Evaluating different evapotranspiration products in the middle Yellow River Basin, China

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

ABSTRACT

Actual evapotranspiration (ETa) 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 ETa 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 ETa by the water balance methods. (2) ETa from LSMs, considering the precipitation input, exhibits the best performance in capturing the reference ETa variation. The MET ETa (AETJUNG), 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 ETa (JRA55), due to overestimations precipitation and radiation. (3) The reference ETa showed decreased and then increased trend. ETa 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 ETa variability in the MYRB, but also significant for improving global ETa 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

INTRODUCTION

Actual evapotranspiration (ETa) is one of the important hydrological processes (Brutsaert 2005; Jung et al. 2010). ETa 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 ETa 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 (Moito & Tao 2015). Acceleration or intensification of the hydrological cycle has occurred with global warming, and ETa has also been affected (Brutsaert & Parlange 1998; Jung et al. 2010). However, the existing observations of ETa 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 ETa at regional or global scales are still lacking.

Different methods have been used to estimate regional ETa, 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
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; El Haj El Tahir et al. 2012; Gokmen et al. 2012). LSMs are excellent in capturing ET_a variation (Xue et al. 2015; 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 \( 3.0 \times 10^5 \) km², 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.
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. 2010) is approximately 1,500 mm, which is three to seven times more than the annual precipitation.

Datasets

Global \( \text{ET}_a \) products

Four different categories of \( \text{ET}_a \) products (Table 1) were evaluated in our study: (1) LSMs, (2) reanalysis, (3) MTE method, and (4) diagnostic models (DMs). Five \( \text{ET}_a \) 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. \( \text{AET}_{\text{Noah1}} \) and \( \text{AET}_{\text{Noah2}} \) came from the Noah model, \( \text{AET}_{\text{CLM}} \) from the community land model (CLM), \( \text{AET}_{\text{VIC}} \) from the variable infiltration capacity (VIC) and \( \text{AET}_{\text{MOS}} \) from the mosaic model (MOS) (Rodell et al. 2004; Ferreira et al. 2013). The \( \text{AET}_{\text{Noah2}} \) 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’ \( \text{ET}_a \) 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, \( \text{AET}_{\text{JRA55}} \) and \( \text{AET}_{\text{JRA25}} \), respectively) (Nakaegwa 2008). The main objective of \( \text{AET}_{\text{JRA55}} \) was to produce a comprehensive atmospheric dataset suitable for studying multi-decadal variability and climate change.
The Jung global land ETa 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. 2013). ETa from ZhangKe (http://www.ntsg.umt.edu/project/et; Zhang et al. 2010) 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 ETa 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. 2009), 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 ETa 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. 2008). 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 ETa (Xu 2001; Gao et al. 2007; Ferreira et al. 2013; Corbari et al. 2014; Long et al. 2014):

$$ET_a = P - R - \Delta S$$

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; \(\Delta S\) 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 \(\Delta S\) played a fundamental role in water, energy, and biogeochemical cycles at regional and global scales (Famiglietti 2015). 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 \(\Delta S\) 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 \(\Delta S\) 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 \(\Delta S\) 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. 2015).

**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:

\[
\text{BIAS} = \frac{\sum_{i=1}^{N} (X_i - Y_i)}{N} \quad (2)
\]

\[
\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (X_i - Y_i)^2}{N}} \quad (3)
\]

\[
\text{CORR} = \frac{\sum_{i=1}^{N} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{N} (X_i - \bar{X})^2 \sum_{i=1}^{N} (Y_i - \bar{Y})^2}} \quad (4)
\]

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–Kendall 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
(SM_{Noah2} and SM_{CLM}) showed similar variations without significant differences ($p > 0.1$). The soil moisture from VIC-IGSNRR (SM\text{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\text{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 ($\Delta S$) 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 2](image2.png)

**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\text{IGSNRR}), GLDAS MOS (SM\text{MOS}), GLDAS Noah (SM\text{Noah}), and GLDAS CLM (SM\text{CLM}).

![Figure 3](image3.png)

**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, \(\Delta S\) 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 (Hobbs 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\(_{\text{Noah}1}\) captured the mean value with the smallest BIAS and largest CORR (\(-0.2\) mm yr\(^{-1}\) and 0.63, respectively), followed by AET\(_{\text{MOS}}\), with a small overestimation (25.0 mm yr\(^{-1}\)) but with a better RMSE/CORR (51.8 mm yr\(^{-1}\)/0.62). AET\(_{\text{VIC}}\), similar to AET\(_{\text{MOS}}\), was approximately overestimated by 6%, and had a good relation with the reference ET\(a\) (CORR = 0.58). Although AET\(_{\text{CLM}}\) and AET\(_{\text{Noah2}}\) 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\(_{\text{CLM}}\) and 0.57 for AET\(_{\text{Noah2}}\)). In particular, note that AET\(_{\text{Noah2}}\) had a longer time series (since 1960) than the other LSMs (since 1979). What is more, the performance of AET\(_{\text{Noah2}}\) 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\(_{\text{Noah2}}\) was

![Figure 4](https://iwaponline.com/hr/article-pdf/48/2/498/365877/nh0480498.pdf)

**Figure 4** | Time series of yearly ET\(a\) from nine products and corresponding precipitation for the MYRB from 1960 to 2013.
recommended for application at the half century time span, and AETNoah1 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 AETJUNG. Therefore, AETJUNG was expected to have a higher capacity for capturing the reference ETa variation than other models; however, the performance of AETJUNG was moderate with underestimation of the reference ETa 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 (AETMODIS and AETZhangKe) both performed adequately, underestimating (19.5% and 18.9%, respectively) the ETa variation with relatively large BIAS/RMSEs (94.0/103.6 mm yr$^{-1}$ for AETMODIS and 86.1/106.1 mm yr$^{-1}$ for AETZhangKe). Although AETMODIS performed poorly with regard to BIAS/RMSE, it accurately reproduced the interannual variation of the reference ETa, with the highest CORR (0.66). Therefore, AETMODIS is recommended when higher spatial resolution is required. In contrast, interannual variation of AETZhangKe was poorly associated with the lowest CORR. However, AETJRA55 from the reanalysis model had poor performance, strongly overestimating the ETa 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 ETa variation, especially for the Noah model. AETJUNG had a good performance in BIAS and RMSE. The DMs (i.e., AETMODIS and AETZhangKe) did not perform well compared to the LSM and MTE models. However, AETJRA55 from the reanalysis model performed poorly in capturing the variation and amplitude of the reference ETa.

**Possible reasons of performance for ETa products**

There are several reasons for the annual differences in ETa, 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 ETa categories, the LSMs, especially the Noah and Mosaic models, captured the variability of the reference ETa 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 2 | Statistical summary of model performance in simulating ETa based on comparison with reference evapotranspiration

<table>
<thead>
<tr>
<th>Model</th>
<th>AET$_{\text{prod}}$ (reference ETa)</th>
<th>BIAS</th>
<th>RMSE</th>
<th>CORR</th>
</tr>
</thead>
<tbody>
<tr>
<td>AETNoah1</td>
<td>462 (462)</td>
<td>0.2</td>
<td>56.3</td>
<td>0.63</td>
</tr>
<tr>
<td>AETNoah2</td>
<td>437 (467)</td>
<td>54.5</td>
<td>58.9</td>
<td>0.57</td>
</tr>
<tr>
<td>AETCLM</td>
<td>390 (462)</td>
<td>72.5</td>
<td>86.2</td>
<td>0.59</td>
</tr>
<tr>
<td>AETMOS</td>
<td>487 (462)</td>
<td>25.0</td>
<td>51.8</td>
<td>0.62</td>
</tr>
<tr>
<td>AETVIC</td>
<td>489 (462)</td>
<td>27.4</td>
<td>56.8</td>
<td>0.58</td>
</tr>
<tr>
<td>AETJUNG</td>
<td>404 (459)</td>
<td>54.7</td>
<td>76.9</td>
<td>0.35</td>
</tr>
<tr>
<td>AETJRA55</td>
<td>587 (467)</td>
<td>120.6</td>
<td>135.0</td>
<td>0.31</td>
</tr>
<tr>
<td>AETMODIS</td>
<td>387 (481)</td>
<td>94.0</td>
<td>103.6</td>
<td>0.66</td>
</tr>
<tr>
<td>AETZhangKe</td>
<td>369 (455)</td>
<td>86.1</td>
<td>106.1</td>
<td>0.08</td>
</tr>
</tbody>
</table>

**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.

### Table 3 | Main forcing fields for the four categories of ETa models: AETNoah1, AETNoah2, AETCLM, and AETMOS included in GLDAS-LSMs, AETJUNG included in MTE, AETJRA55 included in reanalysis dataset, and AETMODIS, AETZhangKe included in DMs

<table>
<thead>
<tr>
<th>Models</th>
<th>Main forcing fields</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLDAS-LSMs</td>
<td>Precipitation, radiation, temperature, humidity, wind, air pressure, soil moisture, satellite</td>
</tr>
<tr>
<td>Eddy covariance method</td>
<td>Flux tower data, FAPAR, radiation, precipitation, temperature, humidity, potential evapotranspiration, land cover</td>
</tr>
<tr>
<td>Reanalysis dataset</td>
<td>Precipitation, radiation, microwave imagers, snow depths, temperature</td>
</tr>
<tr>
<td>DMs</td>
<td>Surface variable: LAI (for AETMODIS), NDVI (for AETZhangKe), FAPAR, enhanced vegetation index, surface albedo, and land cover</td>
</tr>
</tbody>
</table>
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’ ETa: the GLDAS-1, including AETMOS, AETCLM, AETNoah1, and AETVIC-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 AETNoah2 in capturing the reference ETa 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 ETa (Figure 4). This also indicated that the ETa 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 ETa. Thus, the ETa from Noah models (AETNoah1 and AETNoah2), especially for the advanced Noah2 employing consistent forcing data, showed excellent performance in capturing the annual variation of reference ETa.

AETJRA55 was from the Japanese 55-year reanalysis (JRA-55) dataset of the Japan Meteorological Agency (Kobayashi et al. 2015). Although AETJRA55 had adopted a series of strategies to improve the capacity for simulating ETa, 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) (Nakaegawa 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](https://iwaponline.com/hr/article-pdf/48/2/498/365877/nh0480498.pdf) | 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 AETJRA55 and relatively low performance in capturing the variation of the reference ETa. Therefore, great caution should be paid to hydrological variables before using them in reanalysis.

The AETJUNG 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 ETa. In addition, the flux towers were sparse throughout our study region (Jung et al. 2010, Supplementary Figure 1), which may introduce some uncertainties to AETJUNG in capturing the reference ETa variation. Both of the ETa from DMs were underestimated. The AETZhangKe did not consider the evaporation from canopy interception (Zhang et al. 2010), which may be a reason for the difference between the reference ETa and AETZhangKe. The Chinese government adopted a series of strategies to restore the ecosystem since 1999 (Zhang et al. 2013; FAO 2013) 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, ETa could be underestimated. In addition, insufficient accuracy of remotely sensed radiation employed in the AETZhangKe models may contribute to the uncertainty of ETa (Zhang et al. 2012). Although AETMODIS had a limited time span, it showed better performance in capturing the reference ETa variability, which may result from the improvement of the ETa algorithm (Mu et al. 2007, 2011). In addition, the forcing data in the AETMODIS 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 AETMODIS.

### DISCUSSION

#### Comparison of the relationship between precipitation and ETa 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 ETa and considered as the main input parameter in models. Especially in semi-arid and semi-humid regions where water supply limited ETa, the degree of ETa model dependence on precipitation could reflect its performance in estimating ETa (Zhang et al. 2012). The CORR of precipitation and the nine ETa products are listed in Table 4.

All of the ETa 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 1999; 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’ ETa followed the variation of precipitation and responded to changes in soil moisture in the MYRB.

The EC-based AETJUNG and reanalysis AETJRA55 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...
by combining the observation and forecast model results and had numerous uncertainties (Bosilovich et al. 2011). AETJUNG 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 AETMODIS (CORR = 0.62, $P < 0.05$) and low correlation for AETZhangKe (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. 2013; Zhang et al. 2014). Even without precipitation in AETMODIS, there was a general agreement, which may be a result of a series of improvements in the terrestrial evapotranspiration algorithm (Mu et al. 2013, 2014), 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 5(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)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Trend</td>
<td>Mean</td>
</tr>
<tr>
<td>Reference ET$_a$</td>
<td>455.5</td>
<td>−4.54</td>
<td>474.6</td>
</tr>
<tr>
<td>AETNoah1</td>
<td>451.5</td>
<td>−1.86</td>
<td>469.4</td>
</tr>
<tr>
<td>AETNoah2</td>
<td>440.2</td>
<td>−1.20</td>
<td>435.8</td>
</tr>
<tr>
<td>AETCLM</td>
<td>392.3</td>
<td>1.43</td>
<td>384.6</td>
</tr>
<tr>
<td>AETMODIS</td>
<td>490.3</td>
<td>−3.23</td>
<td>475.6</td>
</tr>
<tr>
<td>AETVIC</td>
<td>495.9</td>
<td><strong>3.88</strong>*</td>
<td>469.2</td>
</tr>
<tr>
<td>AETRA55</td>
<td>602.5</td>
<td>1.29</td>
<td>578.3</td>
</tr>
<tr>
<td>AETJUNG</td>
<td>399.6</td>
<td>0.21</td>
<td>411.4</td>
</tr>
<tr>
<td>AETZhangKe</td>
<td>373.8</td>
<td>0.25</td>
<td>359.3</td>
</tr>
<tr>
<td>AETMODIS</td>
<td>−</td>
<td>−</td>
<td>387.0</td>
</tr>
</tbody>
</table>

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 2015), 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. 2015; Yang et al. 2015a), including quantitatively analyzing the relative contributions of different factors (e.g., irrigation, water diversion, global warming, land use change, and CO2 enrichment) to ET_a. The mechanisms influencing the spatio-temporal 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 ETa from Noah (especially for Noah1) and MOS demonstrated the best performance. AETNoah2 is recommended for application in large regions and at a half-century scale, whereas AETNoah1 and AETMOS are recommended for the past three decades and AETMODIS 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 ETa models can significantly improve the performance in capturing the variation of the reference ETa. The AETJUNG model, which was based on eddy covariance, performed relatively well, with a small negative BIAS and a little weak relationship to the reference ETa, which demonstrated its dependence on observation data and independence from precipitation. The two DMs demonstrated fair performance with small underestimation. The AETMODIS model, although a short time span, showed a stronger capacity for capturing the reference ETa than the other models. However, the AETRASS model performed poorly, with the largest positive BIAS and RMSE; this was caused by higher precipitation and radiation input.

The variation of the reference ETa shifted around 1999. Further analysis found that the reference ETa 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 ETa from the Noah model, especially Noah1, better captured the ETa trend and amplitude for the three periods. The AETVIC model best captured the ETa amplitude for the pre-1999 period, while the AETJUNG model was best for the post-1999 and 1983–2013 periods. The underlying reasons for ETa changes may be the implementation of several forest strategies. Comprehensive evaluation and comparison of various ETa products was valuable for the application of ETa products for regional water resource management and guiding future research on ETa estimation. Limited by the coarse temporal and spatial resolution of ETa products in this study, more accurate regional ETa models with higher temporal and spatial resolution forcing data would be necessary to improve the estimation performance of ETa in further work.

ACKNOWLEDGEMENTS

This research was financially supported by the Natural Science Foundation of China (No. 41330529, 41501032, 41571024, and 41501496). 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


IPCC 2007 *Summary for Policymakers: The Physical Science Basis, Contribution of Working Group I to the IPCC Fifth Assessment Report Climate Change*.


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