological data, and the remotely sensed fraction of absorbed photosynthetically active radiation (FPAR). Largely independent of theoretical model assumptions, the MTE method is considered to perform best in capturing ET_a, and has been widely used for validating and parameterizing other models (Cleugh *et al.* 2007; Fisher *et al.* 2008). The remote sensing method provides an unprecedented opportunity to monitor spatiotemporal variability in ET_a in two basic ways: (1) via vegetation information (such as the leaf area index (LAI) or normalized difference vegetation index (NDVI)) and (2) the land surface temperature (LST). The LAI or NDVI, derived from remote sensing data, is used for calculating surface resistance in the Penman–Monteith (PM) algorithm to estimate ET_a (Zhang *et al.* 2010; Mu *et al.* 2011). Remote sensed LST is used to calculate latent heat (LE) as the residual of surface net radiation (R_n), soil heat flux (G), and sensible heat flux (H) (i.e., $LE = R_n - G - H$) (Bastiaanssen 2000; EI Haj EI Tahir *et al.* 2012; Gokmen *et al.* 2012). LSMs are excellent in capturing ET_a variation (Xue *et al.* 2013; Cai *et al.* 2014) and are widely recommended for estimating the hydrological components for large areas, such as surface energy, water fluxes and their response to near-surface atmospheric forcing (Rodell *et al.* 2004; Sheffield *et al.* 2006). Atmospheric reanalysis data, with past observations and state-of-the-art numerical weather prediction systems, can provide various surface hydrological variables, such as soil moisture, precipitation, and ET_a (Hogue *et al.* 2006; Ebita *et al.* 2011). Although these methods provide a potential means to estimate ET_a, it is still a challenge to accurately obtain ET_a due to the differences of models and forcing data (Cai *et al.* 2014; Long *et al.* 2014). In addition, the impact of different model types and forcing data on ET_a products performance needs to be detected. The traditional water balance method provides a useful tool for addressing these questions. Evaluating ET_a products in a closed basin (Rodell *et al.* 2004; Li *et al.* 2014) and analyzing the forcing data and models will benefit the improvement of each model in sensitive regions of climate change.

The middle Yellow River Basin (MYRB), located in the Loess Plateau, endures the most severe soil erosion and water loss in the world (approximately nine to 21 times greater than most major rivers in the world) (Huang *et al.* 2003). To address the environmental crises and promote ecological restoration, the Chinese government has implemented a number of national strategies since 1999, such as the Natural Forest Conversation Program and the Grain to Green Project (GTGP) (Zhang *et al.* 2000; FAO 2010), which have significantly impacted hydrological processes (Zhang *et al.* 2007, 2013). In addition, the MYRB is located in the transition zone of semi-arid and semi-humid regions of China, and is particularly sensitive to climate change (Yang *et al.* 2010; Wang *et al.* 2012; Zhu *et al.* 2015). Various studies have documented significant changes in evapotranspiration in the region (Zhang *et al.* 2011b; Liu *et al.* 2012a). Although the spatiotemporal variation of ET_a can be detected by various ET_a datasets, it is unclear which one is optimal for semi-arid and semi-humid regions. Therefore, the evaluation of ET_a products and investigation into the variation of ET_a in the MYRB are essential for understanding hydrological processes, and are of great value for regional sustainable water resource management.

The main objectives of this paper were to evaluate the performance of different ET_a products and to investigate the causes of different performances in the MYRB, which was not only meaningful for the selection of appropriate ET_a products for regional water resource management, but also significant for improving the ET_a models in semi-arid and semi-humid regions. The study objectives are achieved through the following steps: (1) to estimate the reference ET_a based on the traditional water balance method in the MYRB; (2) to evaluate the performance of different ET_a products used in the MYRB against the reference ET_a; (3) to analyze the reasons for differences in ET_a model performance; and (4) to investigate the performance of each model in capturing the temporal variations of ET_a.

STUDY AREA, DATASETS AND METHODS

Study regions

Our study area is located in the middle of the Yellow River in China, and covers approximately 3.0×10^5 km^2, spanning from 103.95° E to 113.51° E and from 33.69° N to 40.57° N. This area accounts for approximately 38% and 48.4% of the whole Yellow River Basin and Loess Plateau, respectively

(Figure 1). Plateaus, ridges, mounds, and gullies are the dominant landscape feature characterized by loess-paleosol soil with an average depth in excess of 100 m. The vegetation cover is distributed with forest, forest-steppe, typical-steppe, and desert-steppe zones from southeast to northwest, and the land use is predominantly cultivated croplands and improved grasslands. The total annual precipitation ranges from 300 mm in the northwest to approximately 800 mm in the southeast, and approximately half of the precipitation occurs during the rainy season from June to September. Most of the precipitation is in the form of intense rainstorms, resulting in the most severe soil and water loss in the world (Li *et al.* 2008). The average annual pan evaporation (20 cm diameter pan, Liu *et al.* 2011a) is approximately 1,500 mm, which is three to seven times more than the annual precipitation.

Datasets

Global ET_a products

Four different categories of ET_a products (Table 1) were evaluated in our study: (1) LSMs, (2) reanalysis, (3) MTE method, and (4) diagnostic models (DMs). Five ET_a products, coming from LSMs in the Global Land Data Assimilation System (GLDAS) with the spatial resolution of 1° and time range from 1979, except for Noah2 from 1948, were employed. AET_{Noah1} and AET_{Noah2} came from the Noah model, AET_{CLM} from the community land model (CLM), AET_{VIC} from the variable infiltration capacity (VIC) and AET_{MOS} from the mosaic model (MOS) (Rodell *et al.* 2004; Ferreira *et al.* 2013). The AET_{Noah2} was Land Surface Version 2 (for 1948–2010), forced by the Princeton meteorological forcing data (Sheffield *et al.* 2006), while the others were Version 1 forced by constantly updated meteorological data (http://disc.sci.gsfc.nasa.gov/hydrogy/data-holdings). The reanalysis of 55 years' ET_a production, with the spatial resolution of 1.25° and time range from 1958 to 2014, was the second Japanese global atmospheric reanalysis project, and improved on many of the deficiencies of the first 25-year reanalysis (namely, AET_{JRA55} and AET_{JRA25}, respectively) (Nakaegwa 2008). The main objective of AET_{JRA55} was to produce a comprehensive atmospheric dataset suitable for studying multi-decadal variability and climate change

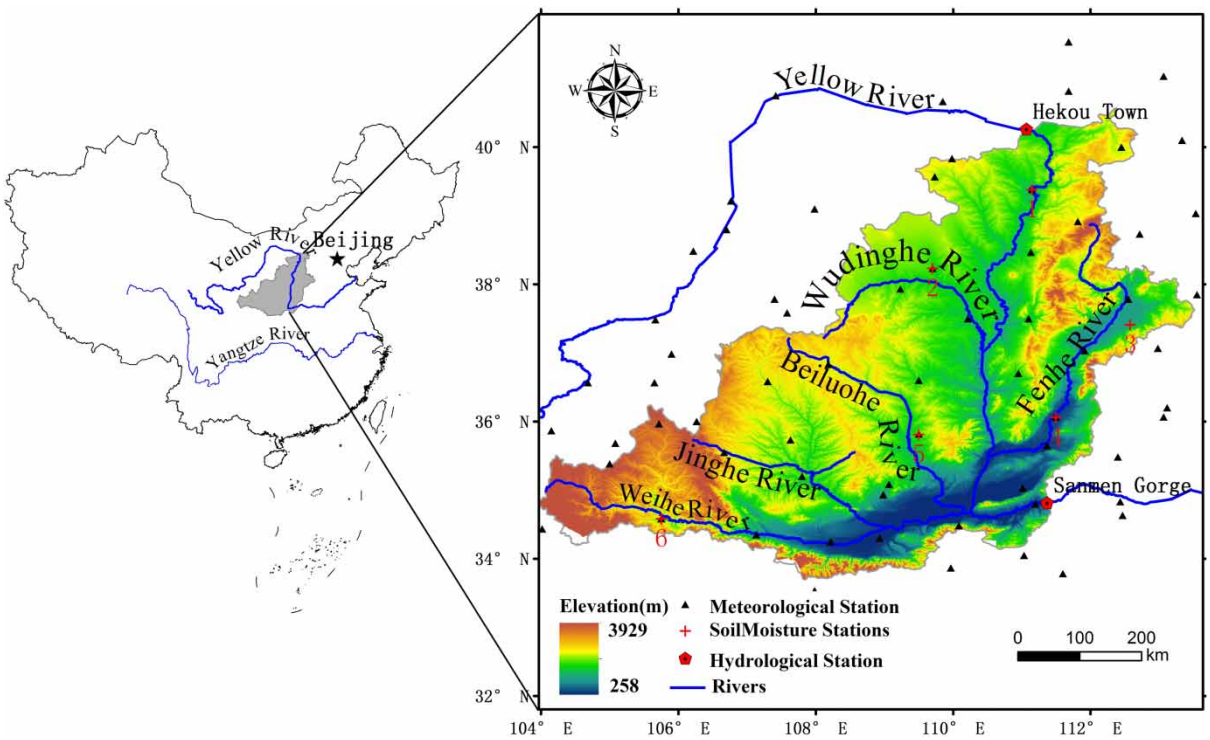

Figure 1 | Sketch map of the study area.

Table 1 | Overview of the evapotranspiration products

Dataset	Category	Spatial resolution	Temporal range	References
AET_{Noah1}	LSM-NOAH-Version1	$1^\circ \times 1^\circ$	1979–2014	Rodell *et al.* (2004)
AET_{Noah2}	LSM-NOAH-Version2	$1^\circ \times 1^\circ$	1948–2010	Rodell *et al.* (2004)
AET_{CLM}	LSM-CLM	$1^\circ \times 1^\circ$	1979–2014	Rodell *et al.* (2004)
AET_{VIC}	LSM-VIC	$1^\circ \times 1^\circ$	1979–2014	Rodell *et al.* (2004)
AET_{MOS}	LSM-mosaic model	$1^\circ \times 1^\circ$	1979–2014	Rodell *et al.* (2004)
AET_{JRA55}	Reanalysis	$1.25^\circ \times 1.25^\circ$	1958–2014	Kobayashi *et al.* (2015)
AET_{JUNG}	Eddy covariance method	$0.5^\circ \times 0.5^\circ$	1982–2011	Jung *et al.* (2010)
$AET_{ZhangKe}$	PM method	8 km × 8 km	1983–2006	Zhang *et al.* (2010)
AET_{MODIS}	PM method	1 km × 1 km	2000–2013	Mu *et al.* (2011)

(http://jra.kishou.go.jp/JRA-55/index_en.html). The Jung global land ET_a product, with the spatial resolution of 0.5° and time range from 1982 to 2011, was data-driven based and is compiled using data from 198 global flux monitoring towers (FLUXNET), remote sensing data, meteorological observations, and the MTE (a machine-learning algorithm, data download from http://www.bgc-jena.mpg.de/geodb/projects/Data.php; Jung *et al.* 2010). ET_a from ZhangKe (http://www.ntsg.umt.edu/project/et; Zhang *et al.* 2011a) and MODIS (http://www.ntsg.umt.edu/project), with the spatial resolution of 8 km and 1 km, respectively, and time range 1983–2006 and 2000–2013, respectively, were DMs and used remote sensed vegetation information as an important input using the PM method (Zhang *et al.* 2010; Mu *et al.* 2011). Additional details on the nine ET_a products employed in our study are shown in Table 1.

Precipitation, streamflow, and soil moisture

Monthly streamflow data from 1960 to 2013 at two gauges, the inlet at Hekou Town and the outlets at Sanmen Gorge, were obtained from the Yellow River Hydrological Bureau. The average annual streamflow was aggregated from monthly data. Monthly meteorological data of 73 national meteorological stations (Figure 1), from the National Climate Center of the China Meteorological Administration (CMA) were used in this study. These data included precipitation, air temperature, wind speed, vapor pressure, and sunshine duration.

The monthly grid precipitation data were averaged from three interpolation results by the inverse distance weighted (IDW) (Qian *et al.* 2006), spline (New *et al.* 2000), and kriging (Xu *et al.* 2006) methods to eliminate the uncertainty from different interpolation methods. Precipitation with a resolution of 0.05° from 1960 to 2014 was used to calculate the reference ET_a with the water balance methods.

Soil moisture at 10 cm, 20 cm, and 50 cm depths was obtained from the agrometeorological stations of the CMA for 1992 to 2013, and the volumetric soil moisture was calculated from the mass percentage using the gravimetric technique (Li *et al.* 2005). These data were used to evaluate the performance of soil moisture products from four models including MOS, Noah2, CLM, and VIC produced by the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences (VIC-IGSNRR, http://hydro.igsnrr.ac.cn/resources.html). The four soil moisture products were summed to the 50 cm depth, and compared with the observed data.

METHODS

Water balance method

For a closed catchment, the water balance method was typically used to calculate ET_a (Xu 2001; Gao *et al.* 2007; Ferreira *et al.* 2013; Corbari *et al.* 2014; Long *et al.* 2014):

$$ET_a = P - R - \Delta S \quad (1)$$

where P was the annual precipitation (mm yr^{-1}) normally obtained from a meteorological station; R was the

streamflow observed at a hydrological station and obtained the streamflow depth (mm yr^{-1}) according to the area of basin; ΔS was the terrestrial water storage change (TWSC, mm yr^{-1}), which was reflected in changes in the moisture storage of surface and subsurface stores, such as glaciers, groundwater, and soil moisture. The ΔS played a fundamental role in water, energy, and biogeochemical cycles at regional and global scales (Famiglietti 2013). However, accurate measurement of TWSC over large areas and long time series was a challenge due to a lack of large-scale, long-term *in situ* observations (Lettenmaier & Famiglietti 2006). Although the satellite data from the Gravity Recovery and Climate Experiment were widely used to explore ΔS in large river basins (Li *et al.* 2014; Long *et al.* 2014), the relatively short duration of this measurement (2002 until now) restricted its application. Some papers had assumed a negligible ΔS for long-term (Hobbins *et al.* 2001; Xue *et al.* 2013), which may be true in some regions. However, the MYRB, located in the semi-arid and semi-humid transition region, was highly sensitive to climate changes and had experienced apparent land use and land cover changes (Huang *et al.* 2003); therefore, the ΔS needed to be tested in our study region. The variation of TWSC can be captured by the total soil moisture change (TSMC) (Lettenmaier & Famiglietti 2006; Yang *et al.* 2015b), and some papers have documented that the soil moisture derived from the land-surface model had better performance than those derived from others models (Chen *et al.* 2013).

Evaluation criteria of ET_a products

The ET_a of each model was calculated from the gridded data as the weighted arithmetic mean using the proportion of the area of a certain grid included in the study region to the total area of the MYRB as weights. Next, the model ET_a was evaluated against the reference ET_a via several criteria. The evaluation statistics included the following: the bias (BIAS), root-mean-square error (RMSE), and correlation coefficient (CORR), and are defined as follows:

$$\mathrm{BIAS} = \sum_{i=1}^{N} (X_i - Y_i)/N \tag{2}$$

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{N} (X_i - Y_i)^2/N} \tag{3}$$

$$\mathrm{CORR} = \frac{\sum_{i=1}^{N} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N} (X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{N} (Y_i - \overline{Y})^2}} \tag{4}$$

where N represented the number of years during the study period, and X_i, Y_i were the annual model simulation ET_a and the reference ET_a, respectively.

Statistical methods

The rank-based non-parametric Mann–Kendall statistic test (Kendall 1948) has been commonly used to detect trends due to its robustness for non-normally distributed data, which have been frequently encountered in hydro-climatic time-series (Liu *et al.* 2011a, 2011b, 2012b, 2013). Assuming a normal distribution at the significance level of $P = 0.05$, a positive Mann–Kendal statistic Z larger than 1.96 indicates a significant increasing trend, while a Z lower than -1.96 indicates a significant decreasing trend. Critical Z values of ± 1.64 and ± 2.58 are used for the significance of $P = 0.1$ and 0.01, respectively. The difference for factors between 1960–1990 and 1991–2013 was analyzed using one-way analysis of variance (ANOVA), the least significant difference test, and the F test ($P < 0.05$) in SPSS 19.0.

RESULTS

Estimation of reference ET_a based on the water balance method

Figure 2 shows the variation of the four soil moisture products against observations. Compared to the observations, the soil moisture from MOS (SM_{MOS}) exhibited a relatively high magnitude of interannual variation and underestimated all stations except for station 2 (Figure 2). Furthermore, it exhibited poor performance in capturing the variation of soil moisture. Soil moisture from Noah2 and CLM

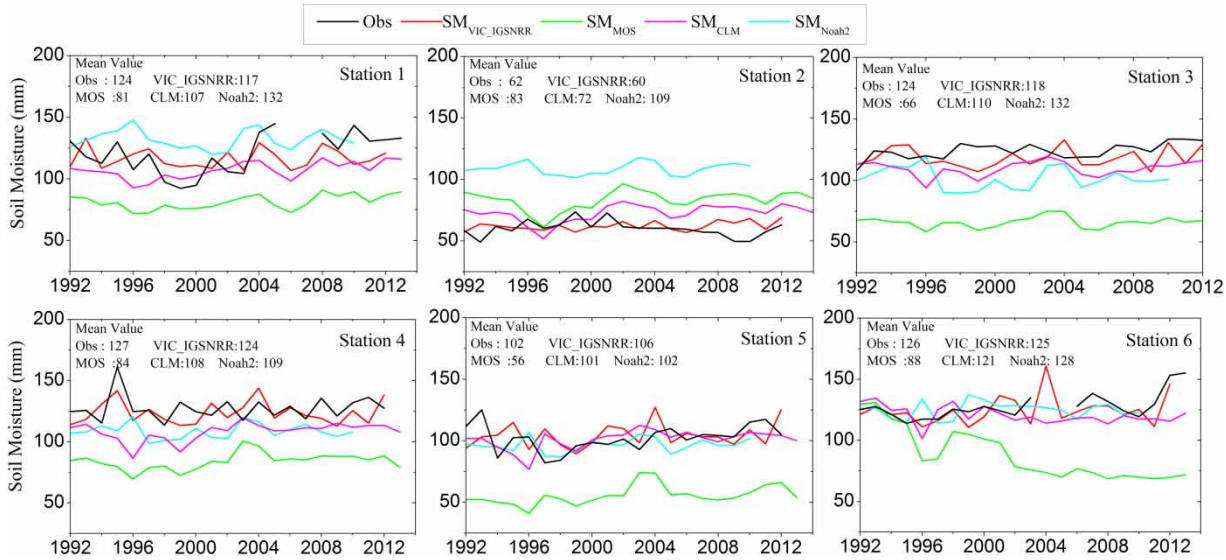

Figure 2 | Comparisons of annual soil moisture between observations at agrometeorological stations and model estimations at the 50 cm depth from 1992 to 2012. The four soil moisture models include VIC-IGSNRR (SM_{IGSNRR}), GLDAS MOS (SM_{MOS}), GLDAS Noah (SM_{Noah}), and GLDAS CLM (SM_{CLM}).

(SM_{Noah2} and SM_{CLM}) showed similar variations without significant differences ($p > 0.1$). The soil moisture from VIC-IGSNRR ($SM_{VIC-IGSNRR}$) had better performance in capturing both the variation and amplitude at all stations except for a small overestimation at station 5. These results agreed well with previous studies that the VIC model demonstrated a better performance in simulating the soil moisture than other models (Maurer *et al.* 2002; Lettenmaier & Famiglietti 2006). The better performance of $SM_{VIC-IGSNRR}$ was reasonable as it incorporated more grounded hydrological and meteorological observations in China (Zhang *et al.* 2014), which enhanced the accuracy of the soil moisture. Therefore, the VIC-IGSNRR model was applied to evaluate TWSC (ΔS) for the period from 1960 to 2012 at a 2 m depth. The annual TWSC and reference ET_a are shown in Figure 3.

TSMC can be assumed to be negligible at multiyear or a longer time scale (Figure 3(a)). However, this assumption cannot be applied to evaluate ET_a at finer temporal scales (e.g., annual or monthly), as neglecting the interannual

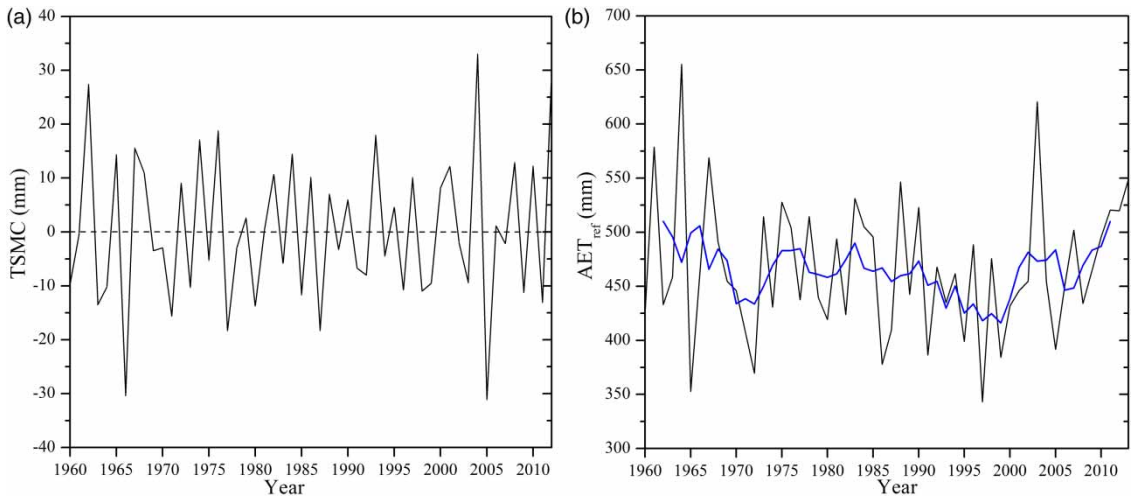

Figure 3 | Time series of annual (a) TSMC estimated from VIC-IGSNRR and (b) ET_a estimated from the water balance method; the (blue) line shows the 5-year moving average. Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/nh.2016.120.

TWSC in the water balance method would significantly undermine the robustness for validation of interannual ET_a variation (Han *et al.* 2015). The variations of TWSC in our study region ranged from -31.1 mm yr^{-1} to 32.9 mm yr^{-1}, accounting for 6% to 7% of the multiyear average precipitation. Thus, ΔS cannot be neglected when applying the water balance method to estimate the reference ET_a in the MYRB. After that, the reference ET_a was calculated based on the water balance method following Equation (1). The estimated annual reference ET_a showed a large variation for the last 50 years (Figure 3(b)). The multiyear average reference ET_a was approximately 466.9 mm yr^{-1}, accounting for more than 90% of the precipitation.

Evaluation performance of different ET_a products

The water balance method has been widely applied to estimate the ET_a at basin or regional scale (Hobbins *et al.* 2001; Xue *et al.* 2013). Comparisons of the ET_a from the four category models and the reference ET_a are shown in Figure 4. The statistical information of model performance is listed in Table 2. All the five LSMs accurately reproduced the annual variation with the reference ET_a, with a CORR larger than 0.57, a range from -72.5 mm yr^{-1} to 27.4 mm yr^{-1} for BIAS, and from 51.8 mm yr^{-1} to 86.2 mm yr^{-1} for RMSE. AET_{Noah1} captured the mean value with the smallest BIAS and largest CORR (-0.2 mm yr^{-1} and 0.63, respectively), followed by AET_{MOS}, with a small overestimation (25.0 mm yr^{-1}) but with a better RMSE/CORR (51.8 mm yr^{-1}/0.62). AET_{VIC}, similar to AET_{MOS}, was approximately overestimated by 6%, and had a good relation with the reference ET_a (CORR = 0.58). Although AET_{CLM} and AET_{Noah2} consistently were underestimated with BIAS values of -72.5 mm yr^{-1} and -54.5 mm yr^{-1}, respectively, they associated well with relatively high CORR (0.59 for AET_{CLM} and 0.57 for AET_{Noah2}). In particular, note that AET_{Noah2} had a longer time series (since 1960) than the other LSMs (since 1979). What is more, the performance of AET_{Noah2} was improved when the start year was changed to 1979 (BIAS/RMSE/CORR -18.9 mm yr^{-1}/48.5 mm yr^{-1}/0.61). This demonstrated that the LSM-Noah model was the best one in depicting the variation of reference ET_a, which was similar to the result of previous researchers (Xue *et al.* 2013; Li *et al.* 2014). Therefore, ET_{Noah2} was

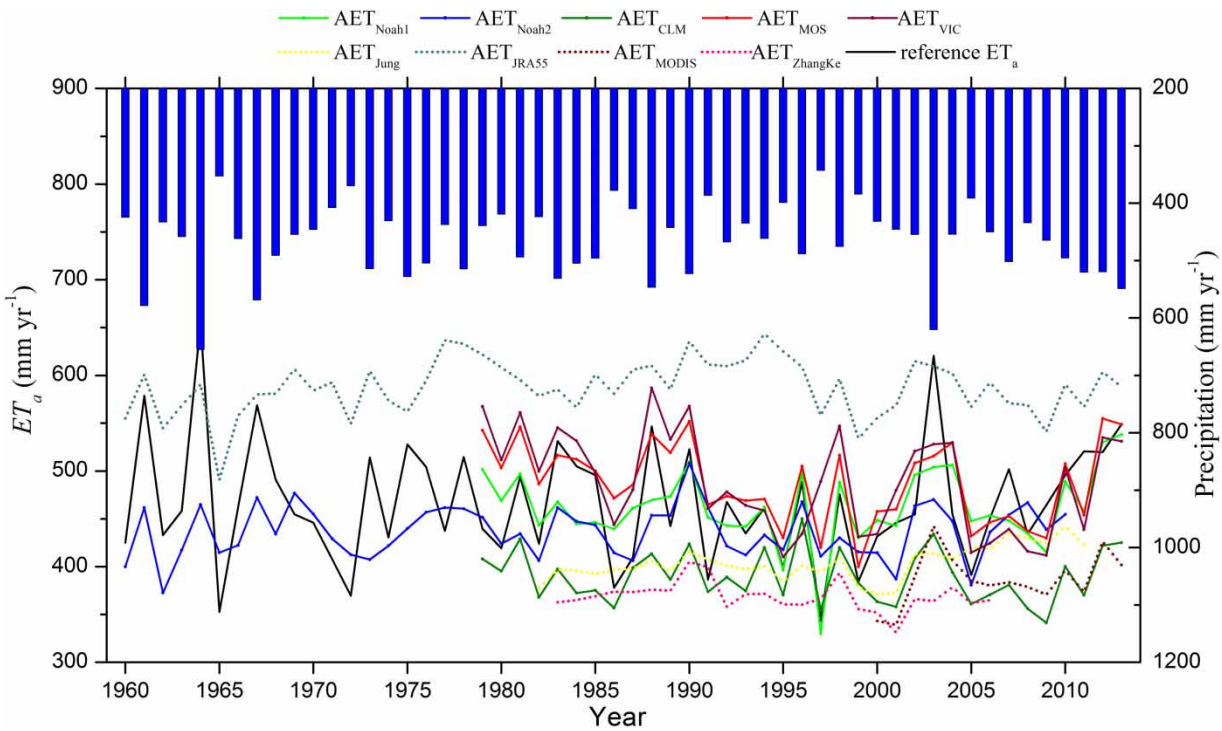

Figure 4 | Time series of yearly ET_a from nine products and corresponding precipitation for the MYRB from 1960 to 2013.

Table 2 | Statistical summary of model performance in simulating ET_a, based on comparison with reference evapotranspiration

	AET_{prod} (reference ET_a)	BIAS	RMSE	CORR
AET_{Noah1}	462 (462)	**−0.2**	56.3	0.63
AET_{Noah2}	437 (467)	−54.5	58.9	0.57
AET_{CLM}	390 (462)	−72.5	86.2	0.59
AET_{MOS}	487 (462)	25.0	**51.8**	0.62
AET_{VIC}	489 (462)	27.4	56.8	0.58
AET_{JUNG}	404 (459)	−54.7	76.9	0.35
AET_{JRA55}	587 (467)	120.6	135.0	0.31
AET_{MODIS}	387 (481)	−94.0	103.6	**0.66**
$AET_{ZhangKe}$	369 (455)	−86.1	106.1	0.08

Note: The comparison was carried out according to the temporal span shown in Table 1. Smaller BIAS and RMSE and larger CORRs are in bold font.

recommended for application at the half century time span, and AET_{Noah1} for over the past three decades in the MYRB.

The MTE method, based on eddy covariance and highly accurate for measuring exchanges of carbon dioxide, water vapor, and energy between the biosphere and atmosphere as well as validating the other models (Fisher *et al.* 2008; Vinukollu *et al.* 2011), was employed in the AET_{JUNG}. Therefore, AET_{JUNG} was expected to have a higher capacity for capturing the reference ET_a variation than other models; however, the performance of AET_{JUNG} was moderate with underestimation of the reference ET_a by 11.8% and had a larger BIAS/RMSE (−54.7/76.9 mm yr^{-1}) and a relatively low CORR (0.35). The two DMs (AET_{MODIS} and $AET_{ZhangKe}$) both performed adequately, underestimating (19.5% and 18.9%, respectively) the ET_a variation with relatively large BIAS/RMSEs (−94.0/103.6 mm yr^{-1} for AET_{MODIS} and −86.1/106.1 mm yr^{-1} for $AET_{ZhangKe}$). Although AET_{MODIS} performed poorly with regard to BIAS/RMSE, it accurately reproduced the interannual variation of the reference ET_a, with the highest CORR (0.66). Therefore, AET_{MODIS} is recommended when higher spatial resolution is required. In contrast, interannual variation of $AET_{ZhangKe}$ was poorly associated with the lowest CORR. However, AET_{JRA55} from the reanalysis model had poor performance, strongly overestimating the ET_a by 25.7% with the largest BIAS/RMSE (120.6/135.0 mm yr^{-1}).

Therefore, compared to the other three categories, the LSMs performed the best in capturing the reference ET_a variation, especially for the Noah model. AET_{JUNG} had a good performance in BIAS and RMSE. The DMs (i.e., AET_{MODIS} and $AET_{ZhangKe}$) did not perform well compared to the LSM and MTE models. However, AET_{JRA55} from the reanalysis model performed poorly in capturing the variation and amplitude of the reference ET_a.

Possible reasons of performance for ET_a products

There are several reasons for the annual differences in ET_a, including model types (e.g., LSM for GLDAS, machine-learning algorithm for the MTE method, the PM method for the DMs), forcing data (listed in Table 3) and parameter optimization. Compared to other ET_a categories, the LSMs, especially the Noah and Mosaic models, captured the variability of the reference ET_a fairly well. In the LSMs, the impact of soil moisture stress on bare soil evaporation and plant transpiration was considered. Soil evaporation happens only when the top soil can draw water from lower layers (Chen *et al.* 1996). Plant transpiration was also constrained by canopy interception and conductance (Koster & Suarez 1994). The Noah and Mosaic models could successfully simulate the energy and water flux, as reported by several studies (Chen *et al.* 1996; Koster & Suarez 1996). The VIC model was a macro-scale, semi-distributed, hydrological model (Liang *et al.* 1994), which calculated the infiltration, evaporation, and runoff variation within

Table 3 | Main forcing fields for the four categories of ET_a models: AET_{Noah1}, AET_{Noah2}, AET_{CLM}, and AET_{MOS} included in GLDAS-LSMs, AET_{JUNG} included in MTE, AET_{JRA55} included in reanalysis dataset, and AET_{MODIS}, $AET_{ZhangKe}$ included in DMs

Models	Main forcing fields
GLDAS-LSMs	Precipitation, radiation, temperature, humidity, wind, air pressure, soil moisture, satellite
Eddy covariance method	Flux tower data, FAPAR, radiation, precipitation, temperature, humidity, potential evapotranspiration, land cover
Reanalysis dataset	Precipitation, radiation, microwave imagers, snow depths, temperature
DMs	Surface variable: LAI (for AET_{MODIS}), NDVI (for $AET_{ZhangKe}$), FAPAR, enhanced vegetation index, surface albedo, and land cover
	Meteorological reanalysis data: radiation, air temperature

each grid cell (Franchini & Pacciani 1991). Therefore, it depicted the hydrological process with high accuracy and has been widely used in large river basins around the world (Sheffield *et al.* 2006; Livneh *et al.* 2013). This model also exhibited good performance in the MYRB.

In addition, the different forcing data used in GLDAS Version 1 and 2 may impact the performance of the LSMs' ET_a: the GLDAS-1, including AET_{MOS}, AET_{CLM}, AET_{Noah1}, and $AET_{VIC\text{-}IGSNRR}$, was forced with a combination of NCEP's Global Data Assimilation System, disaggregated CMAP precipitation and Air Force Weather Agency radiation. The GLDAS-2 was forced with the Global Meteorological Forcing Dataset from Princeton University (Sheffield *et al.* 2006). To further clarify this situation, we selected two key variables (precipitation and downward shortwave radiation flux) from Noah2 and JRA55 models to detect their impact on model performance. For the Noah2 model, the annual precipitation (Pre_Noah2) was compared with that from CMA (Pre_CMA, Figure 5(a)). Pre_Noah2 exhibited a similar variation as Pre_CMA with only a small underestimation (2.7%) and high CORR (0.93), which may explain the excellent performance of AET_{Noah2} in capturing the reference ET_a variation. The annual radiation from Noah2 (Ra_Noah2) was compared with the radiation from the Institute of Tibetan Plateau Research at the Chinese Academy of Sciences (Ra_ITPCAS), which incorporated radiation observations from ground stations, and had been validated by various independent surface radiation data sources (i.e., quality-controlled radiation data from CMA, the high-resolution data collected through Global Energy and Water Cycle Experiment Asian Monsoon Experiment–Tibet, and the data collected through the Coordinated Energy and Water Cycle Observations Project Asia–Australia Monsoon Project–Tibet) (Chen *et al.* 2011). The results showed that although the Ra_Noah2 was underestimated by 8.9%, it still followed the same annual variation of Ra_ITPCAS (Figure 5(b)) and did not cause a large underestimation of ET_a (Figure 4). This also indicated that the ET_a was limited by water supply and is not sensitive to radiation in semi-arid and semi-humid transition regions. Despite the slight underestimation of Ra_Noah2, the high CORR and low bias may result in an adequate performance in capturing the variation of reference ET_a. Thus, the ET_a from Noah models (AET_{Noah1} and AET_{Noah2}), especially for the advanced Noah2 employing consistent forcing data, showed excellent performance in capturing the annual variation of reference ET_a.

AET_{JRA55} was from the Japanese 55-year reanalysis (JRA-55) dataset of the Japan Meteorological Agency (Kobayashi *et al.* 2015). Although AET_{JRA55} had adopted a series of strategies to improve the capacity for simulating ET_a, it still used the same forcing field as JRA-25 (i.e., atmospheric parameters such as pressure, temperature, humidity and wind, precipitation, downward solar and long wave radiation fluxes, and total cloud cover) (Nakaegwa 2008; Kobayashi *et al.* 2015), which can cause uncertainties in the reanalysis dataset. Further analysis of the precipitation (Pre_JRA55 in Figure 5(a)) and radiation (Ra_JRA55 in

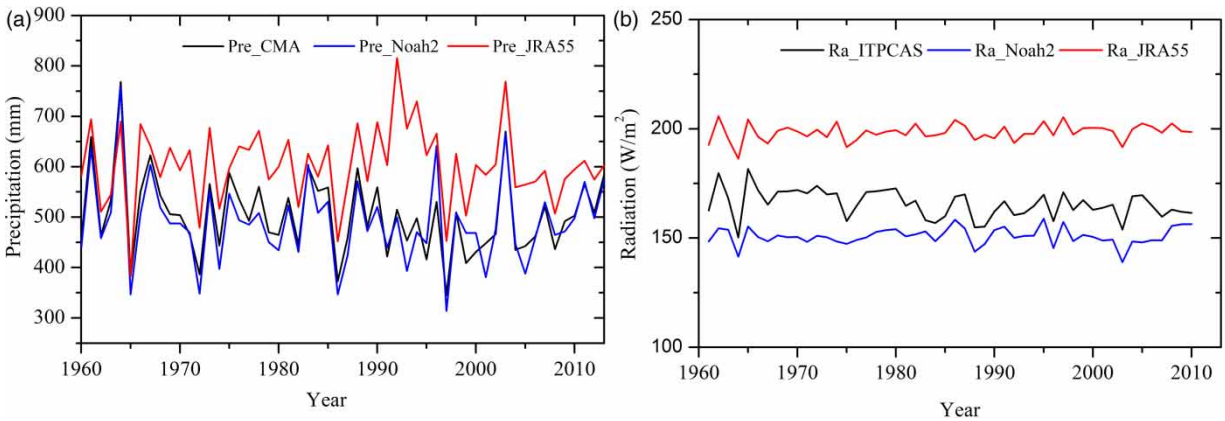

Figure 5 | Comparison of (a) annual precipitation from China Meteorological Administration (Pre_CMA), GLDAS with Noah LSM-2 (Pre_Noah2), and Japanese 55-year reanalysis (Pre_JRA55); (b) annual radiation from Institute of Tibetan Plateau Research, Chinese Academy of Sciences (Ra_ITPCAS), Ra_Noah2, and Ra_JRA55.

Figure 5(b)) showed that both were overestimated (19.8% and 19.7%, respectively) and had relatively low CORR (0.67 and 0.66, respectively). This can explain the large overestimation of AET_{JRA55} and relatively low performance in capturing the variation of the reference ET_a. Therefore, great caution should be paid to hydrological variables before using them in reanalysis.

The AET_{JUNG} was trained with *in situ* flux tower measurement with an average duration of only 2 years (Jung *et al.* 2010). These measurements were strongly dependent on spatial gradients when reproducing the trends and variation in ET_a. In addition, the flux towers were sparse throughout our study region (Jung *et al.* 2010, Supplementary Figure 1), which may introduce some uncertainties to AET_{JUNG} in capturing the reference ET_a variation. Both of the ET_a from DMs were underestimated. The $AET_{ZhangKe}$ did not consider the evaporation from canopy interception (Zhang *et al.* 2010), which may be a reason for the difference between the reference ET_a and $AET_{ZhangKe}$. The Chinese government adopted a series of strategies to restore the ecosystem since 1999 (Zhang *et al.* 2000; FAO 2010) which had significantly increased vegetation coverage (Chen *et al.* 2015; Wang *et al.* 2015) with a higher LAI and water storage capacity of the canopy (Zhang *et al.* 2007). Therefore, without considering the canopy interception, ET_a could be underestimated. In addition, insufficient accuracy of remotely sensed radiation employed in the $AET_{ZhangKe}$ models may contribute to the uncertainty of ET_a (Zhang *et al.* 2012). Although AET_{MODIS} had a limited time span, it showed better performance in capturing the reference ET_a variability, which may result from the improvement of the ET_a algorithm (Mu *et al.* 2007, 2011). In addition, the forcing data in the AET_{MODIS} model maintained the spatial coherence. For example, the FPAR, LAI, albedo, enhanced vegetation index (EVI), and land cover were all from a single MODIS product, which could also explain the improvement in the estimation accuracy of AET_{MODIS}.

DISCUSSION

Comparison of the relationship between precipitation and ET_a products

Evaporation from canopy interception and the uppermost soil layer and transpiration by the plant were all highly related to precipitation. Precipitation was essential for deriving ET_a and considered as the main input parameter in models. Especially in semi-arid and semi-humid regions where water supply limited ET_a, the degree of ET_a model dependence on precipitation could reflect its performance in estimating ET_a (Zhang *et al.* 2012). The CORR of precipitation and the nine ET_a products are listed in Table 4.

All of the ET_a from LSMs showed significant correlation ($p < 0.01$, Table 4) with precipitation and larger CORR (>0.52), which indicates that LSMs were more dependent on precipitation than the other methods. Precipitation, one of the key forcing fields in LSMs (Table 3), was improved significantly by using the near real-time data, which were derived from the geostationary satellite infrared cloud-top temperature measurements and various merged microwave observation datasets (Turk *et al.* 2000). In addition, the water balance method and improved techniques (i.e., adjusting the soil moisture and precipitation) were applied to constrain the LSMs (Koster & Suarez 1996; Rodell *et al.* 2004). Direct evaporation from the bare soil in the LSMs was closely related to the soil water content via the soil water diffusivity equation for moisture transport (Zeng & Decker 2009), which provided an effective and direct coupling between groundwater and surface water. Therefore, the variation of LSMs' ET_a followed the variation of precipitation and responded to changes in soil moisture in the MYRB.

The EC-based AET_{JUNG} and reanalysis AET_{JRA55} exhibited low CORR with precipitation (0.30 and 0.28, respectively). The precipitation variable in the JRA55 model was derived from the reanalysis dataset, which was produced

Table 4 | Pearson correlation between ET_a from global products and precipitation

Variables	AET_{Noah1}	AET_{Noah2}	AET_{CLM}	AET_{MOS}	AET_{VIC}	AET_{JUNG}	AET_{JRA55}	AET_{MODIS}	$AET_{ZhangKe}$
Precipitation	0.57**	0.52**	0.54**	0.63**	0.59**	0.30	0.28*	0.62*	0.12

Note: Significance levels are indicated by *$P > 0.05$ and **$P < 0.01$, respectively.

by combining the observation and forecast model results and had numerous uncertainties (Bosilovich *et al.* 2011). AET_{JUNG} was a data-driven estimation of ET_a and had low dependence on precipitation. There was a large difference in CORR with precipitation for the two DMs, high correlation for AET_{MODIS} (CORR = 0.62, $P < 0.05$) and low correlation for $AET_{ZhangKe}$ (CORR = 0.12, $P > 0.05$). Both of the DMs used remotely sensed vegetation information (i.e., LAI or NDVI) and the PM method forced by meteorological data without precipitation (Mu *et al.* 2009; Zhang *et al.* 2010). Even without precipitation in AET_{MODIS}, there was a general agreement, which may be a result of a series of improvements in the terrestrial evapotranspiration algorithm (Mu *et al.* 2007, 2011), for instance, separating dry canopy surfaces from wet ones and dividing soil surface into saturated wet surface and moist surface, which was explicitly linked to precipitation. Therefore, we could conclude that the more dependent on precipitation of ET_a models, the better the performance in capturing the reference ET_a in semi-arid and semi-humid transition regions. The correlation between precipitation and ET_a from models can be considered as a standard to evaluate models' performance to some extent.

Trend of reference ET_a and performance of ET_a products

The reference ET_a exhibited large variations from 1960 to 2013 in the MYRB (Figure 3(b)). To obtain the temporal trend of the reference ET_a and detect which products can capture the tendencies, the study time span was defined from 1983 to 2013, considering the time consistency of each ET_a product (Table 1). In addition, there was a point (around 1999, Figure 3(b)) where the reference ET_a trend changed (i.e., decreasing from 1983 to 1998 and then increasing from 1999 to 2013). Some previous studies showed global land ET_a decreasing from the 1980s to the late 1990s (Jung *et al.* 2010; Mueller *et al.* 2013; Zeng *et al.* 2014) and then increasing significantly during the last two decades of the 20th century (Jung *et al.* 2010; Zhang *et al.* 2012). Therefore, 1999 was chosen as the change year to detect ET_a temporal variation across the MYRB. The annual mean ET_a and trend of the reference ET_a associated with ET_a products are listed in Table 5.

The annual average reference ET_a changed abruptly around 1999 with a mean value of 464.7 mm yr^{-1} from 1983 to 2013, and the mean values for 1983–1998 (pre-1999) and 1999–2013 (post-1999) were 455.5 mm yr^{-1} and 474.6 mm yr^{-1}, respectively (Table 5). Although the ET_a from products showed significant interannual variability (Figure 4), there was no significant difference between pre-1999 and post-1999 by ANOVA ($p > 0.1$) test. The annual trend in the reference ET_a slope was found to change from negative (−4.54 mm yr^{-2}) for the pre-1999 period to positive (6.50 mm yr^{-2}) for the post-1999 period (Table 5). Although there was a large increase in amplitude during the post-1999

Table 5 | Mean (mm yr^{-1}) and trend (mm yr^{-2}) for the nine ET_a products and reference ET_a during the three periods (1983–1998, 1999–2013, and 1983–2013)

	1983–1998		1999–2013		1983–2013	
Period ET_a	**Mean**	**Trend**	**Mean**	**Trend**	**Mean**	**Trend**
Reference ET_a	455.5	−4.54	474.6	6.50***	464.7	1.04
AET_{Noah1}	451.5	−1.86	469.4	3.16	460.1	0.97
AET_{Noah2}	440.2	−1.20	435.8	3.47	438.3	0.18
AET_{CLM}	392.3	1.43	384.6	1.22	388.6	−0.04
AET_{MOS}	490.3	−3.23	475.6	4.17	483.2	−0.68
AET_{VIC}	495.9	**−3.88**	469.2	1.71	483.0	−1.63
AET_{JRA55}	602.5	1.29	578.3	1.04	590.8	−0.88*
AET_{JUNG}	399.6	0.21	411.4	**5.14*****	404.9	**1.10*****
$AET_{ZhangKe}$	373.8	0.25	359.3	3.01	368.9	−0.62
AET_{MODIS}	–	–	387.0	2.3	382.5	2.30

Note: The significant level of ET_a trend was detected based on Mann–Kendall test. The trend values close to reference ET_a trend are in bold font.
Significance levels are indicated by *$p < 0.1$, and ***$p < 0.01$, respectively.

period, it was offset by the negative trend of pre-1999, resulting in no significant trend for the MYRB during the entire period from 1983 to 2013 (1.04 mm yr^{-2}, $p > 0.1$).

Regarding the ET_a products for the pre-1999 period, except for AET_{CLM}, all the ET_a from LSMs, particularly AET_{VIC}, captured the reference ET_a negative trend, whereas the other products showed a positive trend. For the post-1999 period, all the ET_a products reproduced the reference ET_a's positive trend, especially for AET_{JUNG} and AET_{MOS} (5.14 and 4.17 mm yr^{-2}, respectively). For the entire period (1983 to 2013), the ET_a from Noah, JUNG, and MODIS showed positive trends, while the other methods showed negative trends. AET_{JUNG} performed best in capturing the trend amplitude. The Noah model, in general, exhibited good performance in capturing the ET_a trend for all the periods, while AET_{VIC} was best for pre-1999, and AET_{JUNG} was best for capturing the ET_a amplitude in both the post-1999 and 1983–2013 periods.

Further analysis revealed that the annual average ET_a made up 92.7% of the total precipitation for the pre-1999 period and 96.4% for the post-1999 period. The abrupt change in ET_a may have resulted from the series of strategies mandated by the Chinese Government since 1999 that aimed to restore the ecosystem, prevent severe soil erosion, and ensure sustainable development. These strategies led to increasing the land surface roughness (Liu & Zhang 2013) and then changed the streamflow and ET_a in the MYRB (Huang *et al.* 2003; Zhang *et al.* 2007).

Uncertainties and suggestions

Potential uncertainties from various sources may undermine the confidence of this study. First, the reference ET_a, based on the water balance method and with the assumption that the deep groundwater storage was not changed at annual scale, can cause uncertainties. Soil moisture was influenced by field capacity, soil bulk density, location (longitude and latitude), meteorological elements, natural factors, and human activities, all of which could cause various uncertainties (Li *et al.* 2005; Liu *et al.* 2008). The validation of soil moisture for all layers needs to be tested in this region; however, we only used six moisture stations at the 50 cm depth from 1992 to 2012. In addition, human activities, such as reservoir construction, irrigation and extraction of groundwater, may impact deep groundwater. Second, the time scale difference used for evaluating ET_a performance may impact conclusions. For instance, the AET_{Noah2} started in 1948 was only evaluated from 1960. In contrast, the ET_a products from other LSMs were evaluated from 1979, and an even shorter time scale for AET_{MODIS} and $AET_{ZhangKe}$.

Despite the above uncertainties, the performance of the nine ET_a products was evaluated based on the water balance method in the MYRB. We also recommend that more attention be paid to the ET_a and runoff response to human activities and climate change (Zhang *et al.* 2013; Yang *et al.* 2015a), including quantitatively analyzing the relative contributions of different factors (e.g., irrigation, water diversion, global warming, land use change, and CO_2 enrichment) to ET_a. The mechanisms influencing the spatiotemporal variation in ET_a and its components (e.g., interception, soil evaporation, and transpiration) at different time scales (Koster *et al.* 2004; Good *et al.* 2015) should be further investigated; this will be beneficial for water resources management and irrigation scheduling in the MYRB. In addition, projections of the ET_a spatiotemporal trend in the future under different scenarios (Meehl *et al.* 2007; IPCC 2013) were useful for informing policy, such as whether to restart the GTGP in the Loess Plateau (Chen *et al.* 2015). More reliable knowledge about ET_a will be helpful to better understand hydrological processes and improve regional sustainable development.

CONCLUSION

The multiyear actual evapotranspiration (reference ET_a) was evaluated with the water balance method, which considered the soil moisture change derived from the VIC-IGSNRR model. The performance of the four categories of global ET_a products was investigated in the MYRB. The results indicated that the total water storage change (ΔS), estimated from the soil moisture of the VIC-IGSNRR model, varied significantly and the TWSC cannot be neglected when evaluating the ET_a by the water balance method at annual scale. The ET_a from LSMs performed excellently in capturing the variation of the reference ET_a among the four categories of products.

The ET_a from Noah (especially for Noah1) and MOS demonstrated the best performance. AET_{Noah2} is recommended for application in large regions and at a half-century scale, whereas AET_{Noah1} and AET_{MOS} are recommended for the past three decades and AET_{MODIS} for the past 10-year scale. In addition, a good correlation was found between LSMs and precipitation, indicating that in arid and semi-arid regions, considering the precipitation in ET_a models can significantly improve the performance in capturing the variation of the reference ET_a. The AET_{JUNG} model, which was based on eddy covariance, performed relatively well, with a small negative BIAS and a little weak relationship to the reference ET_a, which demonstrated its dependence on observation data and independence from precipitation. The two DMs demonstrated fair performance with small underestimation. The AET_{MODIS} model, although a short time span, showed a stronger capacity for capturing the reference ET_a than the other models. However, the AET_{JRA55} model performed poorly, with the largest positive BIAS and RMSE; this was caused by higher precipitation and radiation input.

The variation of the reference ET_a shifted around 1999. Further analysis found that the reference ET_a decreased for the pre-1999 period (-4.54 mm yr^{-2}) and increased for the post-1999 period (6.50 mm yr^{-2}). However, the increase during the post-1999 period was offset by the decrease during the pre-1999 period, resulting in no significant change during the entire period from 1983 to 2013. Generally, the ET_a from the Noah model, especially Noah1, better captured the ET_a trend and amplitude for the three periods. The AET_{VIC} model best captured the ET_a amplitude for the pre-1999 period, while the AET_{JUNG} model was best for the post-1999 and 1983–2013 periods. The underlying reasons for ET_a changes may be the implementation of several forest strategies. Comprehensive evaluation and comparison of various ET_a products was valuable for the application of ET_a products for regional water resource management and guiding future research on ET_a estimation. Limited by the coarse temporal and spatial resolution of ET_a products in this study, more accurate regional ET_a models with higher temporal and spatial resolution forcing data would be necessary to improve the estimation performance of ET_a in further work.

ACKNOWLEDGEMENTS

This research was financially supported by the Natural Science Foundation of China (No. 41330529, 41501032, 41571024, and 41301496). We thank two anonymous reviewers for their comments and suggestions to improve the quality of the manuscript and Xiaomang Liu, Guotao Dong, Kaiwen Wang, and Xiaocong Liu for their great help regarding the clarity of this paper.

REFERENCES

Bastiaanssen, W. G. M. 2000 SEBAL based sensible and latent heat fluxes in the irrigated Gediz Basin, Turkey. *J. Hydrol.* **229**, 87–100.

Bonan, G. B., Pollard, D. & Thompson, S. L. 1992 Effects of boreal forest vegetation on global climate. *Nature* **359**, 716–718.

Bosilovich, M. G., Robertson, F. R. & Chen, J. 2011 Global energy and water budgets in MERRA. *J. Climate* **24**, 5721–5739.

Brutsaert, W. 2005 *Hydrology: An Introduction*. Cambridge University Press, Cambridge, UK.

Brutsaert, W. & Parlange, M. 1998 Hydrologic cycle explains the evaporation paradox. *Nature* **396**, 30. DOI: 10.1038/23845.

Cai, X. T., Yang, Z. L., Xia, Y. L., Huang, M. Y., Wei, H. L., Leung, R. & Ek, M. B. 2014 Assessment of simulated water balance from Noah, Noah-MP, CLM, and VIC over CONUS using the NLDAS test bed. *J. Geophys. Res. Atmos.* **119**, 13751–13770.

Chen, F., Mitchell, K., Schaake, J., Xue, Y., Pan, H.-L., Koren, V., Duan, Q. Y., Ek, M. & Betts, A. 1996 Modeling of land surface evaporation by four schemes and comparison with FIFE observations. *J. Geophys. Res. Atmos.* **101**, 7251–7268.

Chen, Y., Yang, K., He, J., Qin, J., Shi, J., Du, J. & He, Q. 2011 Improving land surface temperature modeling for dry land of China. *J. Geophys. Res. Atmos.* **116**, D20104.

Chen, Y., Yang, K., Qin, J., Zhao, L., Tang, W. & Han, M. 2013 Evaluation of AMSR-E retrievals and GLDAS simulations against observations of a soil moisture network on the central Tibetan Plateau. *J. Geophys. Res. Atmos.* **118**, 4466–4475.

Chen, Y., Wang, K., Lin, Y., Shi, W., Song, Y. & He, X. 2015 Balancing green and grain trade. *Nat. Geosci.* **8**, 739–741.

Cleugh, H. A., Leuning, R., Mu, Q. & Running, S. W. 2007 Regional evaporation estimates from flux tower and MODIS satellite data. *Remote Sens. Environ.* **106**, 285–304.

Corbari, C., Mancini, M., Su, Z. & Li, J. 2014 Evapotranspiration estimate from water balance closure using satellite data for the Upper Yangtze River basin. *Hydrol. Res.* **45**, 603–614.

Ebita, A., Kobayashi, S., Ota, Y., Moriya, M., Kumabe, R., Onogi, K., Harada, Y., Yasui, S., Miyaoka, K. & Takahashi, K. 2011 The Japanese 55-year Reanalysis 'JRA-55': an interim report. *Scientific Online Letters on the Atmosphere Sola.* **7**, 149–152.

EI Haj EI Tahir, M., Wang, W., Xu, C., Zhang, Y. & Singh, V. 2012 Comparison of methods for estimation of regional actual evapotranspiration in data scarce regions: Blue Nile Region, Eastern Sudan. *J. Hydrol. Eng.* **17**, 578–589.

Famiglietti, J. S. 2013 Remote sensing of terrestrial water storage, soil moisture and surface waters. *Geophysical Monograph Series* **150**, 197–207.

FAO 2010 *Global Forest Resources Assessment 2000 Main Report*. FAO Forestry Paper.

Ferreira, V. G., Gong, Z., He, X., Zhang, Y. & Andam-Akorful, S. A. 2013 Estimating total discharge in the Yangtze River Basin using satellite-based observations. *Remote Sensing* **5**, 3415–3430.

Fisher, J. B., Tu, K. P. & Baldocchi, D. D. 2008 Global estimates of the land–atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites. *Remote Sens. Environ.* **112**, 901–919.

Franchini, M. & Pacciani, M. 1991 Comparative analysis of several conceptual rainfall-runoff models. *J. Hydrol.* **122**, 161–219.

Gao, G., Chen, D., Xu, C. & Simelton, E. 2007 Trend of estimated actual evapotranspiration over China during 1960–2002. *J. Geophys. Res. Atmos.* **112**, 1–8.

Gao, G., Xu, C., Chen, D. & Singh, V. 2012 Spatial and temporal characteristics of actual evapotranspiration over Haihe River basin in China. *Stoch. Env. Res. Risk A* **26**, 655–669.

Gokmen, M., Vekerdy, Z., Verhoef, A., Verhoef, W., Batelaan, O. & van der Tol, C. 2012 Integration of soil moisture in SEBS for improving evapotranspiration estimation under water stress conditions. *Remote Sens. Environ.* **121**, 261–274.

Good, S. P., Noone, D. & Bowen, G. 2015 Hydrologic connectivity constrains partitioning of global terrestrial water fluxes. *Science* **349**, 175–177.

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

Hemakumara, H., Chandrapala, L. & Moene, A. F. 2003 Evapotranspiration fluxes over mixed vegetation areas measured from large aperture scintillometer. *Agri. Water Manage.* **58**, 109–122.

Hetherington, A. M. & Woodward, F. I. 2003 The role of stomata in sensing and driving environmental change. *Nature* **424**, 901–908.

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**, 1389–1403.

Hogue, T. S., Bastidas, L. A., Gupta, H. V. & Sorooshian, S. 2006 Evaluating model performance and parameter behavior for varying levels of land surface model complexity. *Water Resour. Res.* **42** (8), W08430.

Huang, M., Zhang, L. & Gallichand, J. 2003 Runoff responses to afforestation in a watershed of the Loess Plateau, China. *Hydrol. Process.* **17**, 2599–2609.

IPCC 2013 *Summary for Policymakes: The Physical Science Basis, Contribution of Working Group I to the IPCC Fifth Assessment Report Climate Change*.

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.

Kendall, M. G. 1948 *Rank Correlation Methods*. C Griffin, London.

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. Meteorol. Soc. Jpn* **93**, 5–48.

Koster, R. & Suarez, M. 1994 The components of a 'SVAT'scheme and their effects on a GCM's hydrological cycle. *Adv. Water Resour.* **17**, 61–78.

Koster, R. & Suarez, M. 1996 Energy and water balance calculations in the Mosaic LSM. *NASA Tech. Memo.* **104606**, 59.

Koster, R. D., Dirmeyer, P. A., Guo, Z., Bonan, G., Chan, E., Cox, P., Gordon, C. T., Kanae, S., Kowalczyk, E. & Lawrence, D. 2004 Regions of strong coupling between soil moisture and precipitation. *Science* **305**, 1138–1141.

Lakshmi, V., Hong, S., Small, E. E. & Chen, F. 2011 The influence of the land surface on hydrometeorology and ecology: new advances from modeling and satellite remote sensing. *Hydrol. Res.* **42**, 95–112.

Lettenmaier, D. P. & Famiglietti, J. S. 2006 Hydrology: water from on high. *Nature* **444**, 562–563.

Li, H., Robock, A., Liu, S., Mo, X. & Viterbo, P. 2005 Evaluation of reanalysis soil moisture simulations using updated Chinese soil moisture observations. *J. Hydrometeorol.* **6**, 180–193.

Li, R., Yang, W., Li, B., Yang, Q., Yu, C., Zou, H., Zhang, Z., Zhou, P., Jia, H. & Liang, Y. 2008 *Research and Prospect on the Loess Plateau of China*. Science Press, Beijing, China (in Chinese).

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. Atmos.* **119**, 13079–13095.

Liang, X., Lettennmaier, D. P., Wood, E. F. & Burges, S. J. 1994 A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *J. Geophys. Res. Atmos.* **99**, 14415–14428.

Liu, X. & Zhang, D. 2013 Trend analysis of reference evapotranspiration in Northwest China: The roles of changing wind speed and surface air temperature. *Hydrol. Process.* **27**, 3941–3948.

Liu, S., Mao, L., Mo, X., Zhao, W. & Lin, Z. 2008 Analysis of spatial variability of soil moisture and its driving force factors in the Shanxi-Henan Region along the Yellow River. *Climate Environ. Res.* **12**, 645–657.

Liu, X., Luo, Y., Zhang, D., Zhang, M. & Liu, C. 2011a Recent changes in pan-evaporation dynamics in China. *Geophys. Res. Lett.* **38**, L13404.

Liu, X., Zheng, H., Zhang, M. & Liu, C. 2011b Identification of dominant climate factor for pan evaporation trend in the Tibetan Plateau. *J. Geogr. Sci.* **21**, 594–608.

Liu, C., Zhang, D., Liu, X. & Zhao, C. 2012a Spatial and temporal change in the potential evapotranspiration sensitivity to meteorological factors in China (1960–2007). *J. Geogr. Sci.* **22**, 3–14.

Liu, X., Liu, C., Luo, Y., Zhang, M. & Xia, J. 2012b Dramatic decrease in streamflow from the headwater source in the central route of China's water diversion project: climatic variation or human influence? *J. Geophys. Res. Atmos.* **117**, D06113.

Liu, X., Liu, W. & Xia, J. 2013 Comparison of the streamflow sensitivity to aridity index between the Danjiangkou Reservoir basin and Miyun Reservoir basin, China. *Theor. Appl. Climatol.* **111**, 683–691.

Livneh, B., Lin, C., Mishra, V., Andreadis, K., Maurer, E. P. & Lettenmaier, D. P. 2013 A long-term hydrologically based dataset of land surface fluxes and states for the conterminous U.S.: update and extensions. *J. Climate* **26**, 9384–9392.

Long, D., Longuevergne, L. & Scanlon, B. R. 2014 Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites. *Water Resour. Res.* **50**, 1131–1151.

Ma, N., Zhang, Y., Xu, C. & Szilagyi, J. 2015 Modeling actual evapotranspiration with routine meteorological variables in the data-scarce region of the Tibetan Plateau: comparisons and implications. *J. Geophys. Res. Biogeosci.* **120**, 1638–1657.

Maurer, E., Wood, A., Adam, J., Lettenmaier, D. & Nijssen, B. 2002 A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States. *J. Climate* **15**, 3237–3251.

Meehl, G., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchell, J., Stouffer, R. & Taylor, K. 2007 The WCRP CMIP3 Multimodel Dataset: a new era in climate change research. *BAMS* **88**, 1383–1394.

Moiwo, J. P. & Tao, F. 2015 Contributions of precipitation, irrigation and soil water to evapotranspiration in (semi)-arid regions. *Int. J. Climatol.* **35**, 1079–1089.

Mu, Q., Heinsch, F. A., Zhao, M. & Running, S. W. 2007 Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. *Remote Sens. Environ.* **111**, 519–536.

Mu, Q., Jones, L. A., Kimball, J. S., McDonald, K. C. & Running, S. W. 2009 Satellite assessment of land surface evapotranspiration for the pan-Arctic domain. *Water Resour. Res.* **45**, W09420.

Mu, Q., Zhao, M. & Running, S. W. 2011 Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sens. Environ.* **115**, 1781–1800.

Mueller, B., Hirschi, M., Jimenez, C., Ciais, P., Dirmeyer, P. A., Dolman, A. J., Fisher, J. B., Jung, M., Ludwig, F. & Maignan, F. 2013 Benchmark products for land evapotranspiration: LandFlux-EVAL multi-dataset synthesis. *Hydrol. Earth Syst. Sci.* **17**, 3707–3720.

Nakaegwa, T. 2008 Reproducibility of the seasonal cycles of land-surface hydrological variables in Japanese 25-year reanalysis. *Hydrol. Res. Lett.* **2**, 56–60.

New, M., Hulme, M. & Jones, P. 2000 Representing twentieth-century space-time climate variability. Part II: development of 1901–96 monthly grids of terrestrial surface climate. *J. Climate* **13**, 2217–2238.

Oki, T. & Kanae, S. 2006 Global hydrological cycles and world water resources. *Science* **313**, 1068–1072.

Qian, T., Dai, A., Trenberth, K. E. & Oleson, K. W. 2006 Simulation of global land surface conditions from 1948 to 2004. Part I: forcing data and evaluations. *J. Hydrometeorol.* **7**, 953–975.

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, J. P., Lohmann, D. & Toll, D. 2004 The global land data assimilation system. *BAM Meteorol. Soc.* **85**, 381–394.

Sheffield, J., Goteti, G. & Wood, E. F. 2006 Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J. Climate* **19**, 3088–3111.

Shen, Y., Kondoh, A., Tang, C., Zhang, Y., Chen, J., Li, W., Sakura, Y., Liu, C., Tanaka, T. & Shimada, J. 2002 Measurement and analysis of evapotranspiration and surface conductance of a wheat canopy. *Hydrol. Process.* **16**, 2173–2187.

Shutov, V., Gieck, R. E., Hinzman, L. D. & Kane, D. L. 2006 Evaporation from land surface in high latitude areas: a review of methods and study results. *Hydrol. Res.* **37**, 393–411.

Trajkovic, S. 2010 Testing hourly reference evapotranspiration approaches using lysimeter measurements in a semiarid climate. *Hydrol. Res.* **41**, 38–49.

Turk, F. J., Rohaly, G., Hawkins, J., Smith, E., Grose, A., Marzano, F., Mugnai, A. & Levizzani, V. 2000 Analysis and assimilation of rainfall from blended SSM/I, TRMM and geostationary satellite data. In: *Proc. 10th Conf. Satellite Meteorology and Oceanography*, pp. 66–69.

Vinukollu, R. K., Wood, E. F., Ferguson, C. R. & Fisher, J. B. 2011 Global estimates of evapotranspiration for climate studies using multi-sensor remote sensing data: evaluation of three process-based approaches. *Remote Sens. Environ.* **115**, 801–823.

Wang, Q. X., Fan, X. H., Qin, Z. D. & Wang, M. B. 2012 Change trends of temperature and precipitation in the Loess Plateau Region of China, 1961–2010. *Global Planet. Change* **92**, 138–147.

Wang, S., Fu, B., Piao, S., Lu, Y., Ciais, P., Feng, X. & Wang, Y. 2015 Reduced sediment transport in the Yellow River due to anthropogenic changes. *Nature Geosci.* **9**, 38–41.

Xu, C. 2001 Statistical analysis of parameters and residuals of a conceptual water balance model–methodology and case study. *Water Resour. Manage.* **15**, 75–92.

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.

Xu, C., Gong, L., Jiang, T., Chen, D. & Singh, V. 2006 Analysis of spatial distribution and temporal trend of reference

evapotranspiration and pan evaporation in Changjiang (Yangtze River) catchment. *J. Hydrol.* **327**, 81–93.

Xue, B., Wang, L., Li, X., Yang, K., Chen, D. & Sun, L. 2013 Evaluation of evapotranspiration estimates for two river basins on the Tibetan Plateau by a water balance method. *J. Hydrol.* **492**, 290–297.

Yang, T., Xu, C.-Y., Shao, Q., Chen, X., Lu, G.-H. & Hao, Z.-C. 2010 Temporal and spatial patterns of low-flow changes in the Yellow River in the last half century. *Stoch. Env. Res. Risk A* **24**, 297–309.

Yang, S., Xu, K., Milliman, J., Yang, H. & Wu, C. 2015a Decline of Yangtze River water and sediment discharge: Impact from natural and anthropogenic changes. *Scientific Rep.* **5**. DOI: 10.1038/srep12581.

Yang, T., Wang, C., Chen, Y., Chen, X. & Yu, Z. 2015b Climate change and water storage variability over an arid endorheic region. *J. Hydrol.* **529**, 330–339.

Yin, Y., Wu, S. & Zhao, D. 2013 Past and future spatiotemporal changes in evapotranspiration and effective moisture on the Tibetan Plateau. *J. Geophys. Res. Atmos.* **118**, 10850–10860.

Zeng, X. & Decker, M. 2009 Improving the numerical solution of soil moisture-based Richards equation for land models with a deep or shallow water table. *J. Hydrometeorol.* **10**, 308–319.

Zeng, Z., Wang, T., Zhou, F., Ciais, P., Mao, J., Shi, X. & Piao, S. 2014 A worldwide analysis of spatiotemporal changes in water balance-based evapotranspiration from 1982 to 2009. *J. Geophys. Res. Atmos.* **119**, 1186–1202.

Zhang, P., Shao, G., Zhao, G., Master, D. C. L., Parker, G. R., Dunning, J. B. J. & Li, Q. 2000 China's forest policy for the 21st century. *Science* **288**, 2135–2136.

Zhang, X., Zhang, L., McVicar, T. R., Van Niel, T. G., Li, L. T., Li, R., Yang, Q. K. & Wei, L. 2007 Modelling the impact of afforestation on average streamflow in the Loess Plateau, China. *Hydrol. Process.* **22**, 1996–2004.

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**, W09522.

Zhang, K., Kimball, J. S., Kim, Y. & McDonald, K. C. 2011a Changing freeze-thaw seasons in northern high latitudes and associated influences on evapotranspiration. *Hydrol. Process.* **25**, 4142–4151.

Zhang, Q., Xu, C.-Y., Chen, Y. D. & Ren, L. 2011b Comparison of evapotranspiration variations between the Yellow River and Pearl River basin, China. *Stoch. Env. Res. Risk A* **25**, 139–150.

Zhang, Y., Leuning, R., Chiew, F. H. S., Wang, E., Zhang, L., Liu, C., Sun, F., Peel, M. C., Shen, Y. & Jung, M. 2012 Decadal trends in evaporation from global energy and water balances. *J. Hydrometeorol.* **13**, 379–391.

Zhang, D., Liu, X., Liu, C. & Bai, P. 2013 Responses of runoff to climatic variation and human activities in the Fenhe River, China. *Stoch. Env. Res. Risk A* **27**, 1293–1301.

Zhang, X., Tang, Q., Pan, M. & Tang, Y. 2014 A long-term land surface hydrologic fluxes and states dataset for China. *J. Hydrometeorol.* **15**, 2067–2084.

Zhu, Y., Chang, J., Huang, S. & Huang, Q. 2015 Characteristics of integrated droughts based on a nonparametric standardized drought index in the Yellow River Basin, China. *Hydrol. Res.* **47** (2), 454–467. DOI: 10.2166/nh.2015.287.

First received 6 January 2016; accepted in revised form 25 April 2016. Available online 26 May 2016