Agricultural and Forest Meteorology 223 (2016) 151–167

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

Assessment and simulation of global terrestrial latent heat flux by synthesis of CMIP5 climate models and surface eddy covariance observations

Yunjun Yao[a,*], Shunlin Liang[a], Xianglan Li[b], Shaomin Liu[a], Jiquan Chen[c], Xiaotong Zhang[a], Kun Jia[a], Bo Jiang[a], Xianhong Xie[a], Simon Munier[d], Meng Liu[a], Jian Yu[a], Anders Lindroth[e], Andrej Varlagin[f], Antonio Raschi[g], Asko Noormets[h], Casimiro Pio[i], Georg Wohlfahrt[j,k], Ge Sun[l], Jean-Christophe Domec[m,n], Leonardo Montagnani[o,p], Magnus Lund[q], Moors Eddy[m,n], Peter D. Blanken[r], Thomas Grünwald[s], Sebastian Wolf[t], Vincenzo Magliulo[u]

[a] State Key Laboratory of Remote Sensing Science, School of Geography, Beijing Normal University, Beijing, 100875, China
[b] College of Global Change and Earth System Science, Beijing Normal University, Beijing, 100875, China
[c] CGCEO/Geography, Michigan State University, East Lansing, MI 48824, USA
[d] Laboratoire d'études en géophysique et océanographie spatiales, LEGOS/CNES/CN RS/IRD/UPS-UMR5566, Toulouse, France
[e] GeoBiosphere Science Centre, Lund University, Sölvegatan 12, 223 62 Lund, Sweden
[f] A.N.Severtsov Institute of Ecology and Evolution RAS 123103, Leninsky pr.33, Moscow, Russia
[g] CNR-IBIMET, National Research Council, Via Caproni 8, 50145 Firenze, Italy
[h] Dept. Forestry and Environmental Resources, North Carolina State University, 920 Main Campus Drive, Ste 300, Raleigh, NC 27695, USA
[i] CESAM & Departamento de Ambiente e Ordenamento, Universidade de Aveiro, 3810-193 Aveiro, Portugal
[j] Institute for Ecology, University of Innsbruck, Sternwartestr. 15, A-6020 Innsbruck, Austria
[k] European Academy Bolzano, Drususallee 1, 39100 Bolzano, Italy
[l] Eastern Forest Environmental Threat Assessment Center, Southern Research Station, United States Department of Agriculture Forest Services, Raleigh, NC 27606, USA
[m] Climate Change and Adaptive Land and Water Management, Alterra Wageningen UR, PO Box 47, 6700 AA Wageningen, The Netherlands
[n] VU University Amsterdam, Boelelaan 1085, 1081HV Amsterdam, The Netherlands
[o] Faculty of Science and Technology, Free University of Bolzano, Piazza Universita 5, 39100 Bolzano, Italy
[p] Forest Services, Autonomous Province of Bolzano, 39100 Bolzano, Italy
[q] Department of Bioscience, Aarhus University, Frederiksborgvej 399, DK-4000 Roskilde, Denmark
[r] Department of Geography, University of Colorado at Boulder, 260 UCB, Boulder, CO 80309-0260, USA
[s] Technische Universität Dresden, Institute of Hydrology and Meteorology, Chair of Meteorology, D-01062 Dresden, Germany
[t] ETH Zurich, Institute of Agricultural Sciences, Universitaetsstr. 2, 8092 Zurich, Switzerland
[u] CNR-ISAFOM, National Research Council, Via Patacca 85, 80040 Ercolano (Napoli), Italy

ARTICLE INFO

Article history:
Received 19 August 2015
Received in revised form 24 March 2016
Accepted 29 March 2016

Keywords:
Global terrestrial *LE*
CMIP5
GCMs
BMA
Taylor skill score

ABSTRACT

The latent heat flux (*LE*) between the terrestrial biosphere and atmosphere is a major driver of the global hydrological cycle. In this study, we evaluated *LE* simulations by 45 general circulation models (*GCMs*) in the Coupled Model Intercomparison Project Phase 5 (*CMIP5*) by a comparison with eddy covariance (*EC*) observations from 240 globally distributed sites from 2000 to 2009. In addition, we improved global terrestrial *LE* estimates for different land cover types by synthesis of seven best *CMIP5* models and EC observations based on a Bayesian model averaging (*BMA*) method. The comparison results showed substantial differences in monthly *LE* among all *GCMs*. The model *CESM1-CAM5* has the best performance with the highest predictive skill and a Taylor skill score (*S*) from 0.51–0.75 for different land cover types. The cross-validation results illustrate that the *BMA* method has improved the accuracy of the *CMIP5 GCM's LE* simulation with a decrease in the averaged root-mean-square error (*RMSE*) by more than 3 W/m^2 when compared to the simple model averaging (*SMA*) method and individual *GCMs*. We found an increasing trend in the *BMA*-based global terrestrial *LE* (slope of 0.018 $W/m^2\,yr^{-1}$, $p < 0.05$) during the period 1970–2005. This variation may be attributed directly to the inter-annual variations in air temperature (T_a), surface incident solar radiation (R_s) and precipitation (*P*). However, our study highlights a large difference from previous studies in a continuous increasing trend after 1998, which may be caused by

* Corresponding author.
E-mail address: boyyunjun@163.com (Y. Yao).

http://dx.doi.org/10.1016/j.agrformet.2016.03.016

the combined effects of the variations of R_s, T_a, and P on *LE* for different models on these time scales. This study provides corrected-modeling evidence for an accelerated global water cycle with climate change.

1. Introduction

Latent heat flux (*LE*) is the flux of heat from the earth's surface to the atmosphere for soil evaporation, plant transpiration, and evaporation from intercepted precipitation by vegetation canopies. *LE* is a fundamental quantity for understanding ecosystem processes and functions (Sun et al., 2011) and developing general circulation models (*GCMs*) and global climatic forecasting and land surface models (*LSMs*) (Liang et al., 2010; Wang and Dickinson, 2012; Wild et al., 2015; Yao et al., 2013). During the past two decades, eddy covariance (*EC*) measurement system (*e.g. FLUXNET*) has been established to measure *LE* and sensible heat flux (*H*) exchanges between the atmosphere and land surface (Baldocchi et al., 2001; Liu et al., 2013; Twine et al., 2000; Yao et al., 2015). However, these short-term point-based *LE* measurements by *EC* are limited due to the spare coverage. Remote sensing technology has a large spatial coverage but satellites also do not directly measure *LE*, which hampers accurately understanding the long-term variations of terrestrial *LE* due to the substantial uncertainties of the individual datasets.

The Coupled Model Intercomparison Project phase 5 (*CMIP5*) from the latest 5th Intergovernmental Panel on Climate Change (*IPCC*) assessment report (*IPCC-AR5*) provides an opportunity to assess global terrestrial *LE* variations and attributions by coupling land-atmosphere interaction processes (Dirmeyer et al., 2013; Taylor et al., 2012). Relative to the previous beta version, *CMIP5* includes more than 45 *GCMs* from different modeling groups with higher spatial and temporal resolution and multiple models for a single experiment (Taylor et al., 2012).The state-of-the-art *GCMs* that are available through *CMIP5* are now widely used to investigate the theoretical mechanisms of climatic changes (Covey et al., 2003; Miao et al., 2013). Compared to the Coupled Model Intercomparison Project phase 3 (*CMIP3*) in the 4th *IPCC* assessment report (*IPCC-AR4*), the *GCMs* in *IPCC-AR5* have improved many more model types, including the Earth System Models (*ESMs*), with more interactive components, including aerosols, dynamic vegetation, atmospheric physics and carbon and hydrological cycles (Liu et al., 2013; Miao et al., 2014). Most dynamic, physical and chemical algorithms were also improved in the *IPCC AR5* models (Moss et al., 2010; Wild et al., 2013; Wild et al., 2015). These improvements will help the long-term climate forecasts, including global terrestrial inter-annual *LE* prediction.

Recently, a much broader comparison of *CMIP5* (or *CMIP3*) data and other gridded datasets had been attempted to assess *GCM* values of the global terrestrial latent and sensible heat fluxes. As reported by Mueller et al. (2011) and Wang and Dickinson (2012), the standard deviation (*SD*) of the *IPCC AR4* simulations within each category (4.6 W/m^2) is lower than those of the reference datasets, including satellite, reanalysis and *LSMs* datasets (*SD* varying from 4.9 to 5.6 W/m^2). Wild et al. (2015) reported that the *CMIP5* models varied greatly (32–46 W/m^2 for *LE*, 16–43 W/m^2 for *H*) in their calculation of the land mean *LE* and *H*, with a global land (including Antarctica) mean *LE* and *H* of 38 W/m^2 and 32 W/m^2, respectively. These inter-comparison studies, however, focus mainly on evaluating the global annual mean and the errors of the surface *LE* and *H* based on gridded datasets, with a few using eddy covariance observations (Jiménez et al., 2011; Wild et al., 2013). Meanwhile, a large number of *EC* observations from the *FLUXNET* project has the potential to be used as a reference dataset to assess the accuracy of *CMIP5 LE* results. Yet a detailed comparison between *CMIP5* modeled versus global *EC* observed *LE* among different land cover types has not been performed.

To maximize the value of *GCMs* or other multiple datasets, several merging algorithms have been effectively used to estimate global terrestrial climatic and hydrologic variables (*e.g.*, air temperature, *P* and *LE*) with high accuracy (Duan et al., 2007; Miao et al., 2013; Miao et al., 2014; Yang et al., 2012; Yao et al., 2014a; Yao et al., 2014b). Recent studies have demonstrated that even a simple multi-model ensemble, such as simple model averaging (*SMA*), is superior to an individual model (Buser et al., 2009; Duan and Phillips, 2010; Lambert and Boer, 2001; Weigel et al., 2008). Sophisticated multi-model ensemble approaches have acquired the weights of single-model contributions to improve performance by training ground-measured observations. Among these complicated ensemble methods, Bayesian model averaging (*BMA*) is one of the most promising methods that combines simulations from multiple datasets and a unanimous probability density function (*PDF*) (Duan and Phillips, 2010; Raftery et al., 2005; Wu et al., 2012). Some studies have highlighted the application of *GCM* data and the *BMA* method in global *P* and air temperature (T_a) simulations (Miao et al., 2013; Miao et al., 2014). However, there is a lack of similar studies that simulate global terrestrial *LE* by using the *BMA* method driven by *CMIP5* models and surface *EC* observations. As a result, little is accurately understood regarding spatiotemporal characterization of the response of global terrestrial *LE* to climate change over long periods.

In this study, we evaluated *LE* simulations of 45 *CMIP5* models and improved global terrestrial *LE* simulations among different land cover types by synthesis of seven best *CMIP5* models and *FLUXNET EC* observations based on the *BMA* method. Our study has three specific objectives: (1) evaluate *LE* simulations from the state-of-the-art *GCMs* of 45 *CMIP5* models with a comprehensive ground-measured *LE* flux data set; (2) use the *BMA* method to merge the seven best *CMIP5* models to generate a global terrestrial long-term (1970–2005) monthly *LE*; and (3) analyze the spatiotemporal variability in the global terrestrial *LE* and its attributions by comparing the changes of the relevant climatic variables.

2. Data

2.1. CMIP5 GCM latent heat flux simulation

IPCC-AR5 used more than 45 state-of-the-art *GCM* results as part of *CMIP5* for the World Climate Research Programme (*WCRP*). In *CMIP5*, the same numerical experiments were performed by different models of the same protocols, which are accepted for a direct comparison of these models (Guilyardi et al., 2013; Ma et al., 2014). The *LE* simulations of the historical experiments were used for the *GCMs* in this study. All of the *LE* outputs were simulated using the same initial approach, initial time, and rattled physics with an ensemble member set at r1i1p1 (Taylor et al., 2012). Monthly *CMIP5 GCM LE* simulations with 0.56–3.75° spatial resolution were used in this study. Most *CMIP5 GCM LE* simulations spanned from 1850 to 2005, including 45 *GCMs* used in this study. When combining some gridded *GCM* datasets with different spatial resolutions, they were

Table 1
Description of the 45 *CMIP5 GCMs* used in this study, mean annual global terrestrial *LE* (excluding Antarctica) and trend of the global terrestrial mean of *LE* (excluding Antarctica) from 1970 to 2005.

No.	Model name	Source	Spatial resolution	Mean annual global terrestrial LE (excluding Antarctica) (W/m^2)	Trend of the global terrestrial mean of LE (excluding Antarctica) (W/m^2 per decade)
1	ACCESS-1-0	Commonwealth Scientific and Industrial Research Organization	1.88° × 1.24°	43.7	0.14
2	ACCESS-1-3	(CSIRO) and Bureau of Meteorology (BOM), Australia		46.9	0.12
3	BCC-CSM-1-1-m	Beijing Climate Center, China Meteorological	1.13° × 1.13°	36.0	0.28
4	BCC-CSM-1-1	Administration	2.81° × 2.81°	38.0	0.30
5	BNU-ESM	College of Global Change and Earth System Science, Beijing Normal University	2.81° × 2.81°	44.1	0.43
6	CanCM4	Canadian Centre for Climate Modelling and	2.81° × 2.81°	39.3	0.23
7	CanESM2	Analysis		37.5	0.23
8	CCSM4	National Center for Atmospheric Research	1.25° × 0.94°	43.5	0.21
9	CESM1-BGC	Community Earth System	1.25° × 0.94°	43.4	0.12
10	CESM1-CAM5	Model Contributors		40.4	0.07
11	CESM1-FASTCHEM			43.5	0.20
12	CESM1-WACCM		2.50° × 1.88°	43.0	0.17
13	CMCC-CESM	Centro Euro-Mediterraneo per I Cambiamenti Climatici	3.75° × 3.75°	42.8	0.41
14	CMCC-CM		0.75° × 0.75°	36.0	0.25
15	CMCC-CMS		1.88° × 1.88°	39.2	0.34
16	CNRM-CM5-2	Centre National de Recherches Meteorologiques/Centre	1.41° × 1.41°	40.2	0.26
17	CNRM-CM5	Europeen de Recherche et Formation Avancees en Calcul Scientifique		40.6	0.23
18	CSIRO-Mk-3-6-0	Commonwealth Scientific and Industrial Research Organization in collaboration with Queensland Climate	1.88° × 1.88°	39.4	0.14
19	CSIRO-Mk3L-1-2	Change Centre of Excellence	5.63° × 3.21°	37.8	0.01
20	EC-EARTH	EC-EARTH consortium	1.13° × 1.00°	40.1	0.17
21	FGOALS-g2	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences and CESS, Tsinghua University	2.81° × 3.00°	43.8	0.23
22	FIO-ESM	The First Institute of Oceanography, SOA, China	2.81° × 2.81°	44.6	0.18
23	GFDL-CM3	NOAA Geophysical Fluid	2.50° × 2.00°	43.9	0.15
24	GFDL-ESM2G	Dynamics Laboratory		43.0	0.31
25	GFDL-ESM2M			42.7	0.35
26	GISS-E2-H-CC	NASA Goddard Institute for Space Studies	2.50° × 2.00°	47.2	0.08
27	GISS-E2-H			47.2	0.13
28	GISS-E2-R-CC			45.7	0.15
29	GISS-E2-R			45.8	0.25
30	HadGEM2-AO	Met Office Hadley Centre (additional HadGEM2-ES realizations contributed by Instituto Nacional de Pesquisas	1.88° × 1.24°	43.8	0.13
31	HadGEM2-ES	Espaciais)		44.8	0.24
32	inmcm4	Institute for Numerical Mathematics	2.00° × 1.50°	43.6	0.10
33	IPSL-CM5A-LR	Institut Pierre-Simon Laplace	3.75° × 1.88°	37.8	0.24
34	IPSL-CM5A-MR		2.50°×1.26°	37.7	0.17
35	IPSL-CM5B-LR		3.75°×1.88°	36.0	0.25
36	MIROC-ESM-CHEM	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University	2.81° × 2.81°	47.5	0.43
37	MIROC-ESM	of Tokyo), and National Institute for Environmental Studies		48.1	0.26
38	MIROC4h	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	0.56° × 0.56°	41.1	0.21
39	MPI-ESM-LR	Max Planck Institute for Meteorology	1.88° × 1.88°	41.5	0.15
40	MPI-ESM-MR			42.5	0.36
41	MPI-ESM-P			41.4	0.18
42	MRI-CGCM3	Meteorological Research Institute	1.13° × 1.13°	38.0	0.08
43	MRI-ESM1			38.2	0.22
44	NorESM1-M	Norwegian Climate Centre	2.50° × 1.88°	43.0	0.21
45	NorESM1-ME			42.7	0.19
46	SMA	simple model averaging in this study	1.00° × 1.00°	41.9	0.19
47	BMA	Bayesian model averaging in this study	1.00° × 1.00°	39.7	0.18

interpolated to one degree using the bilinear interpolation method. Detailed information on the *CMIP5 GCMs* is summarized in Table 1.

2.2. Reanalysis and satellite datasets

To investigate the impact of the downward surface solar shortwave radiation (R_s), near surface air temperature (T_a) and precipitation (P) on the variations in surface *LE* across different regions, reanalysis and satellite datasets are needed. We used the global monthly surface R_s products at a spatial resolution of one degree for 1984–2005, which were derived from the Global Energy and Water Cycle Experiment (*GEWEX*) Surface Radiation Budget (*SRB*) products because of its higher accuracy when compared to other satellite, reanalysis and *GCM* R_s products (Liang et al., 2010; Zhang et al., 2010a). To identify the land cover types, before 1999, the UMD Global Land Cover Classification (AVHRR: UMD CLCF: 1°) product (Hansen et al., 1998), which was generated from Advanced Very High Resolution Radiometer (AVHRR) satellites between 1981 and 1994, was used and after 1999, the Collection 4 *MODIS* land cover (*MOD12C1*: *CMG*, 0.05°) product (Friedl et al., 2002) was

used. The *MOD12C1* product was interpolated into one degree using nearest neighbor interpolation method. Considering that the European Centre for Medium Range Weather Forecasts (*ECMWF*) has the most accurate T_a among the available datasets (*NASA* Data Assimilation Office, *DAO*, and National Centers for Environmental Prediction Reanalysis, *NCEP*), the monthly T_a data for 1984–2005 were extracted from the *ERA-Interim* reanalysis product with a *4D* variation assimilation system at *T255* horizontal resolution, which were interpolated into one degree using bilinear interpolation (Simmons et al., 2006; Zhao et al., 2006). Monthly *P* datasets with 1° spatial resolution were extracted from Princeton Global Forcing (*PGF*) datasets, which are based on the Climatic Research Unit Time Series version 3.0 (*CRU TS*3.0) dataset with bias correction (Mitchell and Jones, 2005; Rodell et al., 2004; Sheffield et al., 2006). The *PGF P* data used in this study cover a period of 1984–2005 and supply useful climate information for global applications.

2.3. Observations from eddy covariance flux towers

To evaluate and validate the *BMA* method and all of the *CMIP5 GCM LE* simulations, comprehensive data of *LE* observations were collected at 240 *EC* flux tower sites provided by the La Thuile Flux data set (Baldocchi, 2008; Jung et al., 2011), the Coordinated Enhanced Observation Network of China (*CEOP*) for assessing terrestrial water budget over northern China (Jia et al., 2012; Liu et al., 2011; Xu et al., 2013) and the Chinese Ecosystem Research Network (*CERN*) for monitoring agricultural water use efficiency (Li et al., 2005). These flux tower sites spread across six continents (Asia, Europe, Africa, North America, South America and Australia) and encompass nine global land cover types: evergreen broadleaf forest (*EBF*; 16 sites), evergreen needleleaf forest (*ENF*; 64 sites), deciduous needleleaf forest (*DNF*; 6 sites), deciduous broadleaf forest (*DBF*; 28 sites), mixed forest (*MF*; 11 sites), shrubland (*SHR*; 14 sites), cropland (*CRO*; 34 sites), savanna (SAW; 10 sites), and grass and other types (*GRA*; 57 sites) (Fig. 1). The data covers a period of 2000–2009 with each site has at least one year of reliable data. The *EC* data are half-hourly observations, and the monthly data are aggregated from half-hourly or hourly data by site *PIs*. All then flux measurements were conducted based on the *EC* method (Baldocchi et al., 2001; Kaimal and Finnigan, 1994). Because of the energy imbalance problem, we used the method proposed of Twine et al. (2000) to correct the *LE* at all different flux tower sites.

3. Methods

3.1. Skillful score model

We used a Taylor skill score (*S*) (Taylor, 2001) to evaluate the skill of the performances of the *CMIP5 GCM LE* simulations and their ensemble predictions.

$$S = \frac{4(1+R)^4}{(\delta_f + 1/\delta_f)^2(1+R_{\max})^4} \tag{1}$$

where $R_{\max}$ is the maximum correlation coefficient that is set to 1.0 in this study, *R* is the correlation coefficient, and *R*-squared (R^2), is used to measure the degree of association between the simulated and observed *LE*, which is calculated as

$$R^2 = \frac{\left[\sum_{S=1}^{N}(x_s-\bar{x}_s)(x_s-\bar{x}_s)\right]^2}{\sum_{S=1}^{N}(x_s-\bar{x}_s)^2\sum_{o=1}^{N}(x_o-\bar{x}_o)^2} \tag{2}$$

Here, x_s is the monthly simulated *LE*, x_o is the monthly observed *LE*, *N* is the number of samples, $\bar{x}_s$ is the average of x_s and $\bar{x}_o$ is the average of x_o. δ_f is the normalized standard deviation of the *LE* simulations over the standard deviation of the corresponding *LE* observations:

$$\delta_f = \frac{\delta_s}{\delta_o} = \frac{\sqrt{\frac{1}{N}\sum_{s=1}^{N}(x_s-\bar{x}_s)^2}}{\sqrt{\frac{1}{N}\sum_{o=1}^{N}(x_o-\bar{x}_o)^2}} \tag{3}$$

The skill score varies from 0 (least skillful) to 1 (most skillful). R^2 characterizes the coherence between the simulated *LE* and the corresponding observations, and the normalized standard deviations reflect the differences between the simulated and observed magnitudes.

The root-mean-square error (*RMSE*) is another metric used to evaluate the predictive skill. It measures the closeness of the simulations and observations over a month and is expressed as

$$RMSE = \sqrt{\frac{1}{N}\sum_{s=1}^{N}(x_s-x_o)^2} \tag{4}$$

The mean bias is also a metric to assess the predictive skill, which implies the difference between the average simulation and observation.

$$Bias = \frac{1}{N}\sum_{s=1}^{N}(x_s-x_o) \tag{5}$$

3.2. Bayesian model averaging method

A Bayesian model averaging (*BMA*) method was used here to combine the selected *CMIP5 GCM LE* simulations and *EC* observations to improve global terrestrial *LE* simulation. The probability density function (*PDF*) of the *BMA* method for *LE* can be computed as a weighted average of the *PDFs* for each single dataset centered on the bias-corrected simulations (Duan and Phillips, 2010; Hoeting et al., 1999; Yao et al., 2014a). The *BMA* method characterizes the contributions of the individual *LE* datasets to the predicted *LE* based on *EC* observations. Based on the law of total probability, the predictive *PDF* of variable *LE* can be written as

$$p\left(LE|LE^T\right) = \sum_{i=1}^{M} p\left(LE|D_i, LE^T\right) p\left(D_i|LE^T\right) \tag{6}$$

where *M* is the number of the ensemble datasets or algorithms. $p\left(LE|D_i, LE^T\right)$ is the predictive *PDF* of the D_i dataset, calculated using the corresponding observations of LE^T. $p\left(D_i|LE^T\right)$ is the posterior probability of the D_i dataset corrected based on the corresponding LE^T. This term can be considered as a statistical weight W_i and can be expressed as

$$W_i = p\left(D_i|LE^T\right) = \frac{p\left(LE^T|D_i\right)p(D_i)}{\sum_{i=1}^{M} p\left(LE^T|D_i\right)p(D_i)} \tag{7}$$

The maximum likelihood is used to acquire W_i and the expectation maximization (*EM*) algorithm is chosen to calculate the maximum likelihood function (Dempster et al., 1977; Raftery et al., 2005). More details of the *BMA* method were described in Duan and Phillips (2010).

To assess the merged *LE* accuracy, we evaluated the performance of the *BMA* method based on a fivefold cross-validation method. This method divided the samples into five groups with roughly equal numbers (Jung et al., 2011). The *BMA*-based simulated *LE* for each of the five groups was independently validated using the samples of the remaining four groups. We also used Taylor diagrams to evaluate the performance of the *BMA* method to qualify the accuracy of the model outputs (Taylor, 2001). In Taylor diagrams,

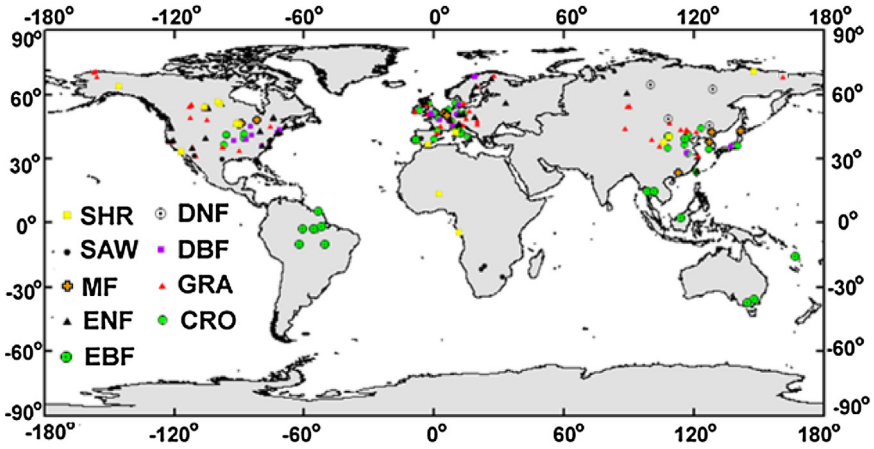

Fig. 1. Locations of the 240 sites used in this study.

the similarity between the simulated and observed *LE* is calculated using their *R* (the cosine of the azimuth angle), their centered root-mean-square difference (*RMSD*) (the radial distance from the observed point) and their standard deviations (the radial distance from the origin). Taylor diagrams are especially beneficial in assessing the relative skill of many different models (Intergovernmental Panel on Climate Change (IPCC) (2001) Climate Change, 2001).

3.3. Simple model averaging method

A simple model averaging (*SMA*) method was used in this study. In the *SMA* method, the weight for the individual dataset or algorithm is set to a constant of $1/M$and computed as

$$LE_{SMA} = \frac{1}{M}\sum_{i=1}^{M} LE_i \quad (8)$$

where LE_{SMA} and LE_i are simulated *LE* values using the *SMA* method and the individual dataset or algorithm, respectively.

3.4. Trend calculation

A simple linear trend equation was employed to detect the trend in the global terrestrial *LE* and other climate variables. The linear regression model is as follows:

$$f(x) = ax + b \quad (9)$$

Here, $f(x)$ is the annual value of *LE* or other climate variables, x is the year and a is the trend of the long-term annual *LE* or other climate variables. The confidence levels of the derived tendencies are computed using Student's *t*-test distribution with *n*-2 degrees of freedom (Pinker et al., 2005). The simple linear trend equation was also used to obtain the trend in the global *LE* and other climate variables pixel by pixel.

4. Results and discussion

4.1. Evaluation of CMIP5 GCM LE simulation

The objective of evaluating *CMIP5 GCM LE* simulations using substantial ground-based observations is to provide a general overview of the predictive ability of *GCMs* for different land cover types. Therefore, the simulated *LE* results at the original *CMIP5 GCM* grid scale were directly compared with ground-based observations. To reduce the uncertainties from the comparison between the simulations and observations, the average value of the observations for more than one site within one pixel was considered as a reference.

Fig. 2 shows the statistical results for each model for different land cover types. At the site scale, 45 *GCMs* illustrate substantial differences in the monthly *LE*. Almost all of the *GCMs* have the highest *S* (more than 0.54) and R^2 (more than 0.52, $p < 0.05$) with the lowest average *RMSEs*, less than 25 W/m^2 for all of the *DNF* sites, compared to those for the other land cover types. This may occur because a few samples for only 6 *DNF* sites artificially highlight the good performance of the *GCMs*. When we selected equal samples for all of the land cover types, most *GCMs* still provide better fits to the flux tower observations for *DNF* sites. Similarly, most *GCMs* exhibit high *S* (more than 0.51) and R^2 (more than 0.50) with a confidence level of $p < 0.05$ for *MF* sites, but the biases are higher relative to the *DNF* sites. This indicates that most *GCMs* still perform well for the variety of vegetation types.

For all of the *EBF* sites, most *GCMs* have poor *LE* modeling performance with an average *S* of less than 0.41, average R^2 of less than 0.40, and average *RMSE* of more than 32 W/m^2. This may be partially attributed to the uncertainty in the *EC* observations and by the smaller intra-annual variability in the tropics and least dependency on radiation input (simple to catch by models in comparison to other drivers) (Fisher et al., 2009; Jung et al., 2011). For all of the *CRO* sites, most *GCMs* have the second lowest *S* (less than 0.49) and R^2(less than 0.47, $p < 0.05$), and a high average *RMSE* of more than 30 W/m^2. This poor agreement may reflect the large differences in the *GCM* parameterizations and couplings for different land cover types. Fig. 2 also demonstrates that most *GCMs* illustrate high predictive skill with average *S* values above 0.45 for other biome (*DBF*, *ENF*, *SHR* and *SAW*) sites. For instance, *CMIP5* can satisfactorily reproduce monthly *LE* simulations for *SAW* sites, with a small average *RMSE* of 22 W/m^2 and high average R^2 of 0.43. This development in the physical model structures of the *CMIP5 GCMs* effectively accounts for the better performance in the *LE* simulations for these land cover types (Wild et al., 1998; Wild et al., 2001; Wild and Roeckner, 2006).

Overall, most *CMIP5 GCMs* overestimate *LE* to some extent, and the bias in the *GCMs*' *LE* simulations deviating from ground observations for all sites varies from 2 W/m^2 to 18 W/m^2. Among all of the *GCMs*, *CESM1-CAM5* has the best performance, with the highest *S* (0.51–0.75) and R^2 (0.46–0.74, $p < 0.01$) and the lowest *RMSE* (16.1–30.2 W/m^2) for different land cover types, followed by *NorESM1* (*NorESM1-M* and *NorESM1-ME*), *IPSL-CM5* (*IPSL-CM5A-LR*, *IPSL-CM5A-MR* and *IPSL-CM5B-LR*) and *EC-EARTH*, which have *S* values above 0.55 and *RMSEs* less than 27 W/m^2. Compared to the other *GCMs*, the improvements in a number of parameterizations and new components (*e.g.,* indirect aerosol effect) in these models provide a more realistic model to improve the simulation of *LE* (Meehl et al., 2013). However, *FGOALS-g2* demonstrates the lowest

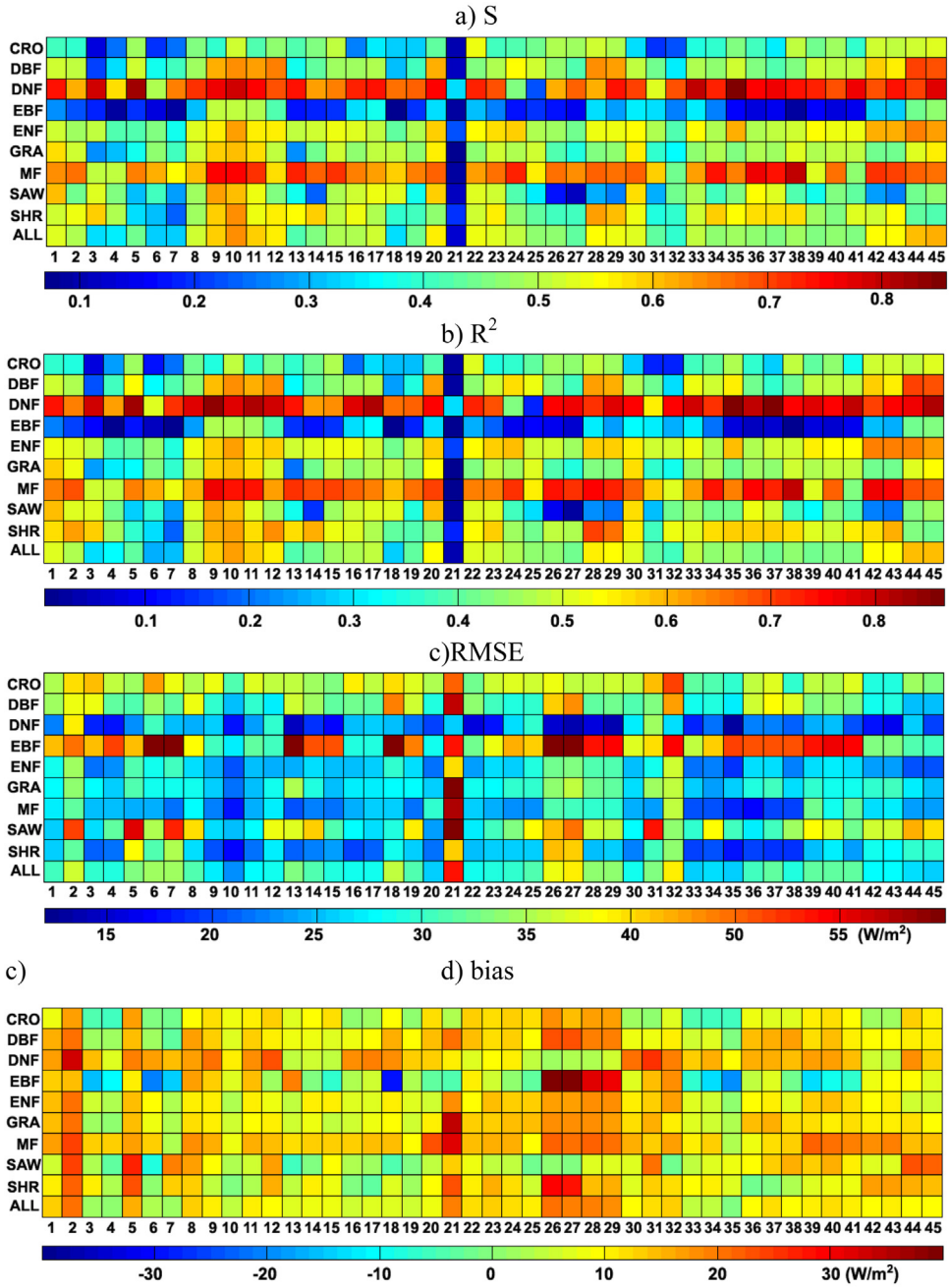

Fig. 2. Diagrams of the statistics (*S*, R^2, *RMSE* and bias) of the comparison between the *LE* simulations from the 45 *GCMs* in *CMIP5* and ground-measurements for different land cover types.

performance over the majority of flux tower sites, likely due to its sensitivity to the parameterization of resistances. Similar conclusions can be drawn for the annual *LE* simulations from the 45 *GCMs* (Fig. 3).

When compared to the satellite and reanalysis *LE* products, the *CMIP5 GCMs* have lower R^2 and larger *RMSEs*. We found that the average R^2 between the *CMIP5 GCMs* and ground observations is approximately 0.50 for all sites, which is obviously lower than the value above 0.60 for the *MODIS* and *ERA-Interim LE* products, which have a 95% level of confidence. This indicates that the *CMIP5 GCMs*' *LE* has a lower accuracy than those from the satellite and reanalysis datasets. Satellites provide the leaf area index (*LAI*) and land cover. Reanalysis datasets generate air temperatures and vapor pressure deficits (*VPD*) by assimilating meteorological observations and other auxiliary data. Therefore, the *CMIP5 GCMs*' *LE* simulations, which completely rely on these variables provided by satellite and reanalysis datasets, have lower accuracy.

4.2. Global LE ensemble from CMIP5 climate models

4.2.1. Cross validation of the BMA method

To reduce the uncertainty in the individual datasets and improve the accuracy of the *LE* simulations, we used the *BMA* method to combine seven *CMIP5 GCMs* (*CESM1-CAM5*, *NorESM1-M*, *NorESM1-ME*, *IPSL-CM5A-LR*, *IPSL-CM5A-MR*, *IPSL-CM5B-LR* and *EC-EARTH*) with high accuracy and ground observations to simulate the global terrestrial *LE*.

At the site scale, we compared the ensemble *LE* using the *BMA* method with those using the *SMA* method and individual *GCMs*. Fig. 4 shows the comparison between the monthly *LE* observations

and *BMA* estimates using a fivefold cross-validation method for each land cover type. It is clear that the *BMA*-based *LE* simulations for different land cover types have lower *RMSEs* and higher R^2 (95% confidence) compared to the *SMA* method and single *GCMs* at all of the flux tower sites. For the *SAW*, *SHR* and *GRA* sites, the *BMA* method has better performance than the *SMA* method and indi-

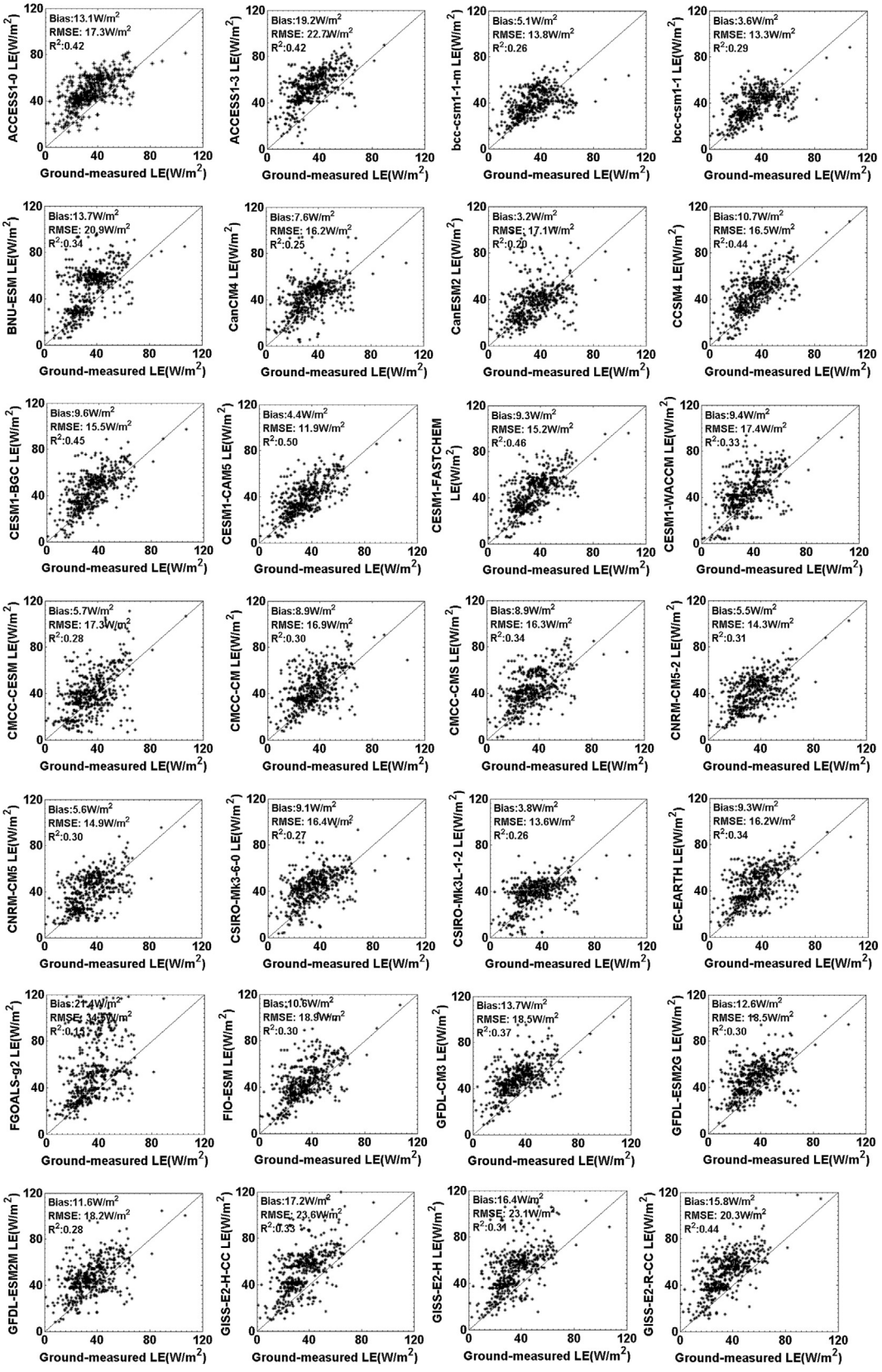

Fig. 3. Comparison of the annual LE observations for all 240 sites and the corresponding *LE* simulations from the 45 *GCMs* in *CMIP5*.

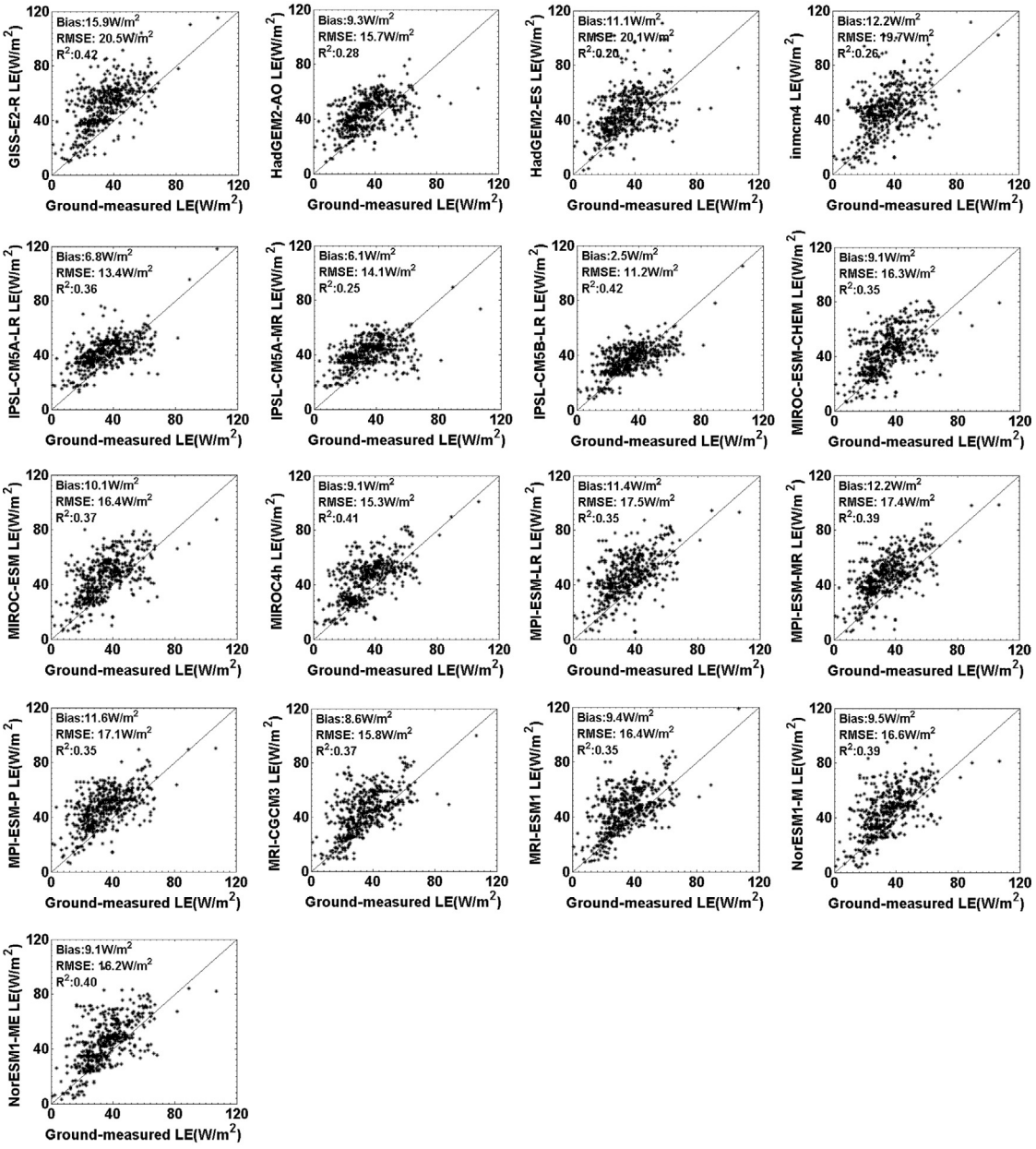

Fig. 3. (*Continued*)

vidual *GCMs*, with lower *RMSEs* (less than 21 W/m^2) and higher R^2 (more than 0.62, $p < 0.01$). For all of the *CRO* sites, the *BMA* method also has a higher R^2, 0.50 (with 99% confidence), and a lower *RMSE*, 24 W/m^2, than the *SMA* method and individual *GCMs*, though it does not present the best performance. Similarly, for all of the *EBF* sites, the *RMSE* of the *BMA*-based *LE* versus ground observations is approximately 26 W/m^2 and the R^2 is approximately 0.44 ($p < 0.01$), but it still showed better performance than the *SMA* method and individual *GCMs*. For other forest sites, the estimated *LE* using the *BMA* method exhibited the lowest *RMSE*, less than 18 W/m^2, and highest R^2, above 0.65 ($p < 0.01$), compared to the *SMA* method and individual *GCMs*. Overall, compared to the other methods, the *BMA* method decreased the *RMSE* by approximately 3 W/m^2 for crop and *EBF* sites and approximately 5 W/m^2 for most forest, savanna, shrub and grass sites and increased the R^2 by more than 0.03 ($p < 0.05$) for most flux tower sites. The good performance of the *BMA* method is mainly attributed to the fact that *BMA* allows weighing by correcting biases of the multiple models to closely match surface *EC* observations.

Fig. 5 compared the annual observed and simulated *LE* by the *BMA* method and *SMA* method. The results showed that the *BMA* method has the best performance, with the highest R^2, 0.58 (99% confidence), and the lowest *RMSE*, 10.5 W/m^2, compared to the *SMA* method. Previous studies found that some *GCMs* tend to overestimate *LE* due to excessive moist advection produced by intense zonal flow from the ocean to interior land (Sheppard and Wild, 2002). The *BMA* method substantially decreased the errors of the *GCMs* through the adjustment of the weights of the individual *GCMs* and incorporating ground-measured *EC* observations.

4.2.2. *Implementation of global terrestrial* LE *ensemble*

To implement global terrestrial *LE* simulations with relatively higher accuracy, we derived the weights of the *BMA* method based on all of the *EC* observations and the seven best *CMIP5 GCM LE* datasets described in Section 4.2.1. Fig. 6 shows the

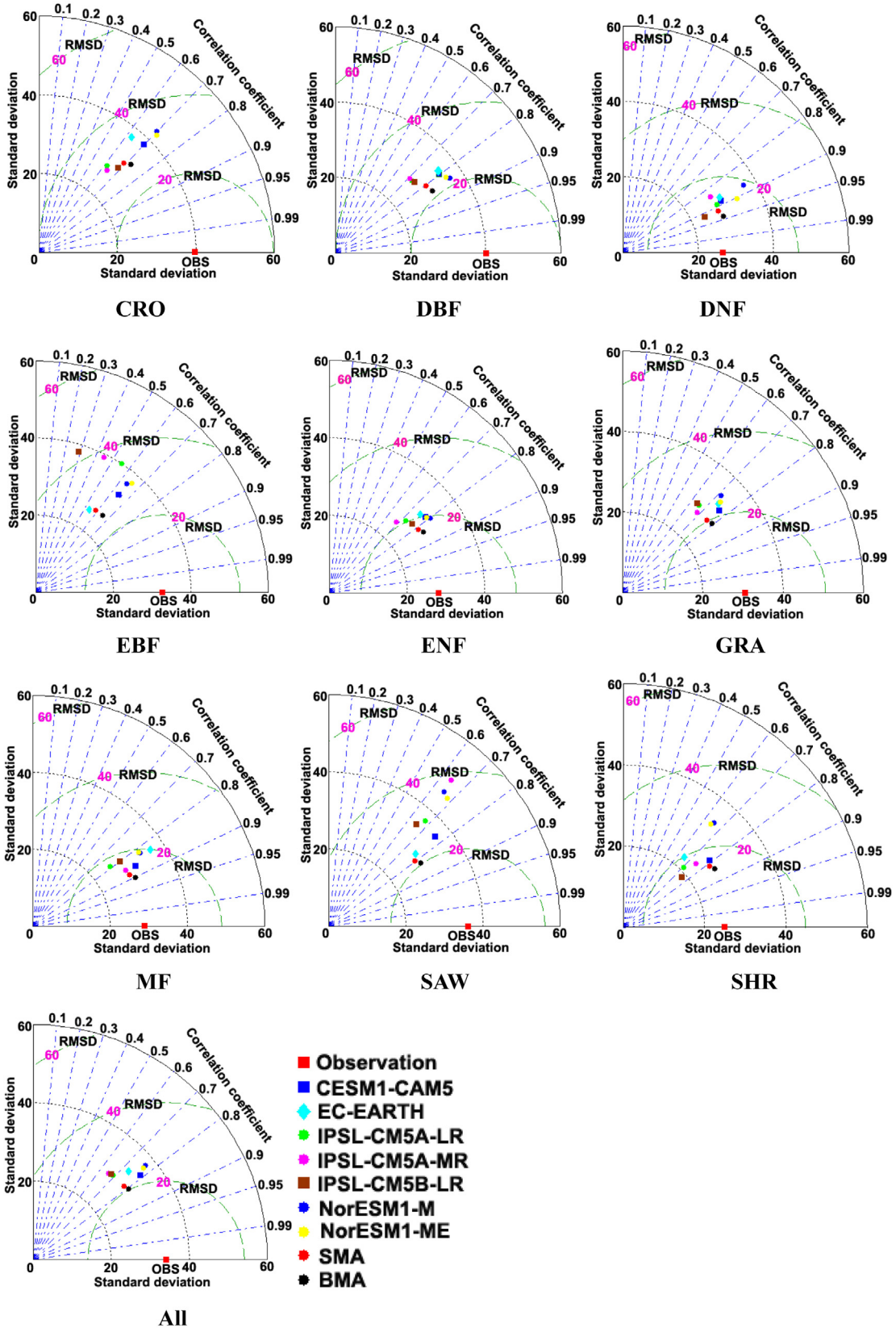

Fig. 4. Taylor diagrams for the monthly *LE* observations and *BMA* estimates using a fivefold cross-validation method for each land cover type.

weights of different *CMIP5 GCM LE* datasets when merging the *LE*. The relative contributions vary for different *GCM LE* datasets. The greatest contributor to the ensemble *LE* is *CESM1-CAM5*, contributing approximately 17.7%, followed by *NorESM1-M* (14.2%), *IPSL-CM5A-LR* (14.0%), *IPSL-CM5B-LR* (13.8%) and *EC-EARTH* (13.8%). The cross-validation also illustrated that *CESM1-CAM5* has the most

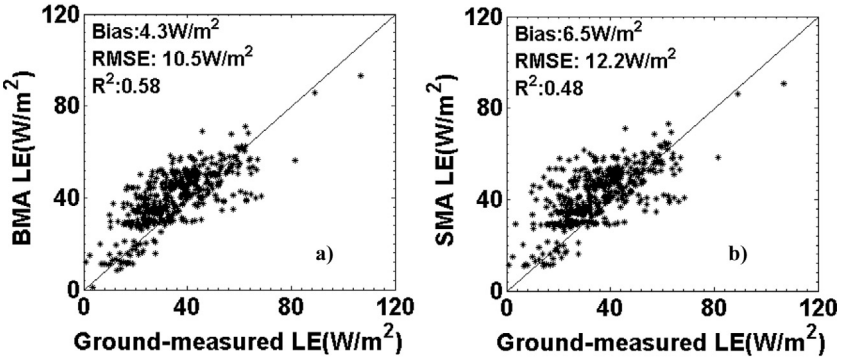

Fig. 5. Comparison of the annual observed and simulated *LE* by the (a) *BMA* method and (b) *SMA* method, respectively.

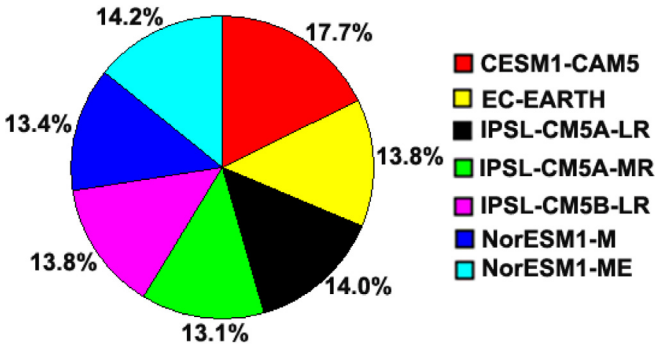

Fig. 6. Weights for the different *CMIP5 GCM LE* datasets for the merged *LE*.

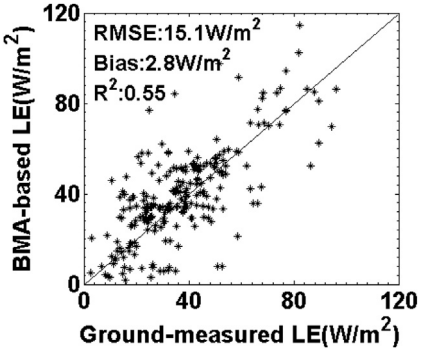

Fig. 7. Comparisons of the *BMA*-based simulated and measured site-averaged monthly *LE* at all sites.

accurate *LE* simulation among the 45 *GCMs*. Therefore, its contribution to the *BMA LE* simulation is greater than that of the other *GCMs*. *IPSL-CM5A-MR* contributes only 13.1% to the merged dataset due its relatively larger errors.

To evaluate the performance of the *BMA* method when predicting global spatial variations in *LE*, we compared the site-averaged *LE* between the ground observations and simulations based on the *BMA* method. Because the 240 *EC* flux tower sites are globally distributed across different continents and different land cover types, comparisons of the site-averaged *LE* can be used to test the performance of the *BMA* method to simulate global spatial variations in *LE*. In Fig. 7, the *RMSE* between the site-averaged ground-measured and simulated *LE* is 15.1 W/m^2 and the corresponding bias is 2.8 W/m^2. The R^2 is approximately 0.55 with 95% confidence. The relatively high accuracy of the among-site variability in the *LE* reflects the good ability of the *BMA* method driven by different *CMIP5 GCM LE* datasets to simulate terrestrial *LE* at the global scale.

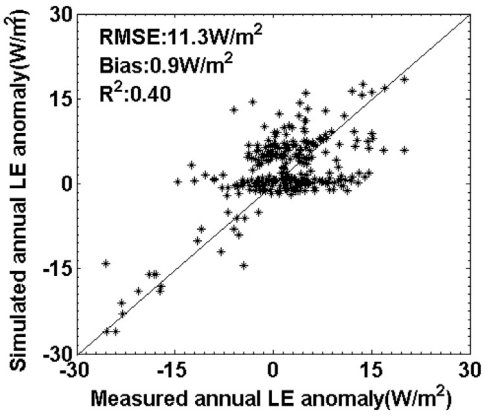

Fig. 8. Comparison of the annual anomalies of the *BMA*-based simulated *LE* and ground measured *LE* from sites with 5 or more years of available data.

To evaluate the ability of the *BMA* method to detecting the long-term variations in the global terrestrial *LE*, we also compared the annual *LE* anomaly for every site between the ground observations and simulations based on the *BMA* method. We only chose the flux tower sites with five or more years of available data. As shown in Fig. 8, the R^2 between the ground-measured and simulated annual *LE* anomaly is 0.40 and the corresponding *RMSE* is 11.3 W/m^2. Overall, the change in the annual *LE* is slightly higher than the observed value, partially due to the missing measured *EC* data. However, the *BMA* method in this study still captures the inter-annual variation in *LE*.

4.3. Spatiotemporal variability in the global terrestrial LE

4.3.1. Spatial distribution of the global terrestrial LE

The *BMA* method driven by ground observations and seven *CMIP5 GCMs* described in Section 4.2.1 was applied to produce global terrestrial *LE* with a spatial resolution of 1° during 1970–2005, as shown in Fig. 9. There are great differences in the spatial distribution of the global terrestrial *LE*: the smallest annual *LE* occurs in the arid and semi-arid regions of temperate climate zones and ice regions in the Arctic, whereas the largest annual *LE* occurs in tropical regions such as the Afrotropical, the Indomalayan and the Neotropical rainforest realms. The annual average global terrestrial *LE* (excluding Antarctica) between 1970 and 2005 estimated by the *BMA* method is approximately 39.7 W/m^2, which is lower than that of the *SMA* method, or 41.9 W/m^2. For different biomes, the highest average *LE*, 84.6 W/m^2, occurs in *EBF*, and *DNF* has the lowest average *LE* of 22.9 W/m^2. Other biomes fall within the range of *EBF* and *DNF* (*SAW*: 64.2 W/m^2, *DBF*: 58.1 W/m^2, *CRO*: 45.8 W/m^2, *MF*: 40.6 W/m^2, *GRA*: 38.4 W/m^2, *ENF*: 29.6 W/m^2

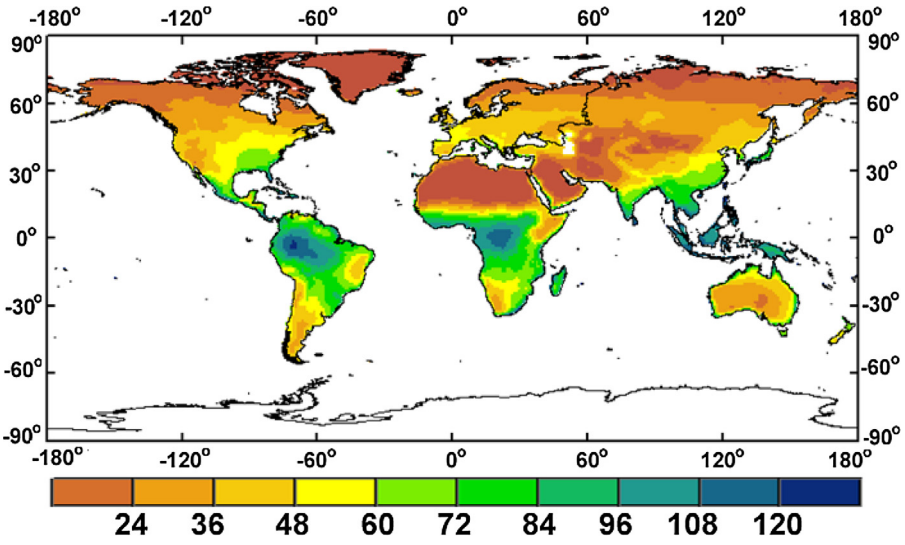

Fig. 9. Spatial distribution of the annual global terrestrial *LE* averaged for 1970–2005 at a spatial resolution of 1° from the *BMA* method driven by seven *CMIP5 GCMs*. Unit: W/m^2.

and *SHR*: 24.9 W/m^2). In general, annual average global terrestrial *LE* derived from multiple datasets varies from 34.1 W/m^2 to 42.7 W/m^2 (Jiménez et al., 2011; Mueller et al., 2011; Wang and Dickinson, 2012). Although many *GCMs* overestimate the annual average global terrestrial *LE* (Table 1), the *BMA* method yields a reasonable result by reducing the uncertainty in the individual *GCMs*.

Fig. 10 shows the great spatial differences in annual global terrestrial *LE* between the *BMA* method and other models. For example, the merged annual *LE* using the *BMA* method is lower in North Africa than those from *CESM1-CAM5* and *EC-EARTH*. Compared to *IPSL-CM5A-LR*, *IPSL-CM5A-MR* and *IPSL-CM5B-LR*, the *BMA* method has higher annual global terrestrial *LE* in the Southern Hemisphere, though the value is lower relative to both *NorESM1-M* and *NorESM1-ME*. Overall, the *BMA* method shows slightly lower annual global terrestrial *LE* compared to the *SMA* method. These spatial dissimilarities may be mainly explained by the weights of the *BMA* method, which incorporate a priori knowledge to correct the bias of the individual datasets (Miao et al., 2013; Wu et al., 2012).

4.3.2. Decadal variations in global terrestrial LE

Although the increasing trends in the *CMIP5 GCM LE* simulations differ greatly (Table 1), the global annual terrestrial *LE* ensemble based on the *BMA* method driven by ground observations and seven *CMIP5 GCMs* increased on average during 1970–2005 with a linear slope of 0.018 W/m^2 yr^{-1} ($p < 0.05$), which is comparable to other studies (Chen et al., 2014; Jung et al., 2010; Zeng et al., 2014) (Fig. 11). This variation is associated with global warming which accelerates the global terrestrial hydrological cycle (Douville et al., 2013; Huntington, 2006; Zeng et al., 2014). Some previous studies documented that the annual global terrestrial *LE* has declined since 1998 and may be attributed to the limitation of soil moisture (*SM*) (Jung et al., 2010; Zeng et al., 2014). However, we found that the merged *LE* increased at the rate of more than 0.01 W/m^2 yr^{-1} ($p > 0.05$) without stopping after 1998, which may be caused by the combined effects of the variations of R_s, T_a, and P on *LE* for different models on these time scales. However, other methods, such as the model tree ensemble (*MTE*) proposed by Jung et al. (2010), ignored the differences between the long-term variations in R_s and T_a by replacing R_s with T_a. In the Northern Hemisphere, there is a significant *LE* increase (0.017 W/m^2 yr^{-1}, $p < 0.05$) between 1970 and 2005. However, in the Southern Hemisphere, there is a no significant increase (0.024 W/m^2 yr^{-1}, $p > 0.05$) during this period. It is clear that the global trend of the *CMIP5 GCM LE* is in line with those from satellite and reanalysis datasets, including *GEWEX-PT* (Yao et al., 2014b), *ERA-Interim*, and *NASA*'s Modern-Era Retrospective Analysis for Research and Applications (*MERRA*) (Wang and Dickinson, 2012). This good agreement is likely because the satellite, reanalysis and *CMIP5* models all use T_a and R_s, which are associated with the acceleration of the global hydrological cycle, as inputs to produce the observed global warming trend (Ma et al., 2014; Simmons et al., 2010).

Fig. 12 shows the spatial pattern of the trends in the global terrestrial *LE* during 1970–2005. More than 66% of the pixels (42% with 95% confidence) show increasing *LE* trends. The largest increasing *LE* trend appears in Europe, eastern North America, central Asia, northern Australia, central and eastern Africa, and the southern regions of South America. In contrast, a widespread decreasing trend occurs in southeastern Asia, western North America, western Asia, southern Africa and eastern South America. To detect the attributions of the variations in the *LE*, the spatial distributions of the trends in the global terrestrial *LE*, T_a, R_s and P between 1984 and 2005 were examined (Fig. 13). In boreal regions, especially in Europe, the spatial pattern of the *LE* trend is almost consistent with that of the trend for T_a. This might originate from the limitation of T_a in high-latitude boreal ecosystems because T_a is the most important parameter in determining *LE* and the variation in *LE* is closely related to the variation in T_a (Nemani et al., 2003). Fig. 13 also shows the similar spatial patterns between the *LE* and *P* trends in semi-arid and arid regions. We should note that *P* is recognized to be a contributor to the *LE* in deserts, where vegetation growth is restricted by the scarce *P* and underground water (Ferguson and Veizer, 2007; Nemani et al., 2003). We also find a strong spatial coherence between the R_s trend changes and *LE* trend changes ($R^2 = 0.56$, $p < 0.05$) in tropical regions (excluding central Africa) over the past 27 years (Fig. 13). This coherence may be explained by the fact that solar radiation plays a dominant role in controlling vegetation growth and increases transpiration through stomatal opening in tropical regions, where water availability is not limiting and fewer clouds during dry periods allow more sunlight to reach the surface (Myneni et al., 2007; Nepstad et al., 1994). However, in central Africa, sparse ground observations impact the correction of the satellite-retrieved R_s, leading to the striking difference between the *LE* and R_s trends.

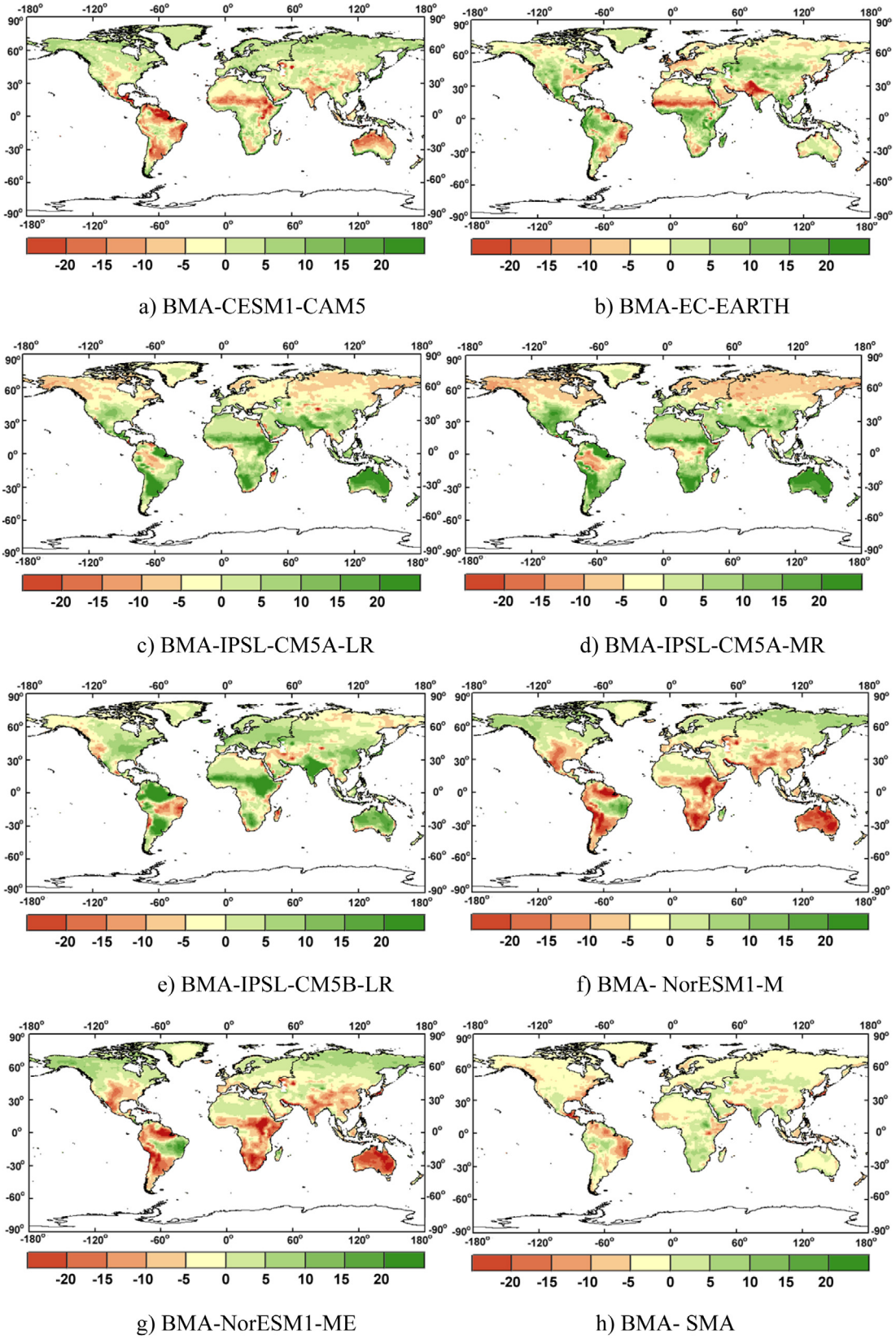

Fig. 10. Spatial differences in the average annual global terrestrial *LE* (1970–2005) between the *BMA* method and other models. Unit: W/m^2.

4.4. Discussion

4.4.1. Uncertainties in evaluating and merging the CMIP5 GCM LE simulations

Evaluating and merging the *CMIP5 GCM LE* simulations using only *EC* measured values will lead to substantial uncertainties, which are discussed below. First, the various uncertainties from the *EC* observations influence the accuracy assessment of the *CMIP5 GCM LE* simulation evaluation and fusion. Although the *EC* measurements are relatively accurate for measuring *LE*, their typical

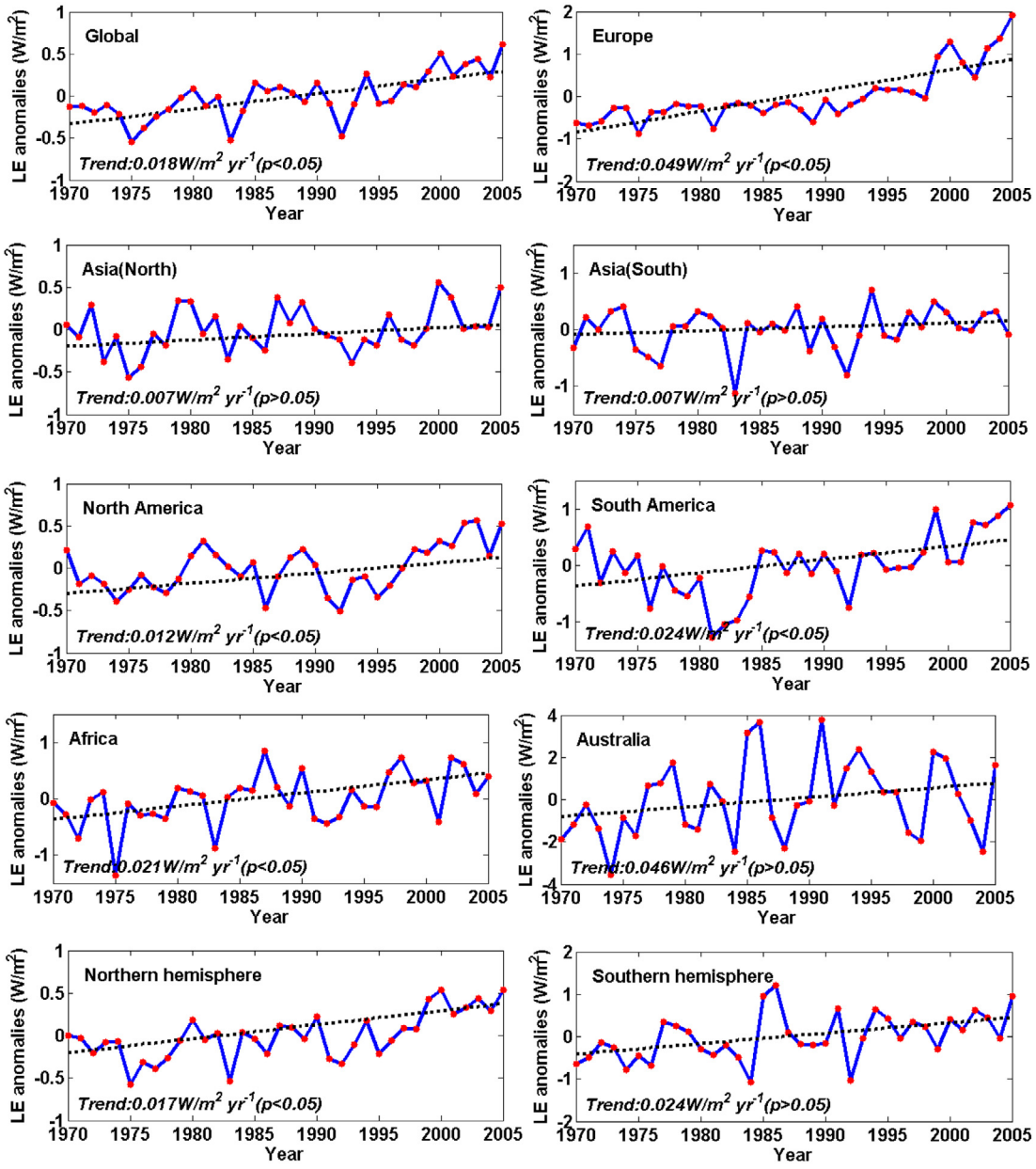

Fig. 11. Regional and global land surface averaged annual *LE* anomalies. The dashed lines refer to the linear trends in the merged *LE* based on the *BMA* method. Unit: W/m^2 per year.

measured errors are approximately 5–20% and the interpretation of their ambiguous values is still required (Foken, 2008; Mahrt, 2010; Wang et al., 2010a; Wang et al., 2010b). The problem is that the energy imbalance in the *EC* method and, generally, the average energy closure ratio ($(LE+H)/(R_n-G)$) for most *FLUXNET* sites is approximately 0.8 (Wilson et al., 2002). Foken (2008) noted that the *EC* method may only measure small eddies, while it cannot capture large eddies in the lower boundary layer, which also contribute to the energy balance. Many approaches, such as frequency correction (Moore, 1986) help to overcome some of the measurement insufficiencies, yet it has been suggested that Bowen ratio (*BR*) be preserved due to our limited understanding of the nature of the energy imbalance (Foken, 2008). Although we used the method proposed by Twine et al. (2000) to correct the *LE* in this study, these corrections still substantially increase the uncertainties of the *EC* ground measured *LE* (Finnigan et al., 2003; Sakai et al., 2001). This leads to large uncertainties when evaluating and merging *CMIP5 GCM LE* simulations.

A second factor is the spatial mismatch between the flux tower site footprints and *CMIP5 GCM* gridded footprints. Generally, the *EC* sites have a footprint of several hundreds meters while the spatial resolution of the *CMIP5 GCM* gridded datasets is more than 700 km (Li et al., 2009; McCabe and Wood, 2006; Zhang et al., 2010b). Thus, the *LE* values of the flux tower sites' footprint cannot represent the *CMIP5 GCM* gridded *LE*. However, in this study, the *EC* ground-measured *LE* was used to characterize the "true" value to evaluate and merge the *CMIP5 GCM* gridded datasets. Such an inaccurate representation will result in large differences between the *LE* observations and the *CMIP5 GCM* gridded datasets.

Third, we notice that the simulation biases in the *CMIP5 GCMs* themselves, misclassification of satellite land cover products and errors that propagate through the resample scale also contribute to the uncertainties in *LE* fusion based on the *BMA* method. Specif-

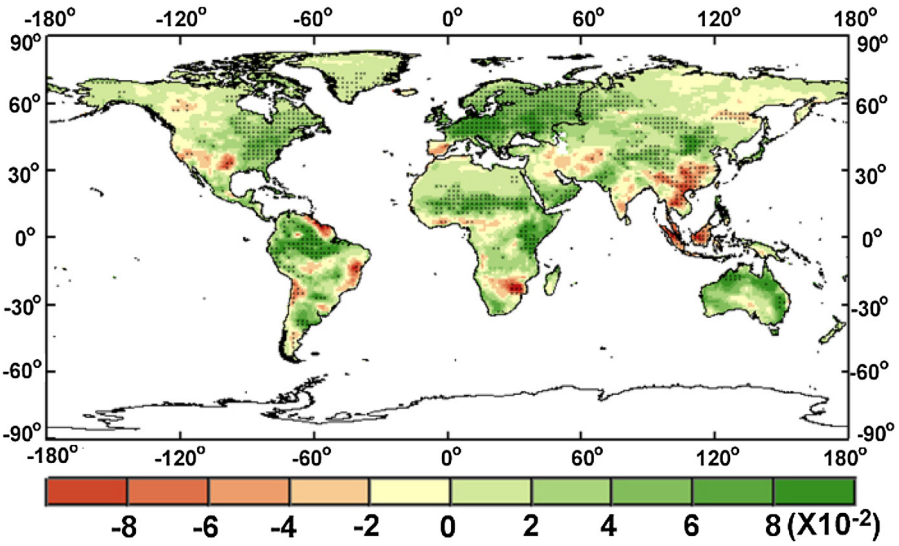

Fig. 12. Map of the linear trend in the merged *LE* based on the *BMA* method driven by seven *CMIP5 GCM* datasets during 1970–2005. The solid dots refer to grids with 95% confidence. Unit: W/m^2 per year.

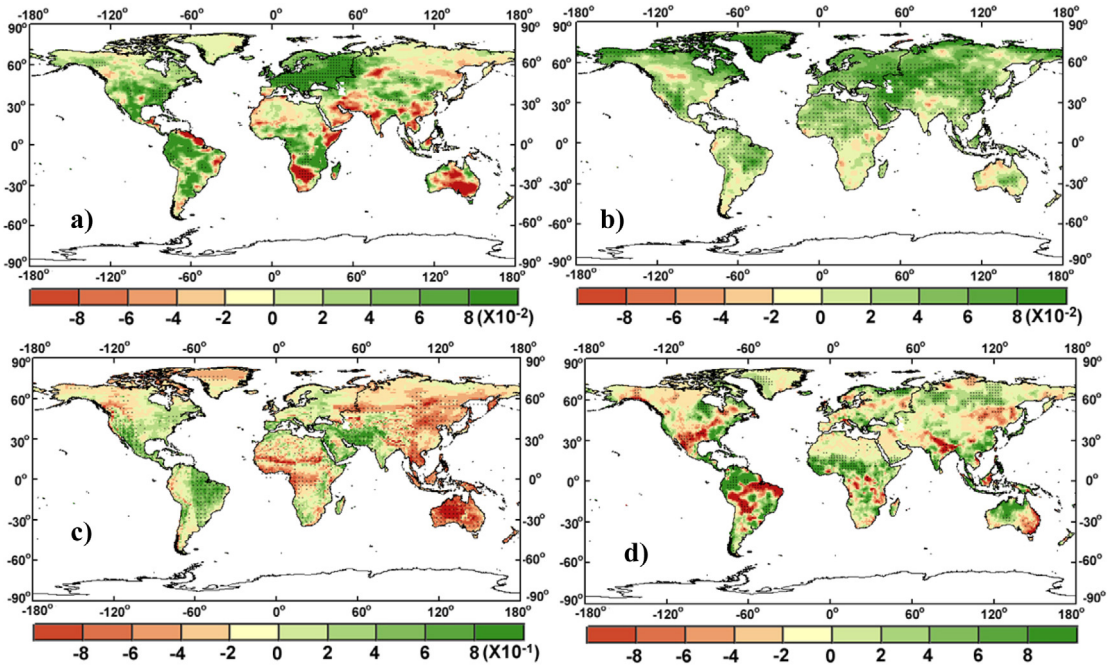

Fig. 13. Maps of the linear trends in the merged (a) *LE* based on the *BMA* method driven by seven *CMIP5 GCM* datasets (Unit: W/m^2 per year), (b) T_a from *ERA-Interim* (Unit: °C per year), c) R_s from *GEWEX* (Unit: W/m^2 per year) and d) *P* from *PGF*(Unit: mm per year) during 1984–2005. The solid dots refer to grids with 95% confidence.

ically, the errors from the monthly *LE* values for the *CMIP5* models cause the input errors of the merged *LE* due to the different *LE* algorithms, physical structures and different vegetation structure, such as the physical parameterizations of the Penman-Monteith (*PM*) equation for different vegetation types, influencing its coupling with atmosphere (Dirmeyer et al., 2013). Previous studies reported that most *CMIP5* climate models had ignored the negative feedbacks between water vapor and clouds, which also indirectly affect the biases of *LE* simulations (Christy et al., 2010; Miao et al., 2014). The accuracy of global terrestrial *LE* estimates is also highly dependent on the errors of the satellite land cover products. Previous studies showed that the accuracy of the *IGBP* layer of both *UMD CLCF* and *MCD12Q1* products are all less than 75% globally (Bartholome and Belward, 2005; Friedl et al., 2002; Hansen et al., 1998). Therefore, the low accuracy of classification also leads to the bias of *LE* ensemble. In addition, the bias of the merged *LE* introduced by the resample scale is a highlighted issue in terms of different *CMIP5 GCM LE* gridded datasets with different spatial resolutions. The *RMSE* of the gridded *LE* from multiple datasets was found to be variable among the various scales, and the improvement in the spatial resolution of *GCMs* influences the merged *LE* results despite its small impacts (Ma et al., 2014; Shi and Liang, 2014). This resample scale process could reduce the accuracy of the merged *LE* based on the *BMA* method.

Finally, we note that the selection of the conditional density function, $p\left(D_i|LE^T\right)$, accounts for the maximum impact on the accuracy of the *LE* fusion because it determines the weights of the *BMA* method. Here, we assumed $p\left(D_i|LE^T\right)$ meets a normal distribution

due to its success application in merging longwave radiation. In contrast, it is reasonable to select a gamma distribution when merging *P* products (Yang et al., 2012). However, *LE* is a complicated variable that couples energy and water cycles and accurately determining $p\left(D_i|LE^T\right)$ is a changeling scientific problem (Duan and Phillips, 2010; Yao et al., 2014a). Selecting the optimal $p\left(D_i|LE^T\right)$ is still considered an extension and elaboration of reliable fusion methods in the future.

4.4.2. Implications for understanding hydroclimate change

The identification of a *GCM*-based mechanism underlying the dynamics of global terrestrial *LE* has important implications for understanding hydroclimate change. Our results emphasize the necessity of deploying the *BMA* method and more observations to reduce the uncertainties in simulating *LE* by combining seven best *CMIP GCMs and* EC observations. For cases with substantial *GCMs*, the *BMA* method is the most suitable for *LE* estimation. *BMA*-based *CMIP5 GCM LE* simulations reconfirmed previous documents that stated that the global terrestrial *LE* has increased over the past three decades (Yao et al., 2012; Zeng et al., 2014). Wang et al. (2010b) attributed the change in this trend to the changes in cloudiness and aerosols in moist regions and the fluctuations of *P* in arid regions.

Despite the good spatial correlations between the *LE* and *P* trends in most regions (especially in arid regions), soil moisture, rather than *P* directly, determines the *LE*, and variations in *P* lead to variations in soil moisture. This will establish a positive feedback process between *LE* and *P*. For instance, large droughts caused by rare *P* will limit the soil moisture supply and reduce *LE*. On the opposite, *LE* positively affects *P* because most atmospheric precipitable water originates from both the evaporation of soil and water bodies, and vegetation transpiration over land surfaces (Eltahir and Bras, 1994; Findell et al., 2011; Jiang et al., 2015). However, this feedback process may be weak because many scientists attributed this rising global land *P* to increasing stratospheric aerosols (Fyfe et al., 2013; Zeng et al., 2014). This feedback process still remains unclear but still provides an example of the acceleration of the global water cycle due to global warming (Koster et al., 2003; Salvucci et al., 2002). Considering the *BMA* method is a data-driven tool and the contributions of each hydroclimate variable are not easily distinguished, the underlying mechanism and effects of hydroclimate variables on the spatiotemporal patterns of global terrestrial *LE* should be better understood by coupling *GCMs* and distributed hydrological models.

5. Summary and conclusions

This study has evaluated *LE* simulations of 45 *CMIP5* models by using globally distributed *FLUXNET EC* observations and then combining the seven best *CMIP5* models based on the *BMA* method. The comparison between the gridded *CMIP5 GCM LE* simulations versus ground observations demonstrated that almost all of the *CMIP5 GCMs* overestimate *LE* to some extent and present positive bias for all of the flux tower sites. Among these models, *CESM1-CAM5* has the best performance, and its *LE* simulation shows high predictive skill for different land cover types.

The *CMIP5 GCM LE* simulation using the *BMA* method was cross-validated, and the results showed that the *BMA* method has the best performance with the highest predictive skill and R^2 and the lowest *RMSE* compared to the *SMA* method and individual *GCMs*. We also validated the annual *LE* anomaly for every site and the among-site variability in the *LE* between ground observations and simulations based on the *BMA* method driven by different *CMIP5 GCM LE* datasets. Our results demonstrated that the merged *LE* using the *BMA* method simulates realistic spatial and inter-annual variations in *LE* at the global scale.

Based on the ground observations and seven best *CMIP5 GCMs*, the *BMA* method was used to simulate the global terrestrial *LE* with a spatial resolution of 1 ◦ between 1970 and 2005. Overall, we found that global annual terrestrial *LE* increased during the period 1970–2005. This increase may originate from the changes in T_a, R_s and *P* associated with global warming. Although consistent with other studies in general, we found large differences for the increasing trend after 1998 among models that are caused by the different input variables. The long-term effect of hydroclimatic variables on *LE* still needs to be determined by coupling *GCMs* and distributed hydrological models.

Acknowledgements

The authors thank the anonymous reviewers for their critical and helpful comments and suggestions. *CMIP5 GCM LE* simulation was downloaded from online (http://cmip-pcmdi.llnl.gov/cmip5/data_portal.html). Monthly *GEWEX* R_s product was obtained from online (http://gewex-srb.larc.nasa.gov/common/php/SRBdataproducts.php). Monthly *ERA- Interim* T_a dataset was downloaded from online (http://www.ecmwf.int/research/era/do/get/era-interim) and monthly *PGF P* product was obtained from online (http://hydrology. princeton.edu/data/pgf/1.0deg/monthly/). This work also used eddy covariance data acquired by the *FLUXNET* community and in particular by the following networks: AmeriFlux (*U.S.* Department of Energy, Biological and Environmental Research, Terrestrial Carbon Program (DE-FG02-04ER63917 and DE-FG02-04ER63911)), AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont,ChinaFlux, Fluxnet-Canada (supported by CFCAS, NSERC, BIOCAP, Environment Canada, and NRCan), GreenGrass, KoFlux, LBA, NECC, OzFlux, TCOS-Siberia, USCCC. We acknowledge the financial support to the eddy covariance data harmonization provided by CarboEuropeIP, FAO-GTOS-TCO, iLEAPS, Max Planck Institute for Biogeochemistry, National Science Foundation, University of Tuscia, Université Laval, Environment Canada and US Department of Energy and the database development and technical support from Berkeley Water Center, Lawrence Berkeley National Laboratory, Microsoft Research eScience, Oak Ridge National Laboratory, University of California-Berkeley and the University of Virginia. Other ground-measured data were obtained from the Coordinated Enhanced Observation Project (*CEOP*) in arid and semi-arid regions of northern China (http://observation.tea.ac.cn/). This work was partially supported by the Natural Science Fund of China (No. 41201331, No. 41331173 and No. 41205104), and the High-Tech Research and Development Program of China (No.2013AA122801). S. Wolf acknowledges support from a Marie Curie International Outgoing Fellowship (European Commission).

References

Baldocchi, D., et al., 2001. FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor and energy flux densities. Bull. Am. Meteorol. Soc. 82, 2415–2434.
Baldocchi, D., 2008. Breathing of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems. Aust. J. Bot. 56, 1–26.
Bartholome, E., Belward, A., 2005. GLC2000: a new approach to global land cover mapping from Earth observation data. Int J. Remote Sens. 6, 1959–1977.
Buser, C., et al., 2009. Bayesian multimodel projection of climate: bias assumptions and interannual variability. Clim. Dyn. 33, 849–868.
Chen, Y., et al., 2014. Comparison of satellite-based evapotranspiration models over terrestrial ecosystems in China. Remote Sens. Environ. 140, 279–293.
Christy, J.R., et al., 2010. What do observational datasets say about modeled tropospheric temperature trends since 1979? Remote Sens. 2, 2148–2169.
Covey, C., et al., 2003. An overview of results from the coupled model intercomparison project. Global Planet. Change 37, 103–133.
Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Ser. B 39, 1–38.

Dirmeyer, P.A., Jin, Y., Singh, B., Yan, X., 2013. Trends in land—atmosphere interactions from CMIP5 simulations. J. Hydrometeorol. 14, 829–849.
Douville, H., Ribes, A., Decharme, B., Alkama, R., Sheffield, J., 2013. Anthropogenic influence on multidecadal changes in reconstructed global evapotranspiration. Nat. Clim. Change 3, 59–62.
Duan, Q., et al., 2007. Multi-model ensemble hydrologic prediction using Bayesian model averaging. Adv. Water Resour. 30, 1371–1386.
Duan, Q., Phillips, T.J., 2010. Bayesian estimation of local signal and noise in multimodel simulations of climate change. J. Geophys. Res. 115, D18123.
Eltahir, E.A.B., Bras, R.L., 1994. Precipitation recycling in the Amazon basin. Q. J. R. Meteorol. Soc. 120, 861–880.
Ferguson, P.R., Veizer, J., 2007. Coupling of water and carbon fluxes via the terrestrial biosphere and its significance to the Earth's climate system. J. Geophys. Res. 112, D24S06.
Findell, K.L., Gentine, P., Lintner, B.R., Kerr, C., 2011. Probability of afternoon precipitation in eastern United States and Mexico enhanced by high evaporation. Nat. Geosci. 4, 434–439.
Finnigan, J.J., Clement, R., Malhi, Y., Leuning, R., Cleugh, H.A., 2003. A re-evaluation of long-term flux measurement techniques: Part I. Averaging and coordinate rotation. Bound. Layer Meteorol. 107, 1–48.
Fisher, J.B., et al., 2009. The land-atmosphere water flux in the tropics. Global Change Biol. 15, 2694–2714.
Foken, T., 2008. The energy balance closure problem: an overview. Ecol. Appl. 18, 1351–1367.
Friedl, M.A., et al., 2002. Global land cover mapping from MODIS: algorithms and early results. Remote Sens. Environ. 83, 287–302.
Fyfe, J.C., von Salzen, K., Cole, J.N.S., Gillett, N.P., Vernier, J.P., 2013. Surface response to stratospheric aerosol changes in a coupled atmosphere-ocean model. Geophys. Res. Lett. 40, 584–588.
Guilyardi, E., et al., 2013. Documenting climate models and their simulations. Bull. Am. Meteorol. Soc. 94, 623–627.
Hansen, M.C., DeFries, R., Townshend, J.R.G., Sohlberg, R., 1998. UMD Global Land Cover Classification, 1 Kilometer, 1.0, Department of Geography. University of Maryland, College Park, Maryland, pp. 1981–1994.
Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging: a tutorial. Stat. Sci. 14, 382–417.
Huntington, T.G., 2006. Evidence for intensification of the global water cycle: review and synthesis. J. Hydrol. 319, 83–95.
Intergovernmental Panel on Climate Change (IPCC) (2001) Climate Change, 2001. Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge.
Jia, Z., Liu, S., Xu, Z., Chen, Y., Zhu, M., 2012. Validation of remotely sensed evapotranspiration over the Hai River Basin, China. J. Geophys. Res. 117, D13113.
Jiang, B., Liang, S., Yuan, W., 2015. Observational evidence for impacts of vegetation change on local surface climate over northern China using the Granger causality. J. Geophys. Res. 120, 1–12.
Jiménez, C., et al., 2011. Global intercomparison of 12 land surface heat flux estimates. J. Geophys. Res. 116, D02102.
Jung, M., et al., 2010. Recent decline in the global land evapotranspiration trend due to limited moisture supply. Nature 467, 951–954.
Jung, M., et al., 2011. Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. J. Geophys. Res. 116, G00J07.
Kaimal, J.C., Finnigan, J.J., 1994. Atmospheric Boundary Layer Flows: Their Structure and Measurement. Oxford University Press, New York, pp. 289.
Koster, R.D., Suarez, M.J., Higgins, R.W., Van den Dool, H.M., 2003. Observational evidence that soil moisture variations affect precipitation. Geophys. Res. Lett. 30, 1241.
Lambert, S.J., Boer, G.J., 2001. CMIP1 evaluation and intercomparison of coupled climate models. Clim. Dyn. 17, 83–106.
Li, Y., Xu, H., Cohen, S., 2005. Long term hydraulic acclimation to soil texture and radiation load in cotton Plant. Cell Environ. 28, 492–499.
Li, Z., Tang, R., Wan, Z., Bi, Y., Zhou, C., Tang, B., Yan, G., Zhang, X., 2009. A review of current methodologies for regional evapotranspiration estimation from remotely sensed data. Sensors 9, 3801–3853.
Liang, S., Wang, K., Zhang, X., Wild, M., 2010. Review on estimation of land surface radiation and energy budgets from ground measurement, remote sensing and model simulations. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 3, 225–240.
Liu, S., et al., 2011. A comparison of eddy-covariance and large aperture scintillometer measurements with respect to the energy balance closure problem. Hydrol. Earth Syst. Sci. 15, 1291–1306.
Liu, S., et al., 2013. Measurements of evapotranspiration from eddy-covariance systems and large aperture scintillometers in the Hai River Basin. China J. Hydrol. 487, 24–38.
Ma, Q., Wang, K.C., Wild, M., 2014. Evaluations of atmospheric downward longwave radiation from 44 coupled general circulation models of CMIP5. J. Geophys. Res. 119, 4486–4497.
Mahrt, L., 2010. Computing turbulent fluxes near the surface: needed improvements. Agric. For. Meteorol. 150, 501–509.
McCabe, M.F., Wood, E.F., 2006. Scale influences on the remote estimation of evapotranspiration using multiple satellite sensors. Remote Sens. Environ. 105, 271–285.
Meehl, G., et al., 2013. Climate change projections in CESM1(CAM5) compared to CCSM4. J. Clim. 26, 6287–6308.
Miao, C., et al., 2014. Assessment of CMIP5 climate models and projected temperature changes over Northern Eurasia. Environ. Res. Lett. 9, 055007.
Miao, C., Duan, Q., Sun, Q., Li, J., 2013. Evaluation and application of Bayesian multi-model estimation in temperature simulations. Prog. Phys. Geogr. 37, 727–744.
Mitchell, T.D., Jones, P.D., 2005. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. Int. J. Climatol. 25, 693–712.
Moore, C.J., 1986. Frequency response corrections for eddy correlation systems. Bound. Layer Meteorol. 37, 17–35.
Moss, R.H., et al., 2010. The next generation of scenarios for climate change research and assessment. Nature 463, 747–756.
Mueller, B., et al., 2011. Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations. Geophys. Res. Lett. 38, L06402.
Myneni, R.B., et al., 2007. Large seasonal swings in leaf area of Amazon rainforests. Proc. Natl. Acad. Sci. U. S. A. 104, 4820–4823.
Nemani, R.R., Keeling, C.D., Hashimoto, H., Jolly, W.M., Piper, S.C., Tucker, C.J., Myneni, R.B., Running, S.W., 2003. Climate-driven increases in global terrestrial net primary production from 1982 to 1999. Science 300, 1560–1563.
Nepstad, D.C., et al., 1994. The role of deep roots in the hydrological and carbon cycles of Amazonian forests and pastures. Nature 372, 666–669.
Pinker, R.T., Zhang, B., Dutton, E.G., 2005. Do satellites detect trends in surface solar radiation? Science 308, 850–854.
Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005. Using Bayesian model averaging to calibrate forecast ensembles. Mon. Weather Rev. 133, 1155–1174.
Rodell, M., et al., 2004. The global land data assimilation system. Bull. Am. Meteorol. Soc. 85, 381–394.
Sakai, R.K., Fitzjarrald, D.R., Moore, K.E., 2001. Importance of low-frequency contributions to eddy fluxes observed over rough surfaces. J. Appl. Meteorol. 40, 2178–2192.
Salvucci, G.D., Saleem, J.A., Kaufmann, R., 2002. Investigating soil moisture feedbacks on precipitation with tests of Granger causality. Adv. Water Resour. 25, 1305–1312.
Sheffield, J., Goteti, G., Wood, E.F., 2006. Development of a 50-year high- resolution global dataset of meteorological forcings for land surface modeling. J. Clim. 19, 3088–3111.
Sheppard, R., Wild, M., 2002. Simulated turbulent fluxes over land from general circulation models and reanalysis compared with observations. Int. J. Climatol. 22, 1235–1247.
Shi, Q., Liang, S., 2014. Surface sensible and latent heat fluxes over the Tibetan Plateau from ground measurements reanalysis, and satellite data. Atmos. Chem. Phys. 14, 5659–5677.
Simmons, A., S. Uppala, D. Dee, and S. Kobayashi, 2006. ERA-Interim: New ECMWF reanalysis produces from 1989. ECMWF Newsletter, No.110, ECMWF, Reading, United Kingdom, pp. 25–35.
Simmons, A.J., Willett, K.M., Jones, P.D., Thorne, P.W., Dee, D.P., 2010. Low-frequency variations in surface atmospheric humidity, temperature, and precipitation: inferences from reanalysis and monthly gridded observational data sets. J. Geophys. Res. 115, D01110.
Sun, G., et al., 2011. Upscaling key ecosystem functions across the conterminous United States by a water-centric ecosystem model. J. Geophys. Res. 116, G00J05.
Taylor, K.E., Stouffer, R.J., Meehl, G.A., 2012. An overview of CMIP5 and the experiment design. Bull. Am. Meteorol. Soc. 93, 485–498.
Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res. 106, 7183–7192.
Twine, T.E., et al., 2000. Correcting eddy-covariance flux underestimates over a grassland. Agric. For. Meteorol. 103, 279–300.
Wang, K., Dickinson, R., 2012. A review of global terrestrial evapotranspiration: observation, modeling, climatology, and climatic variability. Rev. Geophys. 50, RG2005.
Wang, K., Dickinson, R.E., Wild, M., Liang, S., 2010a. Evidence for decadal variation in global terrestrial evapotranspiration between 1982 and 2002: 1. Model development. J. Geophys. Res. 115, D20112.
Wang, K., Dickinson, R.E., Wild, M., Liang, S., 2010b. Evidence for decadal variation in global terrestrial evapotranspiration between 1982 and 2002: 2. Results. J. Geophys. Res. 115, D20113.
Weigel, A.P., Liniger, M.A., Appenzeller, C., 2008. Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts? Q. J. R. Meteorol. Soc. 134, 241–260.
Wild, M., Roeckner, E., 2006. Radiative fluxes in the ECHAM5 general circulation model. J. Clim. 19, 3792–3809.
Wild, M., Ohmura, A., Gilgen, H., Roeckner, E., Giorgetta, M., Morcrette, J.J., 1998. The disposition of radiative energy in the global climate system: GCM-calculated versus observational estimates. Clim. Dyn. 14, 853–869.
Wild, M., Ohmura, A., Gilgen, H., Morcrette, J.J., Slingo, A., 2001. Evaluation of downward longwave radiation in general circulation models. J. Clim. 14, 3227–3239.
Wild, M., et al., 2013. The global energy balance from a surface perspective. Clim. Dyn. 40, 3107–3134.
Wild, M., et al., 2015. The energy balance over land and oceans: an assessment based on direct observations and CMIP5 climate models. Clim. Dyn. 44, 3393–3429.
Wilson, K., et al., 2002. Energy balance closure at FLUXNET sites. Agric. For. Meteorol. 113, 223–243.

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, D22107.
Xu, Z., S. Liu, X., Li, S., Shi, J., Wang, Z., Zhu, T., Xu, W. Wang, 2013. Intercomparison of surface energy flux measurement systems used during the HiWATER-MUSOEXE. J. Geophys. Res. Atmos. 118, 13140–13157.
Yang, C., Yan, Z., Shao, Y., 2012. Probabilistic precipitation forecasting based on ensemble output using generalized additive models and Bayesian model averaging. Acta Meteorol. Sin. 26, 1–12.
Yao, Y., Liang, S., Qin, Q., Wang, K., Liu, S., Zhao, S., 2012. Satellite detection of increases in global land surface evapotranspiration during 1984–2007. Int. J. Digital Earth 5, 299–318.
Yao, Y., et al., 2013. MODIS-driven estimation of terrestrial latent heat flux in China based on a modified Priestley-Taylor algorithm. Agric. For. Meteorol. 171–172, 187–202.
Yao, Y., et al., 2014a. Bayesian multimodel estimation of global terrestrial latent heat flux from eddy covariance, meteorological, and satellite observations. J. Geophys. Res. 119, 4521–4545.
Yao, Y., et al., 2014b. Estimation of the terrestrial water budget over northern China by merging multiple datasets. J. Hydrol. 519, 50–68.
Yao, Y., et al., 2015. A satellite-based hybrid algorithm to determine the Priestley-Taylor parameter for global terrestrial latent heat flux estimation across multiple biomes. Remote Sens. Environ. 165, 216–233.
Zeng, Z., et al., 2014. A worldwide analysis of spatiotemporal changes in water balance-based evapotranspiration from 1982 to 2009. J. Geophys. Res. 119, 1186–1202.
Zhang, X., Liang, S., Wang, K., Li, L., Gui, S., 2010a. Analysis of global land surface shortwave broadband albedo from multiple data sources. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 3, 296–305.
Zhang, K., Kimball, J.S., Nemani, R.R., Running, S.W., 2010b. A continuous satellite-derived global record of land surface evapotranspiration from 1983 to 2006. Water Resour. Res. 46, W09522.
Zhao, M., Running, S.W., Nemani, R.R., 2006. Sensitivity of Moderate Resolution Imaging Spectroradiometer (MODIS) terrestrial primary production to the accuracy of meteorological reanalysis. J. Geophys. Res. 111, G01002.