

## Using Bayesian model averaging to estimate terrestrial evapotranspiration in China



Yang Chen<sup>a</sup>, Wenping Yuan<sup>a,\*</sup>, Jiangzhou Xia<sup>a</sup>, Joshua B. Fisher<sup>b</sup>, Wenjie Dong<sup>a</sup>, Xiaotong Zhang<sup>c</sup>, Shunlin Liang<sup>c,d</sup>, Aizhong Ye<sup>e</sup>, Wenwen Cai<sup>a</sup>, Jinming Feng<sup>f</sup>

<sup>a</sup> State Key Laboratory of Earth Surface Processes and Resource Ecology, Future Earth Research Institute, Beijing Normal University, Beijing/Zhuhai 100875/519087, China

<sup>b</sup> Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA

<sup>c</sup> State Key Laboratory of Remote Sensing Science, School of Geography, Beijing Normal University, Beijing 100875, China

<sup>d</sup> Department of Geographical Sciences, University of Maryland, College Park, MD 20742, USA

<sup>e</sup> College of Global Change and Earth System Science, Beijing Normal University, Beijing 100875, China

<sup>f</sup> Key Laboratory of Regional Climate-Environment Research for Temperate East Asia (RCE-TEA), Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

### ARTICLE INFO

#### Article history:

Received 5 February 2015

Received in revised form 15 May 2015

Accepted 27 June 2015

Available online 2 July 2015

This manuscript was handled by Konstantine P. Georgakakos, Editor-in-Chief, with the assistance of Hamid Moradkhani, Associate Editor

#### Keywords:

Evapotranspiration  
Bayesian model averaging  
Water balance  
Remote sensing  
China

### SUMMARY

Evapotranspiration (ET) is critical to terrestrial ecosystems as it links the water, carbon, and surface energy exchanges. Numerous ET models were developed for the ET estimations, but there are large model uncertainties. In this study, a Bayesian Model Averaging (BMA) method was used to merge eight satellite-based models, including five empirical and three process-based models, for improving the accuracy of ET estimates. At twenty-three eddy covariance flux towers, we examined the model performance on all possible combinations of eight models and found that an ensemble with four models (BMA\_Best) showed the best model performance. The BMA\_Best method can outperform the best of eight models, and the Kling–Gupta efficiency (KGE) value increased by 4% compared with the model with the highest KGE, and decreased RMSE by 4%. Although the correlation coefficient of BMA\_Best is less than the best single model, the bias of BMA\_Best is the smallest compared with the eight models. Moreover, based on the water balance principle over the river basin scale, the validation indicated the BMA\_Best estimates can explain 86% variations. In general, the results showed BMA estimates will be very useful for future studies to characterize the regional water availability over long-time series.

© 2015 Elsevier B.V. All rights reserved.

### 1. Introduction

Evapotranspiration (ET) is one of the most important variables of terrestrial ecosystems as it links water, carbon, and surface energy exchanges. Therefore, accurate estimations of ET in large scale is crucial for understanding the interactions between land surfaces and the atmosphere (Keane et al., 2002), drought and land resource management (Raupach, 2001), and coupling water cycling and ecosystem carbon exchange (Eamus, 2003). Over the past several years, there have been substantial efforts to retrieve ET over large areas. Zeng et al. (2012) estimated global ET with a spatial regression model by integrating precipitation, temperature and satellite-derived normalized difference vegetation index (NDVI) data. Xia et al. (2014) calculated ET over grassland ecosystems of dryland East Asia using regression tree method. Shu et al. (2011) estimated the regional ET over the North China Plain using the data

from Chinese geostationary satellite Fengyun-2C and found spatial variations of ET compare very well to the land use types. However, ET is still the component with the most problem in the water cycle processes because of the complex controlling factors and heterogeneity of the landscape (Lettenmaier and Famiglietti, 2006; Yuan et al., 2010a).

Numerous models are developed for quantifying spatiotemporal variations of ET using remote sensing observations (Cleugh et al., 2007; Mu et al., 2007; Fisher et al., 2008; Leuning et al., 2008; Jung et al., 2009; Yuan et al., 2010b; Zhang et al., 2010; Mu et al., 2011; Vinukollu et al., 2011a; Yang et al., 2012; Baik and Choi, 2015; French et al., 2015; Liu et al., 2015; Tang and Li, 2015). Satellite-based modeling has been an important tool for accurately parameterizing surface biophysical variables because remotely sensed data provide temporally and spatially continuous information over heterogeneous surfaces. In previous studies, the net radiation products, remotely sensed variables (e.g., vegetation index) and meteorological measurements (e.g., vapor pressure and air temperature) were used to calculate the special

\* Corresponding author. Tel.: +86 10 58807715.

E-mail address: [yuanwpcn@126.com](mailto:yuanwpcn@126.com) (W. Yuan).

evapotranspiration. For example, the global land ET was estimated by Vinukollu et al. (2011) using a set of remote sensing and observational based radiation and meteorological forcing datasets such as International Satellite Cloud Climatology Project (ISCCP), Advanced Very High Resolution Radiometer (AVHRR) and Global Meteorological Forcing Data set from Princeton University (PU).

However, there are large model uncertainties revealed by the inter-comparisons of ET estimates (Vinukollu et al., 2011b). For example, Jiménez et al. (2011) indicated that the global annual mean ET between different models and datasets had 50% uncertainties, which induced large uncertainties for the global water and energy cycles. The mean annual ET in China calculated by different models ranged from 535 to 852 mm/year. The major reason for the different models estimations were the differences in model structures and their dominant variables (Chen et al., 2014).

The multi-model ensembles method has increasingly been used to improve model estimations (Hagedorn et al., 2005). The Bayesian Model Averaging (BMA) method, a statistical scheme based on multi-model ensemble, was originally developed as a way to combine different models or forecasts (Hoeting et al., 1999). The contribution of each individual model in the BMA method is weighted by its posterior weight of evidence (Ellison, 2004). BMA has been widely used to study the climate change (Duan and Phillips, 2010), improve the predictions accuracy of hydrology (Duan et al., 2007), weather (Raftery et al., 2005; Wu et al., 2012), forest biomass (Li et al., 2008) and economics (Fernandez et al., 2001). Previous studies indicated better estimations of BMA than other multi-model ensemble methods (Viallefont et al., 2001; Ellison, 2004; Raftery et al., 2005; Slougher et al., 2007). For example, Wang et al. (2012) merged seasonal rainfall forecasts from multiple models using BMA and improved effectively skills of the models. Similarly, BMA method also was used to merge estimates of hydrological flows from multi-model and BMA ensembles decreased estimates bias value and increase correlation coefficient compared with the single best model (Jiang et al., 2012). Moreover, BMA method can quantify the uncertainties from the inputs, model structure and parameters and improve the model accuracy (Najafi et al., 2011). For example,

Najafi et al. (2011) used BMA to merge the hydrologic models variance and quantify the uncertainties, which were useful in evaluating the regional water resources.

This study uses BMA method to improve China terrestrial ET estimates based on eight ET models. The objectives of this study are to: (1) use the BMA method to improve the accuracy of ET estimates; (2) compare two ensemble strategies: ensemble with all models and ensemble with the selected models; (3) examine the performance of the BMA method through a water balance analysis; and (4) analyze the spatiotemporal patterns of ET calculated by the BMA method over China from 1982 to 2009.

## 2. Data

### 2.1. Data at eddy covariance (EC) site

Twenty-three EC sites (Fig. 1, Table 1) were used to examine model performance. The data were collected from Arid/Semi-arid experimental observation synergy and integration, ChinaFlux, AsiaFLUX and LathuileFLUX. The sites included seven major biomes, evergreen needleleaf forests, evergreen broadleaf forests, deciduous needleleaf forests, deciduous broadleaf forests, mixed forests, grasslands and croplands. The eight ET models are driven by 8-day net radiation ( $R_n$ ), solar radiation ( $R_g$ ), relative humidity ( $R_h$ ), air temperature ( $T_a$ ), maximum air temperature ( $T_{max}$ ), atmospheric pressure (Pr), wind speed ( $W_s$ ), vapor pressure deficit (VPD) and Minimum air temperature ( $T_{min}$ ) (see Table 2).

It has been recognized that the sum of latent heat (LE) and sensible heat ( $H$ ) as measured in EC towers is generally less than the available energy (Foken, 2008). LE observations can be corrected with the following formulas (Jung et al., 2010),

$$LE_{cor} = (R_n - G) / (H_{uncor} + LE_{uncor}) \times LE_{uncor} \quad (1)$$

where  $R_n$  is the net radiation,  $G$  is the soil heat flux,  $H_{uncor}$  is uncorrected sensible heat,  $LE_{uncor}$  is uncorrected latent heat and  $LE_{cor}$  is corrected latent heat.

The leaf area index (LAI) and normalized difference vegetation index (NDVI) for the eddy covariance towers were from

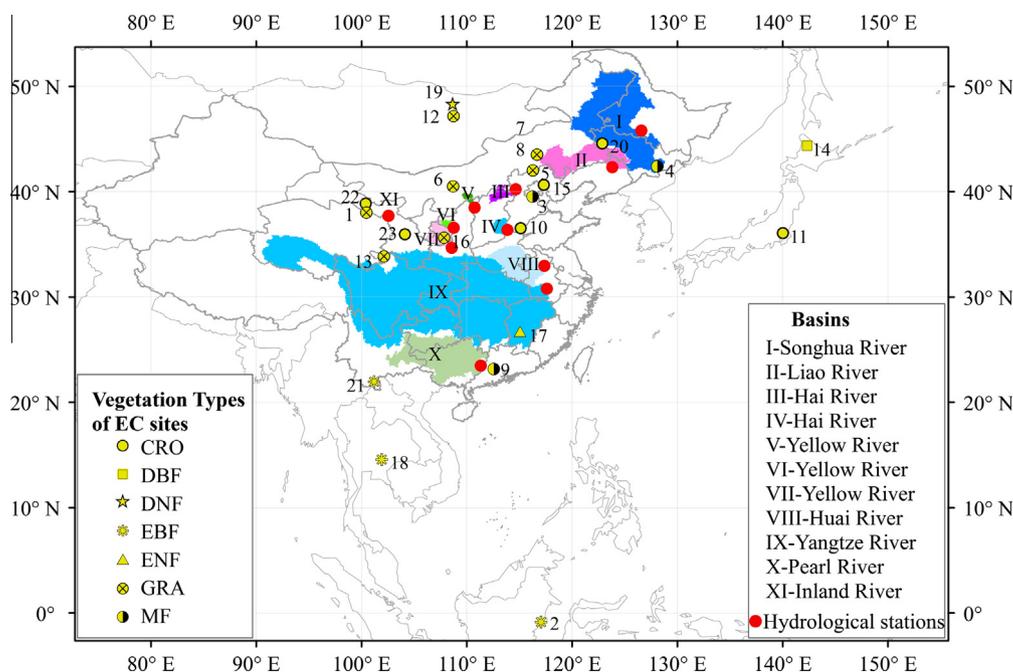


Fig. 1. Locations of the twenty-three EC sites and eleven hydrological stations. Numbers match with the sites ID in Tables 1 and 3. DBF: deciduous broadleaf forests; DNF: deciduous needleleaf forests; EBF: evergreen broadleaf forests; ENF: evergreen needleleaf forests; CRO: croplands; GRA: grasslands; MF: mixed forests.

**Table 1**

Name, location and vegetation types of the 23 EC sites.

ID	Site	Lat (°N)	Long (°E)	Ele (m)	Veg	Tann (°C)	Pann (mm)	Obs periods
1	CN-Aro	38.04	100.46	3033	GRA	2.12	314.90	2008–2009
2	ID-Bks	−0.86	117.05	20	EBF	25.59	2576.27	2002
3	CN-Bed	39.53	116.25	30	MF	11.69	531.85	2005–2006
4	CN-Cha	42.40	128.10	761	MF	3.38	811.35	2003
5	CN-Du2	42.05	116.28	1350	GRA	3.07	425.54	2006
6	CN-Ku1	40.54	108.69	1020	GRA	7.96	277.08	2006
7	CN-Xi1	43.55	116.68	1250	GRA	1.86	378.49	2006
8	CN-Xi2	43.55	116.67	1250	GRA	1.86	378.49	2006
9	CN-Din	23.17	112.57	364	MF	20.55	1829.12	2002–2007
10	CN-Gta	36.52	115.13	1999	CRO	14.09	613.14	2009
11	JP-Mas	36.05	140.03	12	CRO	10.86	1544.03	2002–2003
12	MN-Kbu	47.21	108.74	1235	GRA	0.20	239.70	2003–2008
13	CN-Mqu	33.89	102.14	3423	GRA	5.83	660.44	2009
14	JP-Mbf	44.38	142.32	585	DBF	7.80	755.79	2004–2005
15	CN-Myu	40.63	117.32	350	CRO	9.60	531.74	2008–2009
16	CN-Qya	35.66	107.84	1136	GRA	8.98	537.44	2009
17	CN-Qia	26.74	115.06	79	ENF	18.73	1771.52	2003–2004, 2006–2007
18	TL-Skr	14.57	101.92	543	EBF	25.11	2477.54	2001–2003
19	MN-Skt	48.35	108.65	1630	DNF	−3.10	445.32	2003–2006
20	CN-Tyc	44.57	122.88	184	CRO	6.18	495.86	2009
21	CN-Xsh	21.96	101.20	756	EBF	21.10	1424.37	2004–2007
22	CN-Yke	38.86	100.41	1519	CRO	11.61	254.66	2008–2009
23	CN-Yzh	35.95	104.13	1966	CRO	10.94	422.45	2008

Lat: latitude; Long: longitude; Ele: elevation; Veg: vegetation type; Tann: annual mean temperature; Pann: annual mean precipitation; Obs periods: observation periods. DBF, deciduous broadleaf forests; MF, mixed forests; ENF, evergreen needleleaf forests; EBF, evergreen broadleaf forests; DNF, deciduous needleleaf forests; GRA, grasslands; CRO, croplands. Positive values represent north latitude and east longitude. Negative values represent south latitude and west longitude.

**Table 2**

Summary of the eight ET models and forcing variables.

ID	ET algorithm	Forcing inputs	Outputs	References
1	Artificial neural network (ANN)	$R_n, T_a, R_h, NDVI$	ET	Chen et al., 2014
2	Regression tree (RT)	$R_n, T_a, R_h, NDVI$	ET	Xia et al., 2014
3	Support vector model (SVM)	$R_n, T_a, R_h, NDVI$	ET	Chen et al., 2014
4	Empirical ET algorithm (Reg1)	$R_n, T_a, NDVI$	ET	Wang et al., 2007
5	Semi-empirical ET algorithm (Reg2)	$R_g, T_a, R_h, VPD, W_s, NDVI$	$ET_A, ET_E$	Wang et al., 2010
6	Revised Penman–Monteith model (RRS-PM)	$R_n, T_a, R_h, VPD, LAI, Pr$	$ET_s, ET_c$	Yuan et al., 2010b
7	MODIS ET algorithm (PM-MOD16)	$R_n, T_a, R_h, VPD, LAI, T_{min}$	$ET_s, ET_c, ET_i$	Mu et al., 2011
8	Bio-meteorological approach based on Priestley–Taylor equation (PT-JPL)	$R_n, T_a, T_{max}, VPD, NDVI$	$ET_s, ET_c, ET_i$	Fisher et al., 2008

ET, total evapotranspiration;  $ET_A$ , the atmospheric control component;  $ET_E$ , the energy control component;  $ET_s$ , soil evaporation;  $ET_c$ , canopy transpiration;  $ET_i$ , interception evaporation.

Forcing variables are  $R_n$ : net radiation;  $T_a$ : air temperature;  $T_{min}$ : minimum air temperature;  $T_{max}$ : maximum air temperature;  $R_h$ : relative humidity; NDVI: normalized difference vegetation index;  $R_g$ : solar radiation; VPD: vapor pressure deficit;  $W_s$ : wind speed; LAI: leaf area index; Pr: atmospheric pressure.

Moderate Resolution Imaging Spectroradiometer (MODIS). MODIS ASCII data generated from MODIS Collection 5 data (<http://daac.ornl.gov/MODIS/>) was used in this study. The resolution of the MODIS LAI (MOD15A2) and MODIS NDVI (MOD13A2) is  $1 \times 1$  km. Only the LAI and NDVI values of the pixels containing the towers were used. Quality control (QC) flags, which contain cloud contamination for pixels, were examined reject the data with poor quality. Finally MODIS LAI and NDVI data with the interval of eight days for each flux site were produced.

The runoff measurements at eleven river basins used in the water balance approach were collected from the Hydrological Year Book (Fig. 1, Table 3).

## 2.2. Data at regional scale

To estimate regional ET,  $R_g, R_n, T_a, R_h, W_s, VPD, Pr, T_{max}$  and  $T_{min}$  datasets from the Modern Era Retrospective–Analysis for Research and Applications (MERRA, [Global Modeling and Assimilation Office, 2004](http://globalmodelingandassimilationoffice.nasa.gov)) were used. MERRA is the National Aeronautics and Space Administration (NASA) reanalysis for the satellite era that uses the new version of the Goddard Earth Observing System Data Assimilation System Version 5 (GEOS-5). Details about the

MERRA dataset are available at the NASA website (<http://gmao.gsfc.nasa.gov/research/merra>). The precipitation ( $P$ ) data were obtained from the National Climate Center of China Meteorological Administration. We used the thin-plate spline (Hutchinson 1995, 2004) on the 600 meteorological sites to get the precipitation over China at a spatial resolution of  $0.1^\circ$  latitude and longitude for the period, 1982–2009 (Yuan et al., 2014).

Satellite-based vegetation index (NDVI) was used to indicate the vegetation conditions in the six ET models (i.e. ANN, RT, SVM, Reg1 and Reg2) and calculate the energy partition (PT-JPL). Satellite-based LAI was used to calculate the canopy resistance in the RRS-PM and PM-MOD16. This study used the MODIS LAI and NDVI datasets and Advanced Very High Resolution Radiometer (AVHRR) to calculate the combined LAI and NDVI from 1982 to 2009. The spatial resolution of AVHRR Global Inventory Modeling and Mapping Studies (GIMMS) NDVI is based on a composite of monthly maximum values of biweekly data with a  $0.0727^\circ$  spatial resolution and covers the period from 1982 to 2006. The monthly NDVI data aggregated from the 16-day MODIS NDVI (MOD13A2) data at 1-km spatial resolution from 2000 to 2009 was used in this study. The QC flags were used to judge the quality of the NDVI data and then rejected the poor quality NDVI data. To be consistent with

**Table 3**  
Hydrological stations and river basins used in this study.

ID	Site name	River basin	Lat (°N)	Long (°E)	Ele (m)	Area (km <sup>2</sup> )	Tann (°C)	Pann (mm)	Obs periods
I	Ha erbin	Songhua River	45.79	126.58	118	379,079	15.13	1012.55	1982–2004
II	Tie ling	Liao River	42.32	123.83	59	129,421	12.29	1131.26	1982–2004
III	Shi zhali	Hai River	40.21	114.66	805	22,487	8.06	579.14	1982–2004
IV	Guan tai	Hai River	36.36	113.87	648	16,797	−2.01	564.11	1982–2004
V	Zhang jiashan	Yellow River	34.66	108.58	360	43,173	7.43	472.33	1982–2009
VI	Wen jiachuan	Yellow River	38.48	110.75	760	8476	5.59	464.89	1982–2009
VII	Liu jiahe	Yellow River	36.56	108.76	1153	7429	5.58	518.58	1982–2009
VIII	Beng bu	Huai River	32.95	117.37	11	121,330	6.84	400.26	1999–2008
IX	Da tong	Yangtze River	30.77	117.62	1	1,705,383	6.86	1518.82	2001–2006
X	Wu zhou	Pearl River	23.47	111.33	64	327,006	0.91	355.37	1982–2004
XI	Za musi	Inland River	37.7	102.57	2248	851	8.65	532.73	1982–2005

Lat: latitude; Long: longitude; Ele: elevation; Area: watershed area; Tann: annual mean temperature; Pann: annual mean precipitation; Obs periods: observation periods.

the spatial resolution of the AVHRR NDVI data, MODIS NDVI data were first spatially aggregated to a resolution of approximately 0.0727°. The procedures following the method of Zhang et al. (2008) were used to combine the two series: (1) a simple linear regression was used to regress the monthly MODIS NDVI on corresponding AVHRR NDVI from 2000 to 2006; (2) use the regression equations in step one to adjust the AVHRR NDVI time series and then calculate an integrated AVHRR–MODIS NDVI monthly time series from 1982 to 2009 (Chen et al., 2014). 8-day MODIS LAI (MOD15A2) (Myneni et al., 2002) and monthly AVHRR LAI (Myneni et al., 1997) were used with the same procedures as those of NDVI to generate a continuous LAI dataset from 1982 to 2009.

### 3. Methods

#### 3.1. Evapotranspiration models

Five empirical ET models and three process-based models were used in this study. These eight models were the representatives of current various ET estimate methods and they have been calibrated and validated at regional even global scales. The five empirical models include an artificial neural network (ANN) model, a regression tree (RT) model, a support vector model (SVM) and two models (Reg1 and Reg2) developed by Wang et al. (2007) and Wang et al. (2010). All measurements needed to be split to two datasets for model calibration and validation (Irmak et al., 2003; Tabari et al., 2013; Xu and Singh, 2000; Valipour, 2014a, 2014b, 2015). Half of the measurements from all sites were selected randomly to train the three machine learning methods (ANN, SVM and RT), and validate with the other half of the measurements. The three process-based models are a revised Penman–Monteith model (RRS-PM, Yuan et al., 2010b), MODIS evapotranspiration algorithm (PM-MOD16, Mu et al., 2011) and a bio-meteorological approach developed from Priestley–Taylor equation (PT-JPL, Fisher et al., 2008).

ANN is a machine learning method. It is a modeling tool to solve not only the linear but also the non-linear multivariate regression problems. It has three layers: input layer, hidden layer, and output layer. The ANN adjusts the weight of internal nodes with training data. The back propagation artificial neural and multilayer perceptron (Rumelhart et al., 1986) were used in this study. ANN can represent any arbitrary nonlinear function given sufficient complexity of the trained network, and it can find relationships between different input samples. Most importantly, the ANN is able to generalize a relationship from small subsets of the data while remaining relatively robust in the presence of noisy or missing inputs, and can adapt or learn in response to changing environments (Dawson and Wilby, 1998). There are some of the successful applications of back propagation artificial neural and multilayer perceptron in hydrology (Landeras et al., 2008; Valipour et al., 2012, 2013; Li et al., 2014b).

RT algorithm is a machine-learning method for constructing prediction models. It recursively partitions a dataset into more homogeneous subsets to predict class membership (Xiao et al., 2010). The RT model can handle the non-linear relationship between predictive and target variables. It also can allow both continuous and discrete variables. A commercial software called Cubist which implements a modified regression tree algorithm was used in this study. It is proved to be a very effective approach for producing rule-based models (Xia et al., 2014).

SVM represents a useful technique for nonlinear classification, regression and outlier detection. It is based on the statistical learning theory and to produce a model which predicts the target values of the test data. It can generate rules through a training process with the training data. The characteristic of the SVM is that it can change nonlinear regression to linear regressions (Vapnik et al., 1998). It is widely used to approximate regressions due to its ability to approximate any nonlinear functions, especially when samples are limited.

The Reg1 model estimates ET using a vegetation index, air temperature and surface net radiation as the dominant variables:

$$ET = R_n \cdot (a_0 + a_1 \cdot NDVI + a_2 \cdot T_a) \quad (2)$$

where  $R_n$  is net radiation,  $T_a$  is air temperature and NDVI is normalized difference vegetation index. It expresses the dependence of ET variations on the vegetation in the simplest form that is consistent with the Priestley–Taylor equation while incorporation the influence on vegetation control on ET. Although it uses the simplest form to express the influence of the vegetation, it can predict ET through different surface land cover types and soil moisture contents (Wang et al., 2007).

The Reg2 algorithm partitions total ET into two components, the atmospheric control component ( $ET_A$ ) and then energy control component ( $ET_E$ , Wang et al., 2010):

$$ET_E = \frac{\Delta}{\Delta + \gamma} \cdot R_g \cdot [a_1 + a_2 \cdot NDVI + (1 - R_h) \cdot (a_3 + a_4 \cdot NDVI)] \quad (3)$$

$$ET_A = \frac{\gamma}{\Delta + \gamma} \cdot W_s \cdot VPD \cdot [a_5 + (1 - R_h) \cdot (a_6 + a_7 \cdot NDVI)] \quad (4)$$

$$ET = a_8 \cdot (ET_E + ET_A) + a_9 \cdot (ET_E + ET_A)^2 \quad (5)$$

where the  $\Delta$  is slope of saturation-to-vapor pressure curve,  $\gamma$  is the psychrometric constant,  $R_g$  is the solar radiation, NDVI is normalized difference vegetation index,  $W_s$  is wind speed, VPD is water vapor pressure deficit, and  $R_h$  is relative humidity. This method was developed based on a Penman-based equation. The empirical coefficients were added to include the function of vegetation and soil moisture. The vegetation is diagnosed by NDVI term and soil moisture is diagnosed by  $(1 - R_h)$  term (Wang et al., 2010). This method can estimate ET over different climate conditions.

Although this method is simple, it can investigate long-term ET variation in regional ET over the land.

The RRS-PM algorithm is a Penman–Monteith type model modified by Yuan et al. (2010b). In the RRS-PM model, Beer–Lambert law was used to partition net radiation between the soil surface and the canopy (Ruimy et al., 1999):

$$R_{ns} = R_n \cdot \exp(-k \cdot LAI) \quad (6)$$

$$R_{nc} = R_n - R_{ns} \quad (7)$$

where  $R_n$  is net radiation,  $R_{ns}$  is net radiation to the soil,  $R_{nc}$  is net radiation to the canopy, LAI is leaf area index, and  $k$  is extinction coefficient (0.5). The temperature constraint controlling the stomatal conductance uses the equation expressed by June et al. (2004) and in PT-JPL from Fisher et al. (2008):

$$m_T = \exp\left(-\left(\frac{T_a - T_{opt}}{T_{opt}}\right)^2\right) \quad (8)$$

where  $T_a$  is air temperature,  $T_{opt}$  is optimum temperature. This method was proved to improve the ET estimates effectively than the original RS-PM algorithm (Yuan et al., 2010b; Yuan et al., 2012).

The PM-MOD16 algorithm developed by Mu et al. (2011) is another Penman–Monteith type model. The total ET is sum of soil evaporation ( $ET_s$ ), canopy transpiration ( $ET_c$ ), and interception evaporation ( $ET_i$ ). It was developed based on the Penman–Monteith equation (Monteith, 1965) after adapting by Cleugh et al. (2007):

$$LE = \frac{\Delta R_n + \rho C_p (e_{sat} - e)/r_a}{\Delta + \gamma(1 + r_s/r_a)} \quad (9)$$

where LE is the latent heat flux,  $\rho$  is air density,  $C_p$  is the specific heat capacity of air,  $r_a$  is the aerodynamic resistance,  $r_s$  is the surface resistance. Mu et al. (2011) modified their ET algorithm published in 2007 mainly by simplifying the vegetation cover fraction estimations; calculating total ET as the sum of daytime ET and nighttime ET; adding the calculation of soil heat flux; improving estimates of aerodynamic resistance, boundary layer resistance and stomatal conductance; and separating the dry canopy from the wet one and dividing the soil into moist ( $ET_{s\_moist}$ ) and saturated wet ( $ET_{s\_wet}$ ) surfaces (Mu et al., 2011):

$$ET_c = \frac{(\Delta \cdot R_{nc} + \rho \cdot C_p \cdot (e_{sat} - e)/r_a) \cdot (1 - F_{wet})}{\Delta + \gamma(1 + r_s/r_a)} \quad (10)$$

$$ET_{s\_moist} = \frac{(\Delta \cdot R_{ns} + \rho \cdot C_p \cdot (1 - F_c) \cdot (e_{sat} - e)/r_{as}) \cdot (1 - F_{wet})}{\Delta + \gamma \cdot r_{tot}/r_{as}} \cdot (Rh)^{VPD/\beta} \quad (11)$$

$$ET_{s\_wet} = \frac{(\Delta \cdot R_{ns} + \rho \cdot C_p \cdot (1 - F_c) \cdot (e_{sat} - e)/r_{as}) \cdot F_{wet}}{\Delta + \gamma \cdot r_{tot}/r_{as}} \quad (12)$$

$$ET_s = ET_{s\_moist} + ET_{s\_wet} \quad (13)$$

$$ET_i = \frac{(\Delta \cdot R_{nc} + \rho \cdot C_p \cdot (e_{sat} - e)/r_{hrc}) \cdot F_{wet}}{\Delta + \frac{P_a C_p r_{vrc}}{z \cdot v \cdot r_{hrc}}} \quad (14)$$

$$ET = ET_c + ET_s + ET_i \quad (15)$$

where  $F_c$  is vegetation cover fraction,  $r_{as}$  is the aerodynamic resistance at the soil surface,  $r_{tot}$  is the total aerodynamic resistance to vapor transport,  $r_{hrc}$  is the aerodynamic resistance on the wet canopy surface, and  $r_{vrc}$  is the wet canopy resistance.

The PT-JPL algorithm was developed by Fisher et al. (2008) and was based on the Priestley and Taylor (1972) equation. It contains dynamic coefficients estimated from vegetation indices and atmospheric moisture to estimate actual ET. The total ET is partitioned

into soil evaporation, canopy transpiration, and interception evaporation. Every component is calculated based on the Priestley–Taylor equation by adding the corresponding ecophysiological constraints:

$$ET = ET_s + ET_c + ET_i \quad (16)$$

$$ET_s = (f_{wet} + f_{SM}(1 - f_{wet})) \alpha \frac{\Delta}{\Delta + \gamma} (R_{ns} - G) \quad (17)$$

$$ET_c = (1 - f_{wet}) f_g f_T f_M \alpha \frac{\Delta}{\Delta + \gamma} (R_{nc} - G) \quad (18)$$

$$ET_i = f_{wet} \alpha \frac{\Delta}{\Delta + \gamma} R_{nc} \quad (19)$$

where  $f_{wet}$  is relative surface wetness,  $f_{SM}$  is soil moisture constraint,  $f_g$  is green canopy fraction,  $f_T$  is plant temperature constraint,  $f_M$  is plant moisture constraint.

### 3.2. Bayesian Model Averaging (BMA)

BMA method was used in this study to combine single ET models to estimate terrestrial ET. The BMA method thinks about a dependent variable  $y$ , the training data  $y_T$ , and the ensemble of all members' predictions  $X\{x_1, x_2, \dots, x_K\}$ . In this study,  $y$  refers to the ensemble ET and  $K$  indicates the number of ET models. Based on the law of total probability, the probability density function (PDF) can be expressed as

$$p(y|x_1, x_2, \dots, x_K) = \sum_{k=1}^K p(y|x_k) \cdot p(x_k|y_T) \quad (20)$$

where  $p(y|x_k)$  is the predictive PDF given by the simulation of  $x_k$  alone and  $p(x_k|y_T)$  is the posterior probability of the model prediction  $x_k$ .  $y_T$  is the target data. Identifying  $p(x_k|y_T)$  is a fractional statistical weight  $w_k$ . The magnitude of the weight reflects how well  $x_k$  matches  $y_T$ , and  $\sum_{k=1}^K w_k = 1$ . Eq. (20) can be expressed as

$$p(y|x_1, x_2, \dots, x_K) = \sum_{k=1}^K p(y|x_k) \cdot w_k \quad (21)$$

Before the BMA method is used, it is reasonable to assume that  $p(y|x_k)$  is a Gaussian distribution defined by a mean  $x_k$  and a variance  $\omega_k^2$  (Duan and Phillips, 2010; Raftery et al., 2005).

$$p(y|x_k) = g(y|\theta_k) \quad (22)$$

where  $g$  refers to Gaussian distribution and  $\theta_k = \{x_k, \omega_k^2\}$  donates parameter vector.

Combining Eqs. (21) and (22), the PDF of the BMA probabilistic prediction of  $y$  can be expressed as:

$$p(y|x_1, x_2, \dots, x_K) = \sum_{k=1}^K g(y|\theta_k) \cdot w_k \quad (23)$$

The log-likelihood function  $l$  was used to obtain both Bayesian weights  $w_k$  and parameter vectors  $\theta_k$  because it is more convenient to compute than the likelihood function itself.  $l$  is approximated as

$$l(\theta_1, \theta_2, \dots, \theta_k) = \sum_{(s,t)} \log \left[ \sum_{k=1}^K g(y_{s,t}|\theta_k) \cdot w_k \right] \quad (24)$$

where  $\sum_{(s,t)}$  is the summation of ET observations at points  $s$ , and  $y_{s,t}$  refers to target data at location  $s$  and time  $t$ . The BMA method will estimate the weights  $w_k$  and parameter vectors  $\theta_k$  through maximizing the log likelihood function  $l$  (Duan and Phillips, 2010).

Before the BMA implementation, a simple linear regression was used to remove the bias in each model, and both of ET observations

and simulations were pre-processed using the Box–Cox transformation prior to the BMA procedure to make them close to the Gaussian distribution (Duan et al., 2007; Vrugt and Robinson, 2007; Slougher et al., 2010; Madadgar and Moradkhani, 2014).

### 3.3. Water balance equation

ET estimates were validated over the regional scale based on the regional water balance equation. ET can be estimated by the total precipitation,  $P$ , the runoff,  $R$ , and the change of water storage,  $\Delta S$  (Verstraeten et al., 2008):

$$ET = P - R - \Delta S \quad (25)$$

Over large areas, the  $\Delta S$  was often neglected at the annual time scale so that ET can be estimated from the observation of the precipitation and runoff (Hobbins et al., 2004; Teuling et al., 2009; Mueller et al., 2011). Therefore, the precipitation and runoff were used to calculate ET on the watershed scale.

### 3.4. Statistical analysis

A linear model was used to analyze regional trends in ET ( $z_t = bx_t + y_0$ , Zhang et al., 2009; Chen et al., 2014), where  $t$  is the time,  $b$  is the slope and  $y_0$  is the intercept of the regression line.  $SE(b)$  is the standard deviation of  $b$ . When  $|b/SE(b)| < 1.0$ , the trend is weak; when  $1.0 \leq |b/SE(b)| \leq t_{0.05}$  ( $t_{0.05}$  is the 5% critical value of Student's  $t$ -distribution), the trend is moderate; and when  $|b/SE(b)| \geq t_{0.05}$ , the trend is statistically significant and strong. These categories were further stratified into six classes according to the slopes of the statistical trends: positive weak, positive moderate, positive strong, negative weak, negative moderate, and negative strong.

Four metrics were used to evaluate model performance. The coefficient of determination ( $R^2$ ) represents how much the model can explain the variations in the observations. The root mean square error (RMSE) quantifies the difference between simulations and observations. The Kling–Gupta efficiency (KGE) was used to assess the model performance comprehensively (Gupta et al., 2009). The KGE is calculated as:

$$KGE = 1 - ED \quad (26)$$

$$ED = \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (27)$$

$$\alpha = \sigma_s / \sigma_o \quad (28)$$

$$\beta = \mu_s / \mu_o \quad (29)$$

where  $ED$  is the Euclidian distance from the ideal point,  $r$  is the correlation coefficient between the simulations and the observations,  $\mu_o$  and  $\sigma_o$  are the mean and standard deviation of the observations,  $\mu_s$  and  $\sigma_s$  are the mean and standard deviation of the simulations,  $\alpha$  is a measure of the relative variability in the simulated and observed values, and  $\beta$  is the ratio of the mean values of simulations and observations. Without any simulation errors, the values of the three components are,  $r = 1$ ,  $\alpha = 1$ , and  $\beta = 1$ . In this condition, KGE value is 1.

### 3.5. Model selection

Two ensemble strategies were compared in this study: an ensemble with all models (BMA\_All) and an ensemble with the selected models (BMA\_Best). There are 247 different combinations (the full array of  $C_8^2 + C_8^3 + C_8^4 + C_8^5 + C_8^6 + C_8^7 + C_8^8$ ) of eight models on the premise that there are at least two models participating the ensemble in each strategy. The models were selected according

to the KGE values which comprehensively included the information of the bias, RMSE and  $R^2$ . The best strategy named as BMA\_Best which had the highest KGE value at the calibration EC sites. In this study, BMA-Best is not the combination of the best individual models. This study aims to improve the accuracy of ET estimates, therefore BMA-Best is the combination of several models and achieves the best ET estimates according to KGE value. The weights of all eight ET models were determined by randomly selecting half of the measurements and other half of the measurements were used to validate the BMA model. The weights and models in the BMA\_Best were used to merge ensemble at the validation sites and calculate the regional ET over China.

### 3.6. Box–Cox transformation

Before we applied BMA, a Box–Cox transformation was first used on both the ensemble members and observations. The Box–Cox transformation is given as follows:

$$L_t = \frac{y_t^\lambda - 1}{\lambda}, (\lambda \neq 0) \quad (30)$$

$$L_t = \log(y_t), (\lambda = 0)$$

where  $y_t$  is the original variable,  $L_t$  is the transformed variable,  $\lambda$  is the Box–Cox coefficient. We derive a common optimal estimate of  $\lambda$  for all ensemble members and the observations (Duan et al., 2007).

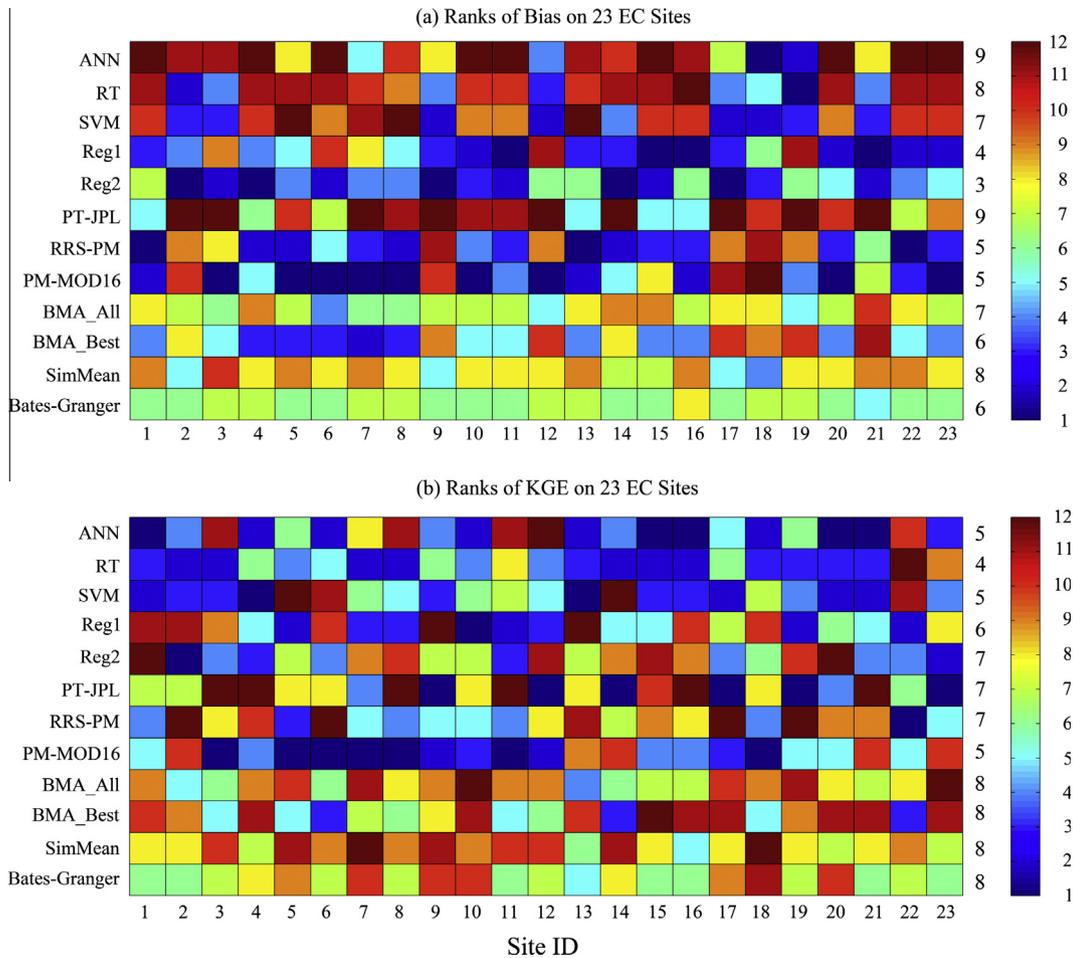
## 4. Results

### 4.1. Model validation

The performances of all 8 models were examined at 23 EC sites. At a given site, the worst model was recorded as ‘one’ according to the bias values and KGE values, and the best model was given ‘twelve’ points (Fig. 2). It should be noted that some models have the middle performance over all 23 EC sites, such as Reg1. On the contrary, other models showed a high performance at some sites but a poor performance at other sites. For example, PM-MOD16 has the largest bias at 6 EC sites and has the smallest bias at 4 EC sites (All of these sites are CRO sites). PT-JPL has the largest KGE on 7 sites (Two MF sites, two GRA sites, two CRO sites and one EBF site) and has the smallest KGE on 6 sites (Six different vegetation types). The ensemble methods ranked better than single models almost on all sites. The KGE of BMA\_Best arranged at first three on almost half sites. These results highlighted the necessity of using the BMA method to estimate ET.

The empirical models showed the negative KGE values at several sites (i.e. CN-Mau, CN-Qya and CN-Tyc). There were less than one-year measurements at the three sites and which have been separated for training and validating models. Only a small quantity of measurements was used to training empirical models, and few of information was included into the models. Therefore, the poor model performance was found at these sites.

The comparison of BMA ET estimates ( $ET_e$ ) and observed ET ( $ET_o$ ) at the EC sites showed that the BMA method performed better than all individual models (Fig. 3). However, ensemble with all models (BMA\_All) was not the best strategy compared with ensemble by selected models (BMA\_Best). The model performance was examined on all possible combinations of multiple models and found that an ensemble with four models (Reg2, PT-JPL, RRS-PM, and PM-MOD16), primarily physically-based, showed the best model performance (Fig. 3). Although the  $R^2$  value for the BMA\_Best was lower (Fig. 3b) and the bias value of BMA\_Best was higher (Fig. 3c) than some single-methods in the ensemble, no single-method was better than BMA\_Best on both  $R^2$  and bias value. In addition, the RMSE of ensemble at most EC sites were



**Fig. 2.** The performance rank of ET models at 23 EC sites according to bias (a) and KGE (b). The site ID indicates 23 EC sites (see Table 1). The color bar indicates the record, and the one and twelve represent the worst and best model performance, respectively. The numbers on the left sides of color bar indicate the average score of model performance. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

much lower than single-methods in ensemble (Fig. 3c). Overall, the BMA\_Best has the best performance considering both of  $R^2$  value, bias, and RMSE comprehensively. The RMSE of BMA\_Best is decreased 4% compared with the lowest RMSE single model. The KGE values (increased by 4% compared with the highest KGE single model) indicated that the BMA ensemble method improved the accuracy of ET estimates (Fig. 3a). The BMA results were compared with the simple mean and Bates–Granger methods (Bates and Granger, 1969). The results showed that BMA had a better result. The frequency distributions of the predictive errors also showed that two BMA estimates substantially improved the estimation accuracy (Fig. 4). For example, BMA\_All estimates increased the percentage of errors around zero and decreased large positive and negative errors.

The BMA\_Best has a good performance at both the calibration and validation EC sites (Fig. 5). The BMA\_Best can explain 80% and 75% of the observed variations of ET at the calibration and validation sites, respectively (Fig. 5), and the KGE and RMSE are 0.85 and 0.60 mm/day at the validation sites. The results showed that the BMA\_Best strategy can capture the ET variance and has a good model performance.

Two ensemble experiments with different strategies showed that the ensemble performance was influenced by the selected models (Fig. 6). In the first experiment, five empirical models (ANN, SVM, RT, Reg1, and Reg2) were merged by BMA method and then the process-based models (RRS-PM, PM-MOD16, and PT-JPL, see Fig. 6a) were added. In the second experiment, the three

process-based models were merged by BMA method and then the empirical models were added (Fig. 6b). The results showed significant differences between groups in these two experiments. The performance of the BMA method had a significant improve with the increasing amount of process-based models and had a significant decline with the increasing amount of empirical models in the ensemble strategy.

The comparison of the water balance principle estimates and BMA estimates over the eleven river basin scales was shown in Fig. 7. On average, the BMA\_Best ET estimates explained 86% of variations of annual ET over the eleven watersheds, and the RMSE is 169 mm/year. According to the  $R^2$  and RMSE, the spatial results calculated from the BMA method were reliable through the verification of a water balance approach.

#### 4.2. Spatial and temporal patterns of ET

The spatial ET calculated from the BMA\_Best strategy indicated the increasing ET from northwest to southeast China (Fig. 8a) and the averaged ET over all of China is  $625 \pm 10$  mm/year. The annual mean ET was found to be lowest in cold and arid regions, intermediate at temperate regions, and highest over the humid tropics ( $<23.43^\circ\text{N}$ ) and sub-tropics ( $23.5\text{--}40^\circ\text{N}$ ). Large areas showed that strongly increasing ET accounted for approximately 40% of China (Fig. 8b). In particular, 61% of the area in China showed positive ET trend and only 39% of China in the northwest showed negative trends. Average ET had different spatial distribution and trend

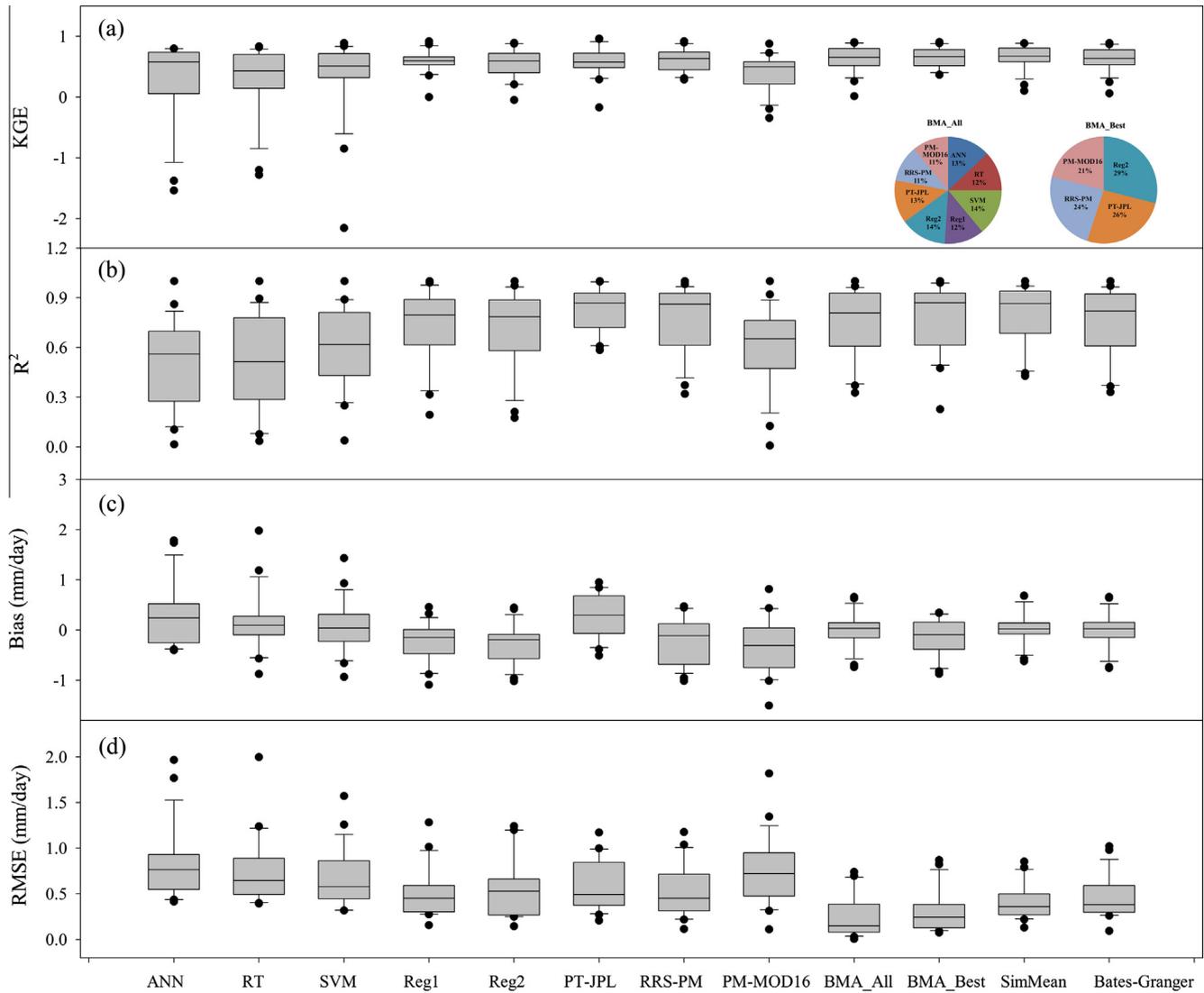


Fig. 3. Comparison of simulations of 8 models and ensemble methods at 23 EC sites. (a) KGE and the weights for BMA\_All and BMA\_Best (Pie charts), (b) R<sup>2</sup>, (c) bias and (d) RMSE.

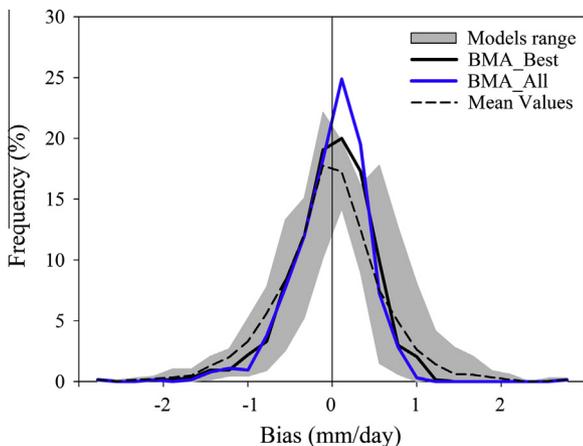


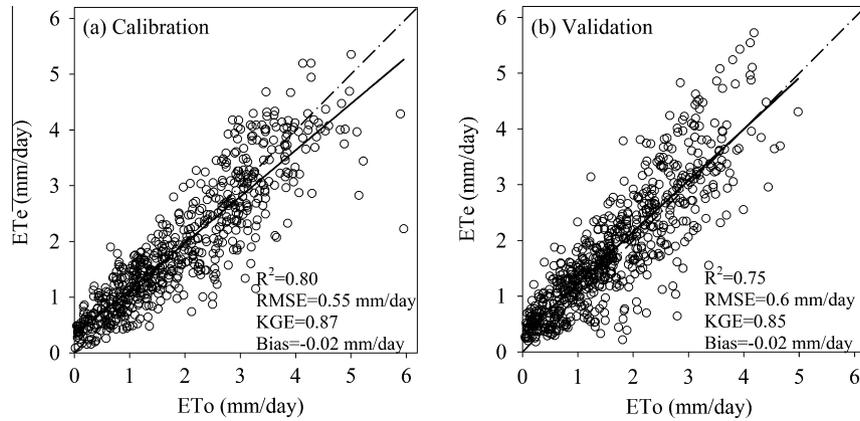
Fig. 4. The frequency distributions of the model errors of BMA\_Best (Merge selected models) and BMA\_All (Merge all models). The grey shades indicate the range of the eight models used in the BMA method (the difference between maximum and minimum values of eight models).

pattern compared with the single models. It was less than PT-JPL in the northern China and larger than RRS-PM and PM-MOD16 in the southern China (Chen et al., 2014). Large model differences occurred in southwest China and Northeast China showed small model differences in annual mean ET (Fig. 8c).

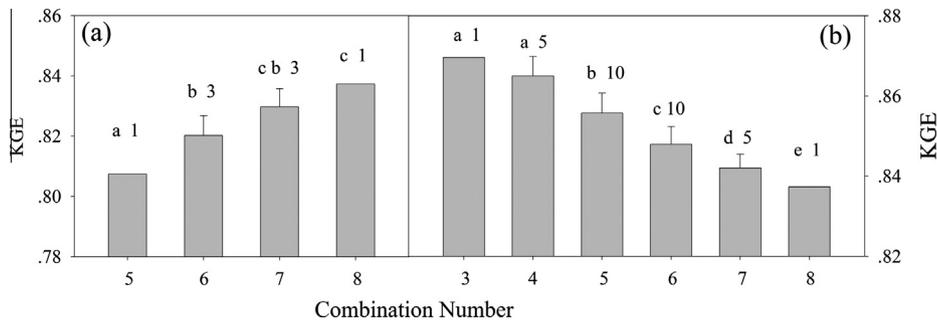
The ensemble ET had a significant increase in China from 1982 to 2009 with the trends of 0.4 mm/year (Fig. 9b). Please note that the El Niño years were not considered when analyzing the ET trend. The eight models have large range for annual mean ET of China and the BMA\_ET is the range of the eight single models (Fig. 9a). The long-term change of the ensemble ET was consistent with some single models such as SVM, Reg1, Reg2 and RRS-PM. They all showed significant increase in ET (Chen et al., 2014). Moreover, ET increased sharply in the El Niño years (i.e. 1990 and 1998).

### 5. Discussion

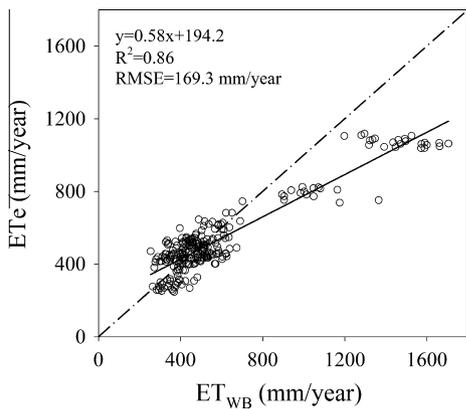
The BMA method successfully improved the ET estimate accuracy with increased KGE compared to the single models by



**Fig. 5.** The 8-day observed evapotranspiration  $ET_o$  at EC sites versus the predicted ET from the ensemble  $ET_e$ . The solid line is the 1:1 line and the short dashed lines are the regression lines.



**Fig. 6.** The comparison of Bayesian Model Averaging (BMA) method performance for various models combinations. (a) Indicates the comparison based on five empirical models (ANN, SVM, RT, Reg1, and Reg2), and (b) based on three processes-based models (RRS-PM, PM-MOD16, and PT-JPL). The numbers in the x-axis refers to the numbers of ET models for BMA ensemble. The letters above the bars indicate the significance of the differences between different BMA ET models combinations. The numbers above each bar indicate the numbers of the BMA ensemble combinations with the same ET model numbers. For example, the number “3” above the second bar in Fig. 6a means there are 3 BMA ensemble combinations of that have 6 ET models.

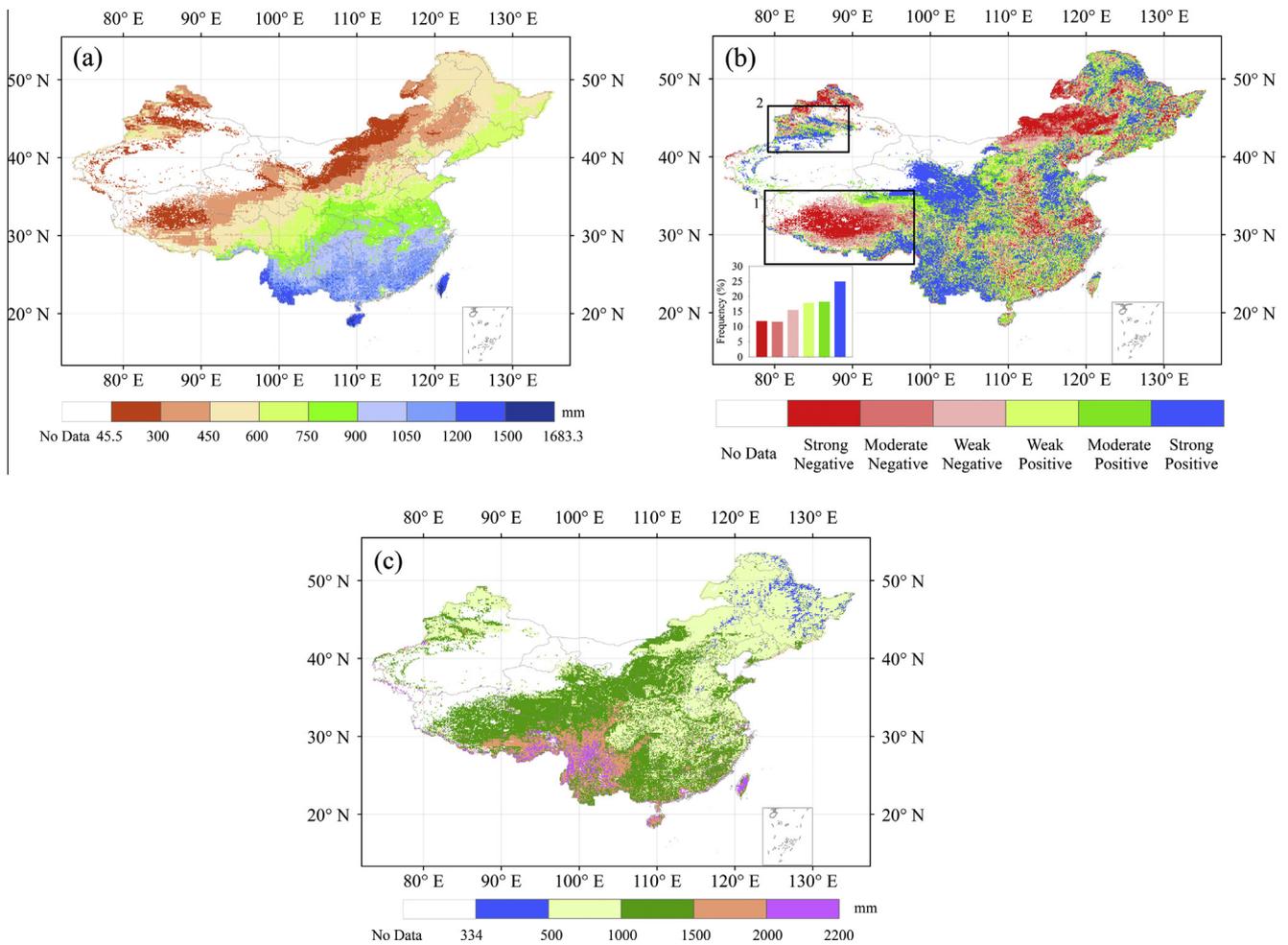


**Fig. 7.** Comparison of the ET estimated by BMA ( $ET_e$ ) and ET calculated by the water balance approach ( $ET_{wb}$ ) over the 11 river basins. Open dots represent annual mean ET for each site-year.

8–23%. The results of this article indicate that the performance of member models strongly impacts the accuracy of the ensemble estimates. Other lines of studies support the conclusion that model performance of ensembles relies on the accuracy of the individual members (Hansen and Salamon, 1990). Fig. 6 showed that the poor models will decrease the accuracy of BMA estimates, therefore numerous studies only selected several models with the high performance for multiple model ensemble simulations (Wichard, 2006; Wang and Overland, 2012).

The performance of the ensemble estimates was not only evaluated at EC sites but also on the spatial scale based on the water balance principle. Eleven river basins with different sizes were used in this study. The results show a good performance over the most of river basins, but the discrepancies between predicted ET by BMA method ( $ET_e$ ) and estimated ET by water balance equation ( $ET_{wb}$ ) still occurred mainly over the several basins especially with large watershed area (i.e. Da tong) (Fig. 7). Uncertainty of ET estimates by BMA method probably is one of the important reasons. Moreover, the accuracy of interpolated precipitation dataset also plays an important role for estimating regional ET based on the water balance equations. Previous study highlighted that the current interpolated precipitation datasets show the low performance, and which will impact the evaluation of regional water balance (Yuan et al., 2014; Fu et al., 2015).

The BMA estimate of annual mean ET over China was  $625 \pm 10$  mm/year, which was comparable to other estimates, such as LandFlux-EVAL ET and MERRA ET. For the eight models in this study, the estimated mean annual ET ranged from 500 to 851 mm/year (Fig. 10). Based on the various ET models, other studies reported 797 mm/year (Liu et al., 2008), 443 mm/year (Zhou et al., 2009), and 500 mm/year (Li et al., 2014a). This study calculated China’s average ET of the LandFlux-EVAL synthesis ET. It included four kinds of merged synthesis products: single ET categories only (LandFlux-EVAL Diagnostic), land surface models ET (LandFlux-EVAL LSM), reanalyses ET (LandFlux-EVAL Reanalyses) and from all the three categories (LandFlux-EVAL synthesis All ET) (Mueller et al., 2013). China’s average ET from four kinds of LandFlux-EVAL ensemble methods changed from 406 mm/year to

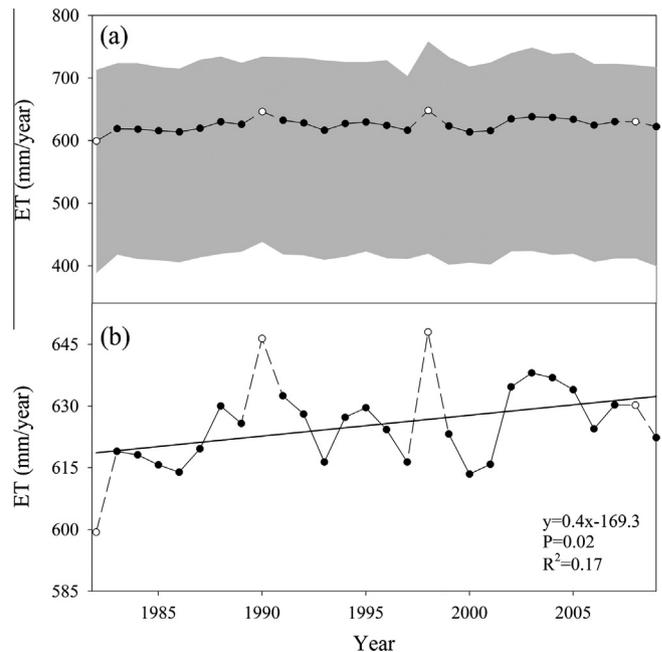


**Fig. 8.** The spatial distribution of ET estimates derived by BMA. (a) Mean annual ET from 1982 to 2009; (b) Long-term ET trend from 1982 to 2009; (c) Difference of maximum and minimum ET estimates among eight models. The rectangle #1 indicates the Tibetan Plateau area, and the rectangle #2 indicates Tianshan Region.

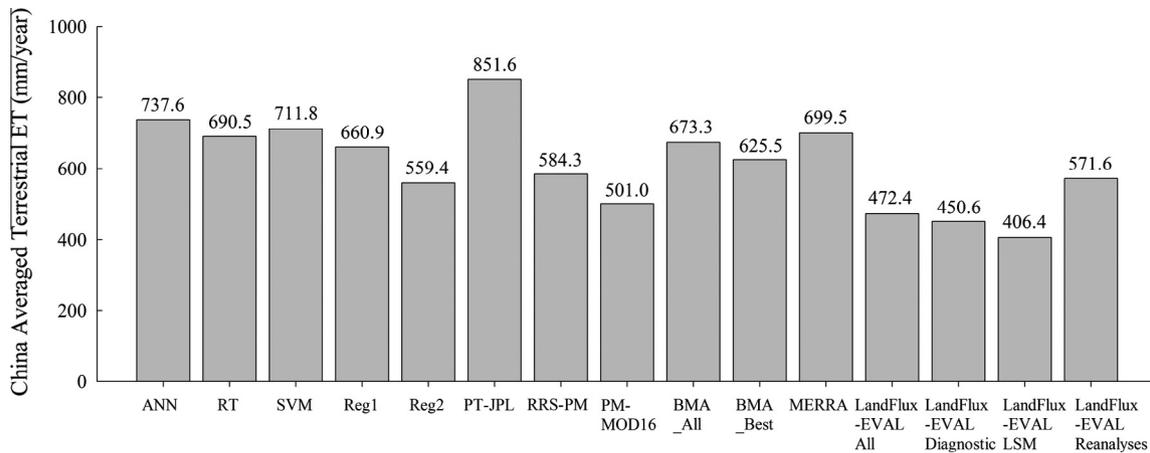
572 mm/year. The annual mean ET was larger than LandFlux-EVAL ensemble methods and mostly closed to the LandFlux-EVAL reanalyses ET which contained the MERRA reanalysis in the ensemble (Mueller et al., 2011). The BMA ensemble results were similar to the MERRA ET. This may be because this study used the MERRA datasets as the input data.

The long-term variance trend of ET agreed with the results of other estimates (Cong et al., 2008; Zhou et al., 2009). For example, Cong et al. (2008) found that the evaporation has increased since 1980 and Zhou et al. (2009) found that China's annual average ET nationwide increased during the 1990s. Fig. 9 clearly demonstrated the largest ET estimates at two El Niño years, 1990 and 1998. El Niño caused heavy rainfall in southern China and the flood in 1998 (Lau and Weng, 2001). The water supply during El Niño was more than that in other years. This contributed to the larger ET estimates during El Niño. Other lines of studies also supported this conclusion on the higher ET over regional scales at El Niño years (Simpson et al., 1993; Dai et al., 1997; Dai and Wigley, 2000; Jung et al., 2010).

The pattern of ET (Fig. 9) at the regional scale is similar to the precipitation pattern in other studies (Qian and Lin, 2005). This is because the change in precipitation played a key role in the change of the estimated ET for most parts of China (Gao et al., 2007). The increasing ET trend in the southern of Tibetan Plateau and Tianshan region may be caused by the increasing melting of snow and glaciers (Xu et al., 2008; Li et al., 2011). The large area in the Tibetan Plateau showed the decreased ET trend. The plain of



**Fig. 9.** The interannual variability of the ET simulations of the ensemble (black dots) and the range of the eight models (grey shadows) from 1982 to 2009. The range indicates the difference of the maximum and minimum ET estimates among the eight models. The open circles in (b) indicate the El Niño years.



**Fig. 10.** The mean annual averaged ET for the different models and methods. MERRA is the reanalysis ET product of Modern Era Retrospective–Analysis for Research and Applications (MERRA) (Bosilovich, 2008). The LandFlux-EVAL synthesis ET included four kinds of merged synthesis products. They were created from single ET categories only (diagnostic ET data sets (LandFlux-EVAL Diagnostic), land surface models (LSMs) ET (LandFlux-EVAL LSM) and reanalyses ET (LandFlux-EVAL Reanalyses)), and from all three categories (LandFlux-EVAL synthesis All ET) (Mueller et al., 2013).

North China has decreasing precipitation (Menon et al., 2002) and increasing ET because of the increasing temperature and radiation (Chen et al., 2014).

Satellite-based data is one of the important inputs of ET models in this study. There are no attempts to improve the quality of the satellite-based data, therefore, any noises of satellite-based data would have been transferred to ET estimates. For example, MODIS NDVI was produced by Terra surface reflectance after correcting for the impacts of cloud contamination, aerosols, and ozone absorption. NDVI noises and errors are inevitable at the flux tower footprint scale, and which will contribute proportionally to ET prediction.

An accurate ET estimation, especially at regional scales, is critical for improving water and land resource management, weather and climate forecasts, drought detection and assessment, predictions of agricultural productivity, and regional hydrological applications. Remote sensing is a useful tool for these studies which need the data over heterogeneous surfaces. In this study, BMA method was used to merge the satellite-based models to get ET productions in China. The Bayesian ET estimates are very useful for future studies to characterize the water availability of terrestrial ecosystems, the assessment of climate change impacts, and to provide guidance to the agriculture in China.

## 6. Conclusions

BMA outperforms the best participating single models of the eight models. Two ensemble strategies were used: BMA\_All and BMA\_Best. The validation results showed that the BMA estimates are closer to the observations than each models individually. BMA\_Best strategy performed better than BMA\_All at most EC sites. The regional ET calculated from the BMA\_Best strategy showed an increasing ET from northwest to southeast China. The results showed an increasing ET trend in most of China from 1982 to 2009. The mean annual ET over terrestrial ecosystem in China is  $625 \pm 10$  mm/year. The regional water balance analysis showed that the regional calculations of ensemble ET were reliable.

## Acknowledgements

This study was supported by the National Science Foundation for Excellent Young Scholars of China (41322005), the National High Technology Research and Development Program of China

(863 Program) (2013AA122003), State Key Laboratory of Environmental Change and Natural Disaster (2015jzhz03) and Program for New Century Excellent Talents in University (NCET-12-0060). We thank the Coordinated Observations and Integrated Research over Arid and Semi-arid China (COIRAS) for providing the eddy covariance flux data. We also acknowledge all the principal investigators, AsiaFlux and ChinaFlux for their volunteer contribution for data distribution. We thank Takahisa Maeda and Samreong Panuthai for providing the data of ID-Bks and TL-Skr in AsiaFlux. This study uses the LandFlux-EVAL merged benchmark synthesis products of ETH Zurich produced under the aegis of the GEWEX and ILEAPS projects (<http://www.iac.ethz.ch/url/research/LandFlux-EVAL/>). We thank Jun Asanuma, Shiping Chen, Shijie Han, Huizhi Liu, Shihua Lu, Akira Miyata, Zhuguo Ma, Takeshi Ohta, Jianwei Tang, Runyuan Wang, Shaoqiang Wang, Tianbao Zhao and Tonggang Zha for using their monitoring data.

## References

- Bates, J., Granger, C., 1969. The combination of forecasts. *Oper. Res. Quart.* 20, 451–468.
- Baik, J., Choi, M., 2015. Evaluation of remotely sensed actual evapotranspiration products from COMS and MODIS at two different flux tower sites in Korea. *Int. J. Remote Sens.* 36 (1), 375–402.
- Bosilovich, M., 2008. NASA's modern era retrospective-analysis for research and applications: integrating Earth observations. *EarthZine* (Online), Retrieved on, 26.
- Chen, Y., Xia, J.Z., Liang, S.L., Feng, J.M., Fisher, J.B., Li, X., Li, X.L., Liu, S.G., et al., 2014. Comparison of satellite-based evapotranspiration models over terrestrial ecosystems in China. *Remote Sens. Environ.* 140, 279–293. <http://dx.doi.org/10.1016/j.rse.2013.08.045>.
- 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 (3), 285–304. <http://dx.doi.org/10.1016/j.rse.2006.07.007>.
- Cong, Z., Yang, D., Lei, Z., 2008. Did evaporation paradox disappear after the 1980s? A case study for China. In: *Geophysical Research Abstracts*. EGU, Munich, Germany.
- Dai, A., Wigley, T.M.L., 2000. Global patterns of ENSO-induced precipitation. *Geophys. Res. Lett.* 27 (9), 1283–1286. <http://dx.doi.org/10.1029/1999GL011140>.
- Dai, A., Fung, I.Y., Del Genio, A.D., 1997. Surface observed global land precipitation variations during 1900–88. *J. Clim.* 10 (11), 2943–2962. [http://dx.doi.org/10.1175/1520-0442\(1997\)010<2943:SOGLPV>2.0.CO;2](http://dx.doi.org/10.1175/1520-0442(1997)010<2943:SOGLPV>2.0.CO;2).
- Dawson, C.W., Wilby, R., 1998. An artificial neural network approach to rainfall-runoff modelling. *Hydrol. Sci. J.* 43, 47–66.
- Duan, Q., Phillips, T.J., 2010. Bayesian estimation of local signal and noise in multimodel simulations of climate change. *J. Geophys. Res.* 115 (D18), D18123. <http://dx.doi.org/10.1029/2009JD013654>.
- Duan, Q., Ajami, N.K., Gao, X., Sorooshian, S., 2007. Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Adv. Water Resour.* 30 (5), 1371–1386. <http://dx.doi.org/10.1016/j.advwatres.2006.11.014>.

- Eamus, D., 2003. How does ecosystem water balance affect net primary productivity of woody ecosystems? *Funct. Plant Biol.* 30 (2), 187–205. <http://dx.doi.org/10.1071/FP02084>.
- Ellison, A.M., 2004. Bayesian inference in ecology. *Ecol. Lett.* 7 (6), 509–520. <http://dx.doi.org/10.1111/j.1461-0248.2004.00603.x>.
- Fernandez, C., Ley, E., Steel, M.F., 2001. Model uncertainty in cross-country growth regressions. *J. Econom.* 16 (5), 563–576.
- 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 (3), 901–919. <http://dx.doi.org/10.1016/j.rse.2007.06.025>.
- Foken, T., 2008. The energy balance closure problem: an overview. *Ecol. Appl.* 18 (6), 1351–1367. <http://dx.doi.org/10.1890/06-0922.1>.
- French, A.N., Hunsaker, D.J., Thorp, K.R., 2015. Remote sensing of evapotranspiration over cotton using the TSEB and METRIC energy balance models. *Remote Sens. Environ.* 158, 281–294.
- Fu, Y., Xia, J.Z., Yuan, W.P., Xu, B., Wu, X.X., Chen, Y., Zhang, H., 2015. Assessment of multiple precipitation products over major river basins of China. *Theor. Appl. Climatol.* <http://dx.doi.org/10.1007/s00704-014-1339-0>.
- Gao, G., Chen, D., Xu, C.Y., Simelton, E., 2007. Trend of estimated actual evapotranspiration over China during 1960–2002. *J. Geophys. Res.* 112 (D11120). <http://dx.doi.org/10.1029/2006JD008010>.
- Global Modeling and Assimilation Office, 2004. File Specification for GEOS-DAS Gridded Output Version 5.3, Rep., NASA Goddard Space Flight Center, Greenbelt, Maryland.
- Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. *J. Hydrol.* 377 (1–2), 80–91. <http://dx.doi.org/10.1016/j.jhydrol.2009.08.003>.
- Hagedorn, R., Doblas-reyes, F.J., Palmer, T.N., 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting – I. Basic concept. *Tellus* 57A, 219–233. <http://dx.doi.org/10.1111/j.1600-0870.2005.00103.x>.
- Hansen, L.K., Salamon, P., 1990. Neural network ensembles. *IEEE Trans. Pattern Anal. Mach. Int.* 12 (10), 993–1001. <http://dx.doi.org/10.1109/34.58871>.
- Hobbins, M.T., Ramirez, J.A., Brown, T.C., 2004. Trends in pan evaporation and actual evapotranspiration across the conterminous US: paradoxical or complementary. *Geophys. Res. Lett.* 30 (13), L13503. <http://dx.doi.org/10.1029/2004GL019846>.
- Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging: a tutorial. *Stat. Sci.* 14 (4), 382–401. <http://dx.doi.org/10.2307/2676803>.
- Hutchinson, M.F., 1995. Interpolating mean rainfall using thin plate smoothing splines. *Int. J. Geogr. Inf. Sci.* 9, 385–403.
- Hutchinson, M.F., 2004. ANUSPLIN version 4.3. Centre for Resource and Environmental Studies, Australian National University, Canberra. <<http://fennerschool.anu.edu.au/publications/software/anusplin.php>>.
- Irmak, S., Irmak, A., Allen, R.G., Jones, J.W., 2003. Solar and net radiation-based equations to estimate reference evapotranspiration in humid climates. *J. Irrig. Drain. Eng. – ASCE* 129 (5), 336–347.
- Jiang, S., Ren, L., Hong, Y., Yong, B., Yang, X., Yuan, F., Ma, M., 2012. Comprehensive evaluation of multi-satellite precipitation products with a dense rain gauge network and optimally merging their simulated hydrological flows using the Bayesian model averaging method. *J. Hydrol.* 452–453 (25), 213–225. <http://dx.doi.org/10.1016/j.jhydrol.2012.05.055>.
- Jiménez, C., Prigent, C., Mueller, B., Seneviratne, S.I., McCabe, M.F., Wood, E.F., Rossow, W.B., et al., 2011. Global intercomparison of 12 land surface heat flux estimates. *J. Geophys. Res.* 116, D02102. <http://dx.doi.org/10.1029/2010JD014545>.
- June, T., Evans, J.R., Farquhar, G.D., 2004. A simple new equation for the reversible temperature dependence of photosynthetic electron transport: a study on soybean leaf. *Funct. Plant Biol.* 31 (3), 275–283. <http://dx.doi.org/10.1071/FP03250>.
- Jung, M., Reichstein, M., Bondeau, A., 2009. Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model. *Biogeosciences* 6, 2001–2013. <http://dx.doi.org/10.5194/bg-6-2001-2009>.
- Jung, M., Reichstein, M., Ciais, P., Seneviratne, S.I., Sheffield, J., Goulden, M.L., Bonan, G., et al., 2010. Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature* 467 (7318), 951–954. <http://dx.doi.org/10.1038/nature09396>.
- Keane, R.E., Ryan, K.C., Veblen, T.T., Allen, C.D., Logan, J., Hawkes, B., 2002. Cascading effects of fire exclusion in rocky mountain ecosystems: a literature review. Rep. RMRS-GTR-91, USDA Forest Service General Technical, Fort Collins, Colorado, pp. 24.
- Landeras, G., Ortiz-Barredo, A., López, J.J., 2008. Comparison of artificial neural network models and empirical and semi-empirical equations for daily reference evapotranspiration estimation in the Basque Country (Northern Spain). *Agr. Water Manage.* 95 (5), 553–565.
- Lau, K.M., Weng, H., 2001. Coherent modes of global SST and summer rainfall over China: an assessment of the regional impacts of the 1997–98 El Niño. *J. Clim.* 14 (6), 1294–1308. [http://dx.doi.org/10.1175/1520-0442\(2001\)014<1294:CMOGSA>2.0.CO;2](http://dx.doi.org/10.1175/1520-0442(2001)014<1294:CMOGSA>2.0.CO;2).
- Lettenmaier, D.P., Famiglietti, J.S., 2006. Water from on high. *Nature* 444, 562–563. <http://dx.doi.org/10.1038/444562a>.
- Leuning, R., Zhang, Y.Q., Rajaud, A., Cleugh, H., Tu, K., 2008. A simple surface conductance model to estimate regional evaporation using MODIS leaf area index and the Penman-Monteith equation. *Water Resour. Res.* 44 (10), W10419. <http://dx.doi.org/10.1029/2007WR006562>.
- Li, X., Liang, S.L., Yuan, W.P., Yu, G., Cheng, X., Chen, Y., Zhao, T., et al., 2014a. Estimation of evapotranspiration over the terrestrial ecosystems in China. *Ecohydrology* 7 (1), 139–149. <http://dx.doi.org/10.1002/eco.1341>.
- Li, X., Zecchin, A.C., Maier, H.R., 2014b. Selection of smoothing parameter estimators for general regression neural networks—applications to hydrological and water resources modelling. *Environ. Modell. Softw.* 59, 162–186.
- Li, Y., Andersen, H.E., Mcgaughey, R., 2008. A comparison of statistical methods for estimating forest biomass from light detection and ranging data. *West. J. Appl. For.* 23 (4), 223–231.
- Li, Z., Li, H., Chen, Y., 2011. Mechanisms and simulation of accelerated shrinkage of continental glaciers: a case study of Urumqi Glacier No. 1 in eastern Tianshan, Central Asia. *J. Earth Sci.* 22, 423–430. <http://dx.doi.org/10.1007/s12583-011-0194-5>.
- Liu, M., Tian, H., Chen, G., Ren, W., Zhang, C., Liu, J., 2008. Effects of land-use and land-cover change on evapotranspiration and water yield in China during 1900–2000. *Water Resour. Assoc.* 44 (5), 1193–1207. <http://dx.doi.org/10.1111/j.1752-1688.2008.00243.x>.
- Liu, R., Wen, J., Wang, X., Zhang, Y., 2015. Evapotranspiration estimated by using datasets from the Chinese FengYun-2D geostationary meteorological satellite over the Yellow River source area. *Adv. Space Res.* 55 (1), 60–71.
- Madadgar, S., Moradkhani, H., 2014. Improved bayesian multi-modeling: integration of copulas and bayesian model averaging. *Water Resour. Res.* 50 (12), 9586–9603.
- Menon, S., Hansen, J., Nazarenko, L., Luo, Y., 2002. Climate effects of black carbon aerosols in China and India. *Science* 297 (5590), 2250–2253. <http://dx.doi.org/10.1126/science.1075159>.
- Monteith, J.L., 1965. Evaporation and environment. *Symp. Soc. Exp. Biol.* 19, 205–234.
- Mu, Q., Zhao, M., Running, S.W., 2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sens. Environ.* 115 (8), 1781–1800. <http://dx.doi.org/10.1016/j.rse.2011.02.019>.
- 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 (4), 519–536. <http://dx.doi.org/10.1016/j.rse.2007.04.015>.
- Mueller, B., Hirschi, M., Jimenez, C., Ciais, P., Dirmeyer, P.A., Dolman, A.J., Fisher, J.B., et al., 2013. Benchmark products for land evapotranspiration: LandFlux-EVAL multi-dataset synthesis. *Hydrol. Earth Syst. Sci. Discuss.* 10 (1), 769–805. <http://dx.doi.org/10.5194/hess-17-3707-2013>.
- Mueller, B., Seneviratne, S.L., Jimenez, C., Corti, T., Hirschi, M., et al., 2011. Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations. *Geophys. Res. Lett.* 38 (6), L06402. <http://dx.doi.org/10.1029/2010GL046230>.
- Myneni, R.B., Ramakrishna, R., Nemani, R., Running, S.W., 1997. Estimation of global leaf area index and absorbed par using radiative transfer models. *IEEE Trans. Geosci. Remote Sens.* 35 (6), 1380–1393. <http://dx.doi.org/10.1109/36.649788>.
- Myneni, R.B., Hoffman, S., Knyazikhin, Y., Privette, J.L., Glassy, J., Tian, Y., et al., 2002. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sens. Environ.* 83 (1–2), 214–231. [http://dx.doi.org/10.1016/S0034-4257\(02\)00074-3](http://dx.doi.org/10.1016/S0034-4257(02)00074-3).
- Najafi, M.R., Moradkhani, H., Jung, I.W., 2011. Assessing the uncertainties of hydrologic model selection in climate change impact studies. *Hydrol. Process.* 25, 2814–2826. <http://dx.doi.org/10.1002/hyp.8043>.
- Priestley, C.H.B., Taylor, R.J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. *Mon. Weather Rev.* 100 (2), 81–92. [http://dx.doi.org/10.1175/1520-0493\(1972\)100<0081:OTAOSH>2.3.CO;2](http://dx.doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2).
- Qian, W., Lin, X., 2005. Regional trends in recent precipitation indices in China. *Meteorol. Atmos. Phys.* 90 (3–4), 193–207. <http://dx.doi.org/10.1007/s00703-004-0101-z>.
- Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005. Using bayesian model averaging to calibrate forecast ensembles. *Mon. Weather Rev.* 133 (5), 1155–1174. <http://dx.doi.org/10.1175/MWR2906.1>.
- Raupach, M.R., 2001. Combination theory and equilibrium evaporation. *Quart. J. Roy. Meteorol. Soc.* 127, 1149–1181. <http://dx.doi.org/10.1002/qj.49712757402>.
- Ruimy, A., Kergoat, L., Bondeau, A., 1999. Comparing global models of terrestrial net primary productivity (NPP): analysis of differences in light absorption and light-use efficiency. *Global Change Biol.* 73 (1), 56–64. <http://dx.doi.org/10.1046/j.1365-2486.1999.00007.x>.
- Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. *Nature* 323, 533–536.
- Shu, Y., Stisen, S., Jensen, K.H., Sandholt, I., 2011. Estimation of regional evapotranspiration over the North China Plain using geostationary satellite data. *Int. J. Appl. Earth Obs.* 13 (2), 192–206. <http://dx.doi.org/10.1016/j.jag.2010.11.002>.
- Simpson, H.J., Cane, M.A., Herczeg, A.L., Zebiak, S.E., Simpson, J.H., 1993. Annual river discharge in southeastern Australia related to El Niño–Southern Oscillation forecasts of sea surface temperatures. *Water Resour. Res.* 29 (11), 3671–3680. <http://dx.doi.org/10.1029/93WR01492>.
- Sloughter, J.M.L., Raftery, A.E., Gneiting, T., Fraley, C., 2007. Probabilistic quantitative precipitation forecasting using bayesian model averaging. *Mon. Weather Rev.* 135 (9), 3209–3220. <http://dx.doi.org/10.1175/MWR3441.1>.
- Sloughter, J.M., Gneiting, T., Raftery, A.E., 2010. Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. *J. Am. Stat. Assoc.* 105, 25–35. <http://dx.doi.org/10.1198/jasa.2009.ap08615>.

- Tang, R., Li, Z.-L., 2015. Evaluation of two end-member-based models for regional land surface evapotranspiration estimation from MODIS data. *Agr. Forest Meteorol.* 202, 69–82.
- Teuling, A.J., Hirschi, M., Ohmura, A., Wild, M., Reichetein, M., Ciais, P., et al., 2009. A regional perspective on trends in continental evaporation. *Geophys. Res. Lett.* 36 (2), L02404. <http://dx.doi.org/10.1029/2008GL036584>.
- Tabari, H., Grismer, M.E., Trajkovic, S., 2013. Comparative analysis of 31 reference evapotranspiration methods under humid conditions. *Irrig. Sci.* 31 (2), 107–117.
- Valipour, M., Banihabib, M., Behbahani, S., 2012. Monthly inflow forecasting using autoregressive artificial neural network. *J. Appl. Sci.* 12 (20), 2139–2147.
- Valipour, M., Banihabib, M.E., Behbahani, S.M.R., 2013. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. *J. Hydrol.* 476, 433–441.
- Valipour, M., 2014a. Application of new mass transfer formulae for computation of evapotranspiration. *J. Appl. Water Eng. Res.* 2 (1), 33–46.
- Valipour, M., 2014b. Evaluation of radiation methods to study potential evapotranspiration of 31 provinces. *Meteorol. Atmos. Phys.*, 1–15.
- Valipour, M., 2015. Importance of solar radiation, temperature, relative humidity, and wind speed for calculation of reference evapotranspiration. *Arch. Agron. Soil. Sci.* 61 (2), 239–255.
- Vapnik, V.N., Naumovich, V., Vapnik, V., 1998. *Statistical learning theory*. Wiley, New York.
- Verstraeten, W., Veroustraete, F., Feyen, J., 2008. Assessment of evapotranspiration and soil moisture content across different scales of observation. *Sensors* 8 (1), 70–117. <http://dx.doi.org/10.3390/s8010070>.
- Viallefont, V., Raftery, A.E., Richardson, S., 2001. Variable selection and Bayesian model averaging in case-control studies. *Stat. Med.* 20 (21), 3215–3230. <http://dx.doi.org/10.1002/sim.976>.
- Vinukollu, R.K., Wood, E.F., Ferguson, C.R., Fisher, J.B., 2011a. Global estimates of evapotranspiration for climate studies using multi-sensor remote sensing data: evaluation of three process-based approaches. *Remote Sens. Environ.* 115 (3), 801–823. <http://dx.doi.org/10.1016/j.rse.2010.11.006>.
- Vinukollu, R.K., Meynadier, R., Sheffield, J., Wood, E.F., 2011b. Multi-model, multi-sensor estimates of global evapotranspiration: climatology, uncertainties and trends. *Hydrol. Process.* 25 (26), 3993–4010. <http://dx.doi.org/10.1002/hyp.8393>.
- Vrugt, J.A., Robinson, B.A., 2007. Treatment of uncertainty using ensemble methods: comparison of sequential data assimilation and Bayesian model averaging. *Water Resour. Res.* 43, 1–15.
- Wang, K., Dickinson, R.E., Wild, M., Liang, S., 2010. Evidence for decadal variation in global terrestrial evapotranspiration between 1982 and 2002: 1. Model development. *J. Geophys. Res.* 115 (D20), D20112. <http://dx.doi.org/10.1029/2009JD013671>.
- Wang, K., Wang, P., Li, Z., Cribb, M., Sparrow, M., 2007. A simple method to estimate actual evapotranspiration from a combination of net radiation, vegetation index, and temperature. *J. Geophys. Res.* 112 (D15), D15107. <http://dx.doi.org/10.1029/2006JD008351>.
- Wang, M., Overland, J.E., 2012. A sea ice free summer Arctic within 30 years: an update from CMIP5 models. *Geophys. Res. Lett.* 39 (18), 2012. <http://dx.doi.org/10.1029/2012GL052868>.
- Wang, Q.J., Schepen, A., Robertson, D.E., 2012. Merging seasonal rainfall forecasts from multiple statistical models through Bayesian model averaging. *J. Clim.* 25 (16), 5524–5537. <http://dx.doi.org/10.1175/JCLI-D-11-00386.1>.
- Wichard, J.D., 2006. Model selection in an ensemble framework. In: *International Joint Conference on Neural Networks, 2006. IJCNN '06*. doi: 10.1109/IJCNN.2006.247012, pp. 2187–2192.
- Wu, H., Zhang, X., Liang, S., Yang, H., Zhou, G., 2012. Estimation of clear-sky land surface longwave radiation from MODIS data products by merging multiple models. *J. Geophys. Res.* 117 (D22), D22107. <http://dx.doi.org/10.1029/2012JD017567>.
- Xia, J., Liang, S., Chen, J., Yuan, W., Liu, S., Li, L., Cai, W., Zhang, L., et al., 2014. Satellite-based analysis of evapotranspiration and water balance in the grassland ecosystems of Dryland East Asia. *PLoS ONE* 9 (5), e97295. <http://dx.doi.org/10.1371/journal.pone.0097295>.
- Xiao, J., Zhuang, Q., Law, B.E., Chen, J., Baldaocchi, D.D., Cook, D.R., et al., 2010. A continuous measure of gross primary production for the conterminous United States derived from MODIS and AmeriFlux data. *Remote Sens. Environ.* 114 (3), 576–591. <http://dx.doi.org/10.1016/j.rse.2009.10.013>.
- Xu, C.Y., Singh, V.P., 2000. Evaluation and generalization of radiation-based methods for calculating evaporation. *Hydrol. Process.* 14 (2), 339–349.
- Xu, Z., Gong, T., Li, J., 2008. Decadal trend of climate in the Tibetan Plateau—regional temperature and precipitation. *Hydrol. Process.* 22 (16), 3056–3065. <http://dx.doi.org/10.1002/hyp.6892>.
- Yang, Y., Shang, S., Jiang, L., 2012. Remote sensing temporal and spatial patterns of evapotranspiration and the responses to water management in a large irrigation district of North China. *Agr. Forest Meteorol.* 164, 112–122.
- Yuan, W., Xu, B., Chen, Z., Xia, J., Xu, W., Chen, Y., Wu, X., Fu, Y., 2014. Validation of China-wide interpolated daily climate variables from 1960 to 2011. *Theor. Appl. Climatol.* L18501, 1–12. <http://dx.doi.org/10.1007/s00704-014-1140-0>.
- Yuan, W., Liu, S., Liang, S., Tian, Z., Liu, H., Young, C., 2012. Estimations of evapotranspiration and water balance with uncertainty over the Yukon River Basin. *Water Resour. Manage.* 26, 2147–2157. <http://dx.doi.org/10.1007/s11269-012-0007-3>.
- Yuan, W., Liu, S., Liu, H., Randerson, J., Yu, G., Tieszen, L., 2010a. Impacts of precipitation seasonality and ecosystem types on evapotranspiration in the Yukon River Basin, Alaska. *Water Resour. Res.* 46, W02514. <http://dx.doi.org/10.1029/2009WR008119>.
- Yuan, W., Liu, S., Yu, G., Bonnefond, J., Chen, J., Davis, K., Desai, A.R., Goldstein, A.H., et al., 2010b. Global estimates of evapotranspiration and gross primary production based on MODIS and global meteorology data. *Remote Sens. Environ.* 114 (7), 1416–1431. <http://dx.doi.org/10.1016/j.rse.2010.01.022>.
- Zeng, Z., Piao, S., Lin, X., Yin, G., Peng, S., Ciais, P., Myneni, R.B., 2012. Global evapotranspiration over the past three decades: estimation based on the water balance equation combined with empirical models. *Environ. Res. Lett.* 7 (1), 014026. <http://dx.doi.org/10.1088/1748-9326/7/1/014026>.
- Zhang, K., Kimball, J.S., Nemani, R.R., Running, S.W., 2010. A continuous satellite-derived global record of land surface evapotranspiration from 1983 to 2006. *Water Resour. Res.* 46 (9), 1–21. <http://dx.doi.org/10.1029/2009WR008800>.
- Zhang, K., Kimball, J.S., Mu, Q., Jones, L.A., Goetz, S.J., Running, S.W., 2009. Satellite based analysis of northern ET trends and associated changes in the regional water balance from 1983 to 2005. *J. Hydrol.* 379 (1–2), 92–110. <http://dx.doi.org/10.1016/j.jhydrol.2009.09.047>.
- Zhang, K., Kimball, J.S., Hogg, E.H., Zhao, M., Oechel, W.C., Cassano, J.J., Running, S.W., 2008. Satellite-based model detection of recent climate-driven changes in northern high-latitude vegetation productivity. *J. Geophys. Res.* 113 (G3), G03033. <http://dx.doi.org/10.1029/2007JG000621>.
- Zhou, L., Wang, S., Chen, J., Feng, X., Ju, W., Wu, W., 2009. The spatial-temporal characteristics of evapotranspiration of China's terrestrial ecosystems during 1991–2000. *Resour. Sci.* 31, 962–972.