# An Empirical Orthogonal Function-Based Algorithm for Estimating Terrestrial Latent Heat Flux from Eddy Covariance, Meteorological and Satellite Observations

**Fei Feng**[1], **Xianglan Li**[1]*, **Yunjun Yao**[2]*, **Shunlin Liang**[2,3], **Jiquan Chen**[4], **Xiang Zhao**[2], **Kun Jia**[2], **Krisztina Pintér**[5,7], **J. Harry McCaughey**[6]

1 State Key Laboratory of Remote Sensing Science, College of Global Change and Earth System Science, Beijing Normal University, Beijing, 100875, China, 2 State Key Laboratory of Remote Sensing Science, School of Geography, Beijing Normal University, Beijing, 100875, China, 3 Department of Geographical Sciences, University of Maryland, College Park, MD, 20742, United States of America, 4 Landscape Ecology & Ecosystem Science (LEES) Lab, Center for Global Change and Earth Observations (CGCEO), Michigan State University, East Lansing, MI, 48823, United States of America, 5 Institute of Botany and Ecophysiology, Szent István University, 2100 Páter K.u.1., Gödöllő, Hungary, 6 Department of Geography, Queen's University, Mackintosh-Corry Hall, Room E112, Kingston, Ontario, Canada, 7 MTA-SZIE Plant Ecology Research Group, 2103, Gödöllő, Hungary

* xlli@bnu.edu.cn (XL); boyyunjun@163.com (YY)

## Abstract

Accurate estimation of latent heat flux (LE) based on remote sensing data is critical in characterizing terrestrial ecosystems and modeling land surface processes. Many LE products were released during the past few decades, but their quality might not meet the requirements in terms of data consistency and estimation accuracy. Merging multiple algorithms could be an effective way to improve the quality of existing LE products. In this paper, we present a data integration method based on modified empirical orthogonal function (EOF) analysis to integrate the Moderate Resolution Imaging Spectroradiometer (MODIS) LE product (MOD16) and the Priestley-Taylor LE algorithm of Jet Propulsion Laboratory (PT-JPL) estimate. Twenty-two eddy covariance (EC) sites with LE observation were chosen to evaluate our algorithm, showing that the proposed EOF fusion method was capable of integrating the two satellite data sets with improved consistency and reduced uncertainties. Further efforts were needed to evaluate and improve the proposed algorithm at larger spatial scales and time periods, and over different land cover types.

## 1. Introduction

Terrestrial latent heat flux (LE), the flux of heat from the Earth's surface to the atmosphere that is associated with soil evaporation and plant transpiration, and is a key component of the hydrological and carbon cycles [1, 2]. Accurate and temporally continuous estimation of LE is

critical for understanding the interactions between the land surface and the atmosphere and improving water use efficiency [3, 4].

Many LE products were developed at various temporal and spatial resolutions during the past several decades [5–11], which were needed to study long-term regional and global climate change [12]. Extensive evaluations of these products were conducted [13–20]. Chen et al. [4] compared eight evapotranspiration (ET) models (equivalent to LE) and found significant inconsistencies among the models, largely due to the driving factors. Long et al. [14] assessed the uncertainty in LE estimates from four land surface models, including two remote sensing-based products (MODIS and advanced very high resolution radiometer (AVHRR)), and Gravity Recovery and Climate Experiment (GRACE), by inferring ET from water budget, and found that uncertainty of remote sensing products was approximately 10–15 mm/month. Furthermore, spatial LE interactions were often ignored in most satellite-based LE models [21]. Ershadi et al. [21] used the surface energy balance system (SEBS) and Landsat images to investigate the effects of aggregation from fine (<100 m) to medium (~1 km) scales. From several common spatial interpolation algorithms, the simple average method preserved most accurately the spatial pattern of LE compared to the nearest neighbors and bilinear or bicubic interpolation methods.

Efforts were made to improve the quality of LE products by developing advanced LE retrieval algorithms [6, 9–11] and using data assimilation methods [22–29]. Data assimilation involved numerical models that incorporate measured data to produce final results for forecasting or analysis [30]. Caparrini et al. [23] used data assimilation to obtain LE, sensible heat and ground heat flux. Similar studies on surface temperature, sensible heat flux, and LE were also performed [25–27]. However, integrating advantages from existing LE products to improve data accuracy and integrity was the main goal. Previous studies showed that averaged LE was more accurate than individual LE models [20, 24]. Cammalleri et al. [22] combined multi-platform remote sensing thermal infrared data to estimate daily field-scale LE data using a spatial and temporal adaptive reflectance fusion model (STARFM). Yao et al. [28] combined five process-based LE algorithms using Bayesian averaging method. However, these methods usually failed to account for spatial and temporal correlations of LE when integrating satellite LE products.

In geosciences, empirical orthogonal function (EOF) method deals with both, temporal and spatial patterns [31, 32]. EOF was first used in meteorology to decompose a space-time field into spatial patterns and associated time indices. Incorporating both spatial and temporal correlations, Chen et al. [33] developed an extended EOF that became a powerful tool to extract dynamic structure, including trends, oscillations, propagating structures and to filter data. Smith et al. [34] used EOF analysis to solve the problem of missing data. Beckers and Rixen [35] developed a "self-consistent" and "parameter-free" EOF interpolation method, data interpolating EOF (DINEOF), which had proven useful for oceanographic data analysis [36]. Wang et al. [37] used hierarchical EOFs (HEOFs) to integrate LAI from MODIS and Carbon cycle and Change in Land Observational Products from an Ensemble of Satellites (CYCLOPES) to improve the quality of satellite based LAI data, which resulted in increase of $R^2$ (from 0.75 to 0.81) and in decrease of root mean square error (rmse) from 1.04 to 0.71.

In this paper, we propose an EOF-based data-fusion method that combines the major spatial and temporal patterns of different LE data to generate a consistent and high accuracy dataset. Errors in satellite based LE products might arise from the use of different driving factors or empirical coefficients. Therefore two process-based LE algorithms were selected to perform the data fusion: the MOD16 algorithm based on the PM approach [38] and the PT-JPL algorithm based on the PT approach [39]. The objectives of this study were to (1) compare the MOD16 and PT-JPL algorithms at FLUXNET sites; (2) evaluate the performance of the proposed EOF

fusion method by comparing it to MOD16, PT-JPL and a simple fusion method; and (3) assess the limitations of proposed method.

## 2. Materials and Methods

### 2.1 The principle of the empirical orthogonal function (EOF)

A fundamental advantage of the EOF-based method was to reconstruct the original data by minimizing the noise and the gaps. EOF incorporated principal component analysis (PCA), but also considered the temporal and spatial characteristics of the data [31]. LE data was stored in a $P \times N$ matrix ($A$), where $s \in [1, P]$ and represents space and $t \in [1, N]$ denotes time. Matrix $A$ was decomposed by singular value decomposition, which is a commonly used method in linear algebra:

$$A = ZSH^T \qquad (1)$$

where $Z$ stands for the left singular vectors (EOFs). $S$ for a diagonal matrix containing the singular values sorted in descending order, and $H$ for the right singular vectors (PCs). EOFs represents the spatial domain, whereas PCs represents the temporal domain. Thus, the spatial and temporal components were separated. Singular-value decomposition was also used to filter noise.

We expanded the EOF to emphasize temporal information [33] by subsetting A in time windows (W) and combining the subsets in a new matrix.

$$(a_1 \, a_2 \ldots a_N) \qquad (2)$$

denotes time series of LE at a specific location, from which a matrix was built using a window length of W,

$$A = \begin{pmatrix} a_1 & a_2 & \cdots & a_{N-W+1} \\ a_2 & a_3 & \ldots & a_{N-W+2} \\ \vdots & \vdots & \ddots & \vdots \\ a_W & a_{W+1} & \cdots & a_N \end{pmatrix} \qquad (3)$$

We applied this process to all spatial points, obtaining matrix $A'$ with dimension $W \times (N - W + 1)$. In this study, the EOF analysis was conducted with this matrix $A'$.

$$A' = \begin{pmatrix} A_{1,1} & A_{1,2} & \ldots & A_{1,N-W+1} \\ A_{1,2} & A_{1,3} & \ldots & A_{1,N-W+2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1,W} & A_{1,W+1} & \ldots & A_{1,N} \\ A_{2,1} & A_{2,2} & \ldots & A_{2,N-W+1} \\ \vdots & \vdots & \ddots & \vdots \\ A_{P,W} & A_{P,W+1} & \ldots & A_{P,N} \end{pmatrix} \qquad (4)$$

Traditional EOF analysis addresses matrices that contain no missing data. However, remote sensing data often does not satisfy this requirement, thus a modified EOF analysis (DINEOF) [35] was employed which uses iterative algorithms to estimate missing data. Before iteration, missing data were replaced with zeroes. Then, the following iterative algorithm was applied to the input matrix data with mean subtraction,

$$(X_a)_{ij} = (ZS_N H_N^T)_{ij} = \sum_{k=1}^{N} p_k (Z_k)_i (H_k^T)_j, (i,j) \in I \qquad (5)$$

where $i$ and $j$ are the location and time of the missing data, respectively; $(X_a)_{ij}$ is the reconstructed data using the leading components ($N$) of the data; and $P$ is the eigenvalue.

Traditional EOF analysis usually employes matrices that contain few spatial points, i.e., the images had coarse spatial resolution. When handling remote sensing data, the number of spatial point increases and the computational time becomes challenging. A modified HEOF was used to solve this problem. HEOFs [37] worked on two levels: coarse and fine-resolution. However, coarse resolution data also required considerable memory capacity. We simplified the original HEOF procedure by dividing the dataset into small subsets and applying an EOF to each of them. Because other subset information could not be used, we used the relative information.

## 2.2 The Framework of EOF fusion

Implementation of the EOF-based algorithm requires the following steps: (1) Forming the necessary matrix for EOF analysis from a time series of the satellite data ($A$). One year's data of MOD16 and PT-JPL was randomly selected. (2) Defining the number of leading components of each LE algorithm and the window length. (3) Intergration of MOD16 and PT-JPL output matrixes by their principal components.

We selected one year (2005) of data for MOD16 and PT-JPL, to test the proposed EOF method. The good overall performance of PT-JPL model was reported previously [19, 40, 41]. In the proposed EOF method 80% of the PT-JPL components and 20% of the MOD16 components were used. From PT-JPL the three leading components explaining about 80% of the total variance were selected (Fig 1) [42]. In the case of MOD16, leading components were selected
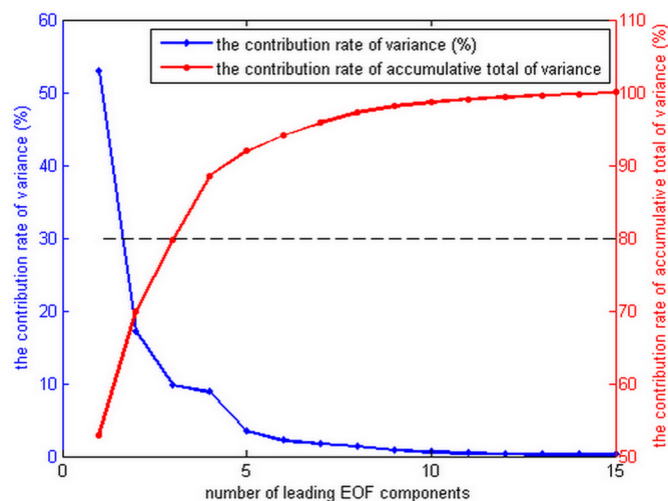


**Fig 1. Coefficient of determination of latent heat fluxes (PT-JPL model) as a function of the number of leading EOF components.** The left Y-axis is the contribution rate of covariance for each single EOF components (blue). The right Y-axis is the contribution rate of cumulative total of variance (red).

from the same position as in PT-JPL, which meant twelve leading components that explain about 20% of the total variance [42]. Low LE values of PT-JPL during spring and winter caused negative LE values during EOF reconstruction process, thus these values were replaced by the main patterns of MOD16. Choosing *W = 1* means that no temporal information is used and the expanded EOF method corresponds to the simple EOF method. To both reduce the computation time and emphasize temporal information the window size was set to 4. The EOF algorithm then was applied to a 200km × 200 km region for an entire year. For the purpose of validation, we included the domain around the FLUXNET sites, where each subset contained one FLUXNET site. We used a simple averaging (SA) model to integrate MOD16 and PT-JPL. The SA method was a simple fusion algorithm taking a constant weight (0.5) for each model. (S1 File)

## 3. Data and Analysis

### 3.1 Satellite Data

Satellite LE products included MODIS with 1 km spatial resolution, University of California Berkeley (UCB) with 0.5 degree spatial resolution, Global Land Evaporation: the Amsterdam Methodology (GLEAM) with 0.5 degree spatial resolution, Atmospheric water balance (AWB) with 2.5 degree spatial resolution, and University of Maryland (MAUNI) with 1 degree spatial resolution. The reanalysis products include the Global Modeling and Assimilation Office's (GMAO)- modern-era retrospective analysis for research and applications (MERRA) with ½ degree latitude × ⅔ degree longitude spatial resolution, interim ERA- the latest global atmospheric reanalysis produced by (ECMWF) with approximately 80 km degree spatial resolution, The National Centers for Atmospheric Prediction/National Center for Atmospheric Research (NCAR/NCEP) with 2.5 degree spatial resolution, MERRA-Land Reanalysis (M-LAND) with ½ degree latitude × ⅔ degree longitude spatial resolution. However, most global LE products had high uncertainties [2] and low resolution. Satellite data was used to estimate land surface variables, which were used as inputs for LE algorithm. Satellite data based LE algorithms were easy to operate for routine, long-term mapping of LE with different spatial scales. However, models structure and physical parameterizations of LE algorithms influenced the accuracy of these products. Dirmeyer et al., 2013 [43] found that model parameterizations in Penman-Monteith equation based LE algorithms influenced the accuracy assessment of LE.

The MODIS LE product algorithm (MOD16) was based on a beta version [5] developed from Cleugh et al. [44] using the PM model [38]. Mu et al. (2011) [11] improved the beta version by: (1) simplifying the calculation of the vegetation cover fraction with FPAR; (2) calculating LE as the sum of daytime and nighttime components; (3) improving calculations of aerodynamic, boundary-layer, and canopy resistance; (4) estimating the soil heat flux using available energy and simplified NDVI; (5) dividing the canopy into wet and dry components; (6) separating moist soil surfaces from saturated wet ones. The MOD16 algorithm was successfully extended to generate MODIS global terrestrial LE product from MODIS land cover, albedo, LAI/FPAR, and a GMAO daily meteorological reanalysis data set [11].

To avoid the complexity of parameterizing aerodynamic and surface resistance, Priestley and Taylor [45] reduced the atmospheric control term in the PM equation and added an empirical factor to design a simple LE algorithm. Based on this algorithm, Fisher et al. [6] proposed a novel PT-based LE algorithm (Priestley-Taylor LE algorithm of Jet Propulsion Laboratory, Caltech, PT-JPL) with atmospheric (RH and VPD) and ecophysiological constraints (FPAR and LAI) to downscale potential ET to actual ET. Total ET was the sum of canopy transpiration (ETc), soil evaporation (ETs) and interception evaporation (ETi). Each component

was calculated using the Priestley–Taylor equation and the corresponding ecophysiological condition.

MOD16 and PT-JPL methods were applied to estimate global terrestrial LE using the daily MERRA data sets with spatial resolution of 1/2 degree latitude × 2/3 degree longitude, 8 day MODIS FPAR/LAI (MOD15A2) product with 1-km spatial resolution, the 16 day MODIS NDVI (MOD13A2) product with 1-km spatial resolution, the 8 day Global Land Surface Satellite (GLASS) LAI product with 1-km spatial resolution [46], annual Land Cover Type product (MCD12Q1) with 500 m spatial resolution and the Shuttle Radar Topography Mission (SRTM30) Digital Elevation Model (DEM) elevation product. Considering the different spatial resolutions of the MERRA data, spatial interpolation based on the cosine [47] was used to match the spatial resolutions of MERRA and MODIS data.

## 3.2 Ground Measurements

The algorithm for EOF integration, MOD16 LE and PT-JPL LE was validated and evaluated using data of 22 EC towers provided by FLUXNET for 2005, as shown in Fig 2. The flux tower sites covered eight major global land-surface biomes: deciduous broadleaf forest (DBF; three sites), deciduous needleleaf forest (DNF; six sites), evergreen broadleaf forest (EBF; two sites), mixed forest (MF; one site), savanna (SAW; one site), shrubland (SHR; one site), cropland (CRO; two sites), and grass and other types (GRA; three sites), as shown in Table 1. The sites were selected according to the following criteria: (a) data being quality controlled; (b) extensive data set with minimal gaps; and (c) availability of all other requireaad input data for simulation using the different models considered for this study. Because of high data availability data from 2005 was selected for this analysis.

These data sets included half-hourly or hourly ground-measured incident solar radiation (Rs), relative humidity (RH), air temperature (Ta), diurnal air-temperature range (DT), wind speed (Ws), vapor pressure (e), sensible heat flux (H), surface net radiation [48], ground heat flux (G), and LE. When available, data sets were gap-filled by site principal investigators (PIs), and daily data was aggregated from half-hourly or hourly data without using additional quality control [49–51]. The more detailed information of the validation data were listed in Table 1 and Fig 2. Although the EC technique is regarded as a good method for measuring heat fluxes, the EC based LE values has to be corrected because of the unclosed energy problem (Twine et al. [52]; Wilson et al. [48]). The method developed by Twine et al. [52] was applied to correct
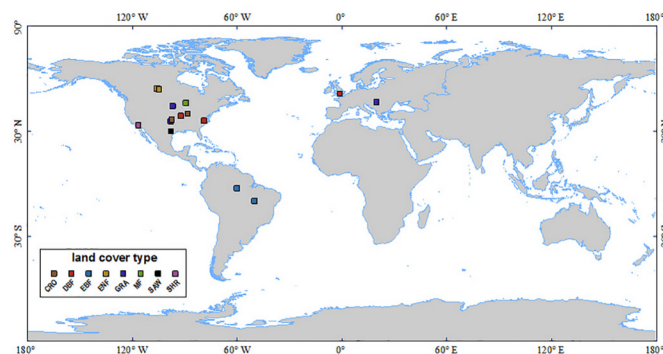


**Fig 2. Spatial distribution of the validation FLUXNET sites used in this study.** The maps were drawn by the MCD12C1 product for 2005.

doi:10.1371/journal.pone.0160150.g002

**Table 1. Characteristics of the selected validation data at the FLUXNET sites (S1 Table).**

| Sites ID | Site name | Latitude | Longitude | IGBP | Available years |
|---|---|---|---|---|---|
| US-ARM | ARM Southern Great Plains control site | 36.61 | -97.49 | CRO | 2000–2013 |
| US-ARC | ARM Southern Great Plains control site | 35.55 | -98.04 | GRA | 2005–2006 |
| US-Bkg | Brookings | 44.35 | -96.84 | GRA | 2004–2010 |
| US-Bo1 | Bondville | 40.01 | -88.29 | CRO | 1996–2010 |
| US-Bo2 | Bondville | 40 | -88.29 | CRO | 2004–2008 |
| US-Dk2 | Duke Forest Hardwoods | 35.97 | -79.1 | DBF | 2001–2008 |
| US-FR2 | Freeman Ranch- Mesquite Juniper | 29.95 | -98 | SAW | 2005–2008 |
| US-MOz | Missouri Ozark Site | 38.74 | -92.2 | DBF | 2004–2013 |
| US-SO2 | Sky Oaks Old | 33.37 | -116.62 | SHR | |
| US-SO3 | Sky Oaks- Young Stand | 33.38 | -116.64 | SHR | 2001–2006 |
| US-SO4 | Sky Oaks New | 33.38 | -116.64 | SHR | |
| US-Syv | Sylvania Wilderness | 46.69 | -89.35 | MF | 2001–2008 |
| HU-Bug | Bugacpuszta | 46.69 | 19.6 | GRA | 2002–2006 |
| UK-PL3 | Pang Lambourne (forest) | 51.45 | -1.27 | DBF | 2005–2006 |
| BR-Ban | Ecotone Bananal Island | -9.82 | -50.16 | EBF | 2003–2006 |
| BR-Ma2 | Manaus—ZF2 K34 | -2.61 | -60.21 | EBF | 1999–2006 |
| CA-Obs | Sask.- SSA Old Black Spruce | 53.99 | -105.19 | ENF | 1999–2005 |
| CA-Ojp | Sask.- SSA Old Jack Pine | 53.92 | -104.69 | ENF | 1999–2005 |
| CA-SF1 | Sask.-Fire 1977 | 54.49 | -105.82 | ENF | 2003–2005 |
| CA-SF2 | Sask.-Fire 1989 | 54.25 | -105.88 | ENF | 2003–2005 |
| CA-SJ2 | Sask.-2002 Harvested Jack Pine | 53.95 | -104.65 | ENF | 2003–2005 |
| CA-Sj3 | Sask.-1975 (Young) Jack Pine | 53.88 | -104.65 | ENF | 2004–2005 |

doi:10.1371/journal.pone.0160150.t001

LE for all flux towers,

$$LE = (R_n - G)/(LE_{ori} + H_{ori}) \times LE_{ori} \qquad (6)$$

Where LE is the corrected latent heat flux, $H_{ori}$ and $LE_{ori}$ were the uncorrected sensible heat flux and latent heat flux, respectively.

## 4. Results and Discussion

### 4.1 Comparison of MOD16 and PT-JPL algorithms

Daily LE estimates from MOD16 and PT-JPL based on both tower-measured meteorology data and MERRA meteorology data were compared. Furthermore, satellite based LE estimates were also checked against measured LE data for the FLUXNET sites, as shown in Fig 3.

The two models had good performance over these sites. The correlation coefficients, $R^2$, of the TP-JPL and MOD16 using in situ meteorology data are higher than 0.6, which corresponds to a good correlation to the measured LE values. Significant differences were found concerning theuncertainties of different algorithms when using in situ and forcing data. A possible reason for the differences might be the scale mismatch issue [53]. The comparison of our validation results with previous studies were in good agreement. Fisher et al. [40] found that PT-JPL had high correlation with ground observations from16 FLUXNET sites. Chen et al. [54] reported that the PT-JPL showed a good performance with $R^2$ equal to 0.8. The MOD16 validation at Brazil EC sites conducted by Ruhoff et al. [55] showed that the correlation coefficient between the ground observation and MOD16 estimates for 8-days average was 0.79, RMSE was 0.78
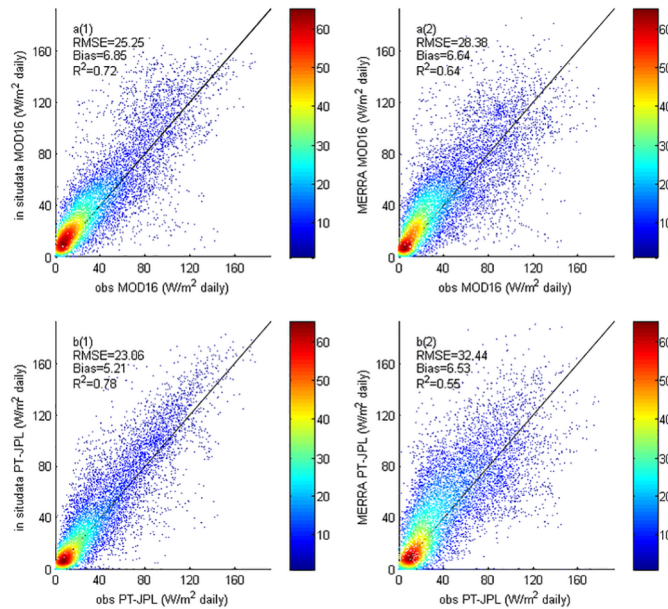
**Fig 3. Comparison of MOD16 and PT-JPL with in situ and meteorological forcing data.** (a1) MOD16 vs in situ data. (a2) MOD16 vs meteorological forcing data. (b1) PT-JPL vs in situ data. (b2) PT-JPL vs meteorological forcing data.

mm day$^{-1}$ and mean bias was 0.54 mm day$^{-1}$. The good performances of MOD16 and PT-JPL might be attributed to good physical basis of these two models. However, previous studies also showed that MOD16 had reduced performance when compared to ground observations. Chen et al. [4] and Ershadi et al. [20] reported reduced performance of MOD16 compared with Chinaflux EC sites and American EC sites. MOD16 validation conducted by Ramoelo et al. [56] suggested that disagreement of MOD16 and flux tower-based ET could be attributed to the parameterization of Penman-Monteith model. Therefore, the parameterization of Penman-Monteith model at different sites or climatic zones might cause the different performances of MOD16.

Small discrepancies in LE were produced by MOD16 and PT-JPL (Fig 3). Both MOD16 and PT-JPL showed positive bias compared with ground measurements. However, as published by Behrangi et al. [57], the MOD16 slightly underestimates LE as compared to PT-JPL. Moreover, when compared with EC towers in Asia, MOD16 had a negative bias (-17.00 mm 8-day$^{-1}$) especially for the cropland sites [58]. This might be due to the different location of the ground observations, i.e. most of EC sites in this study were collected at high latitudes (Fig 1) and in high latitudes, temperature has great impact on LE estimations [59].

We found that vegetation type has a great influence on the performance of MOD16 and PT-JPL (Fig 4). PT-JPL showed the higher R$^2$ (0.96) for MF sites, while MOD16 had the higher R$^2$ (0.75) for the ENF sites. PT-JPL generally had lower bias than MOD16 for CRO, GRA, DBF and ENF, whereas the MOD16 showed a lower bias than PT-JPL for EBF, MF, SAW and SHR. Both algorithms showed negative bias for CRO and GRA sites. The underestimation of MOD16 for cropland site was also reported in the previous study [58]. Yao et al. [28] also found that MOD16 and PT-JPL underestimates LE in the case of cropland and grassland sites. Negative biases in simulated LE by these two algorithms might be attributed to the uncertainty in soil moisture estimation.
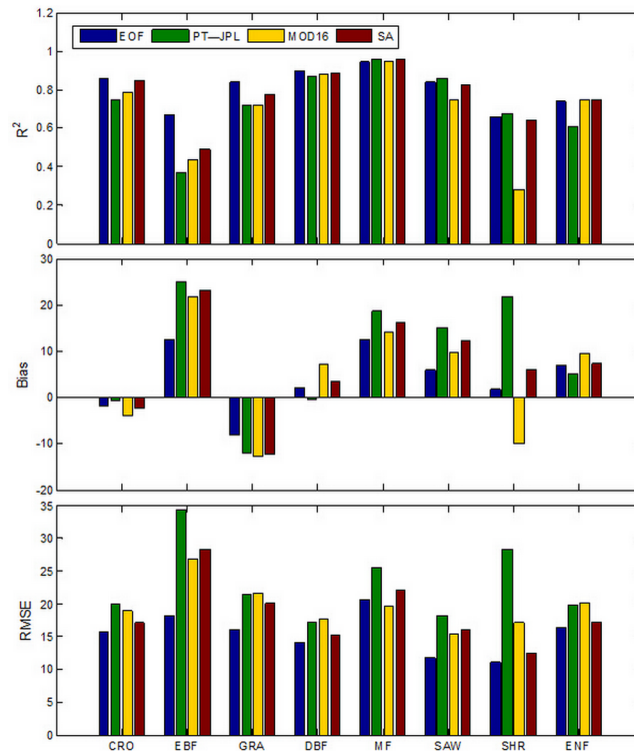
**Fig 4. Direct validation results of EOF-integrated LE for FLUXNET sites at eight biomes: cropland (CRO), evergreen broadleaf forest (EBF), grassland (GRA), deciduous broadleaf forest (DBF), mixed forest (MF), savanna (SAW), shrubland (SHR) and evergreen needleleaf forest (ENF).**

doi:10.1371/journal.pone.0160150.g004

## 4.2 Performance of EOF and SA fusion algorithms

The MOD16, PT-JPL, SA and EOF algorithms exhibited substantial differences when comparing the modelled LE to the LE observed at the 22 EC flux tower sites, as shown in Fig 5. However, MOD16, PT-JPL, SA and EOF algorithms successfully predicted the magnitudes and seasonal variations of the observed LE at the validation sites. Compared with MOD16 and PT-JPL, the fusion algorithms (SA and EOF) showed closer correlation with observed LE, LE predicted by EOF being the best estimate for most sites. Generally, previous studies also showed that fusion methods could produce more accurate LE estimates than the individual LE algorithm. Ershadi et al. [20] found that the ensemble mean of the individual LE models produced the best estimates of LE, with the mean value of the Nash–Sutcliffe efficiency of 0.61 and the root mean squared difference of 64 W/m$^2$. Yao et al. [28] introduced a Bayesian model averaging (BMA) method by merging five process-based LE models. This BMA method showed improved performance compared with individual LE models from 240 FLUXNET EC sites $R^2$ being equal to 0.8, bias equal to 3.5 W/m$^2$ and RMSE equal to 32.8 W/m$^2$.

For validation the 22 sites were categorized according to land cover type, as shown in Fig 4. Both EOF and SA showed higher $R^2$ than MOD16 and PT-JPL for CRO, EBF, GRA and DBF ranging from 0.49 to 0.86. In the case of CRO, EBF and GRA sites, LE estimate by EOF showed the highest correlation ranging from 0.67 to 0.86. In terms of RMSE EOF produced the lowest values ranging from 11.07 to 18.10 W/m$^2$ for all biomes. Except for DBF and ENF sites bias was the lowest for EOF. SA had a relative good performance with bias ranging from -12.37 to
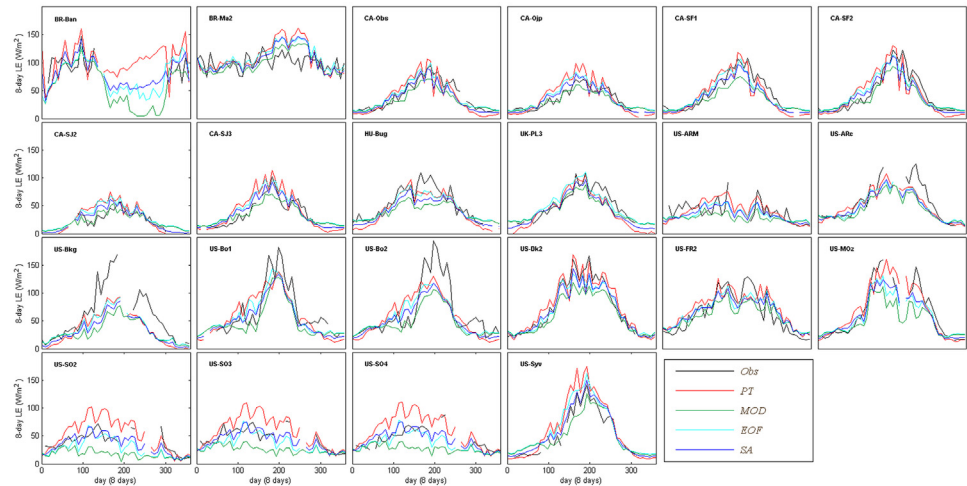
**Fig 5. Validation of the 8-day mean of predicted and observed LE at all sites in 2005.**

doi:10.1371/journal.pone.0160150.g005

23.40 W/m$^2$ and RMSE ranging from 15.39 to 28.37 W/m$^2$ over CRO, GRA and DBF sites. However, SA showed limited improvement for EBF, GRA and MF sites compared with MOD16 and PT-JPL. However there is a variety of individual LE models with different algorithm structures and parameterization, none of them is capable of providing a best LE estimate for all biomes. Hence, the reduced performance of SA might be attributed to the simple constant weights of different LE models [28]. Because EOF took both spatial and temporal information into consideration when reconstructing LE and maintained the main spatial pattern of individual LE model [35], compared to SA, EOF provided improved performance over most vegetation types. The previous study [37] showed that fusion methods based on EOF had substantially improved the accuracy of LAI with R$^2$ increasing from 0.75 to 0.81 and RMSE decreasing from 1.04 to 0.71. The simple structure of EOF fusion might partly explain this improvement. The major advantage of EOF was that it avoided using the measured LE values which were used by many of the fusion methods [28]. Therefore the sensitivity of this parameterization to errors in the input data was substantially lessened [60]. Another advantage was that it was easy to implement and did not require the use of a precalculated covariance model and estimation error matrix [37].

Considering all vegetation types (all validation sites), MOD16, PT-JPL, EOF and SA explained 77%, 74%, 84% and 81% of the variation of the 8 day average LE estimates, respectively (Fig 6), with MOD16 slightly overestimating and PT-JPL overestimating LE. All methods showed positive bias ranging from 3.67 W/m$^2$ to 7.19 W/m$^2$ and RMSE was in the range of 14.83 W/m$^2$ and 21.84 W/m$^2$. The proposed EOF method showed then lowest bias and RMSE and the highest R$^2$ (0.84).

To compare spatial patterns of LE in the case of the four algorithms, we randomly selected images around US-SO2 site (Fig 7). Similar tendencies of LE predicted by the four algorithms were found, LE was decreasing regularly with time. As expected, for a given day the largest difference was between LE predicted by SA and EOF. Compared to SA, EOF showed relatively low LE in upper-left corner of the images, which was more consistent with the LE predicted by MOD16. EOF and SA both showed high LE in bottom left corner of the images. However, the SA showed distinct average results of MOD16 and PT-JPL.
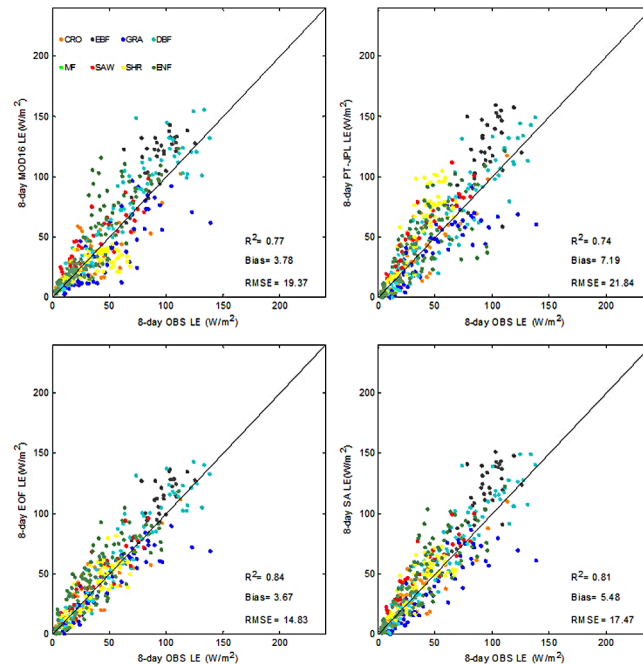
**Fig 6. Validation of EOF, MOD16 and PT-JPL LE methods across different land use types.** 8 day average LE prediction is compared to ground measurements. The solid line is the 1:1 line.

## 4.3 Limitations of the proposed EOF algorithm

The proposed EOF algorithm requires fewer input parameters than geostatistical approaches, e.g. a precalculated covariance model. However, EOF does have some limitations, and the computation cost is very high due to matrix calculations and iterations. Uncertainties that are limiting the use of the EOF model are the following:
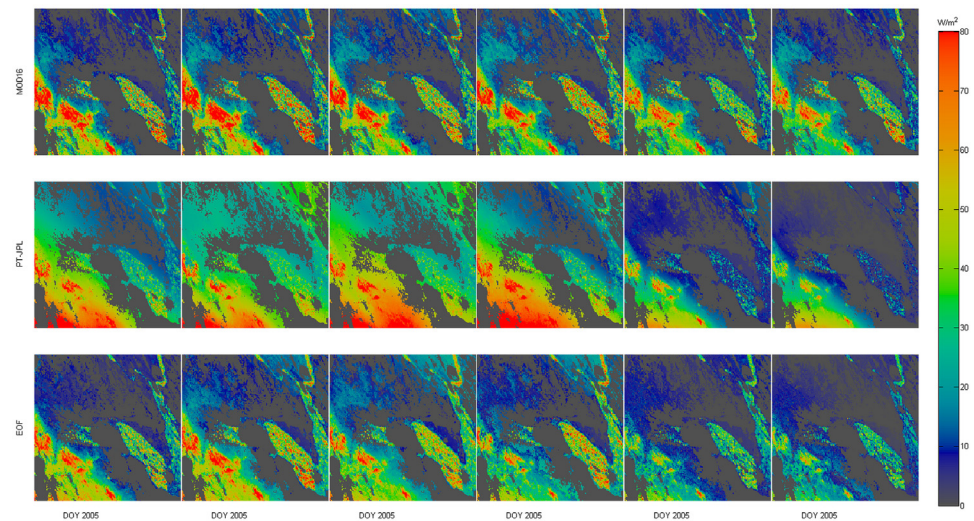


**Fig 7. EOF maintained major pattern of PT-JPL and removed the extreme values as compared with the SA, MOD16 and PT-JPL methods during the periods from February 10, 2005, to March 22, 2005, at the US-SO2 AmeriFlux Site.** The color bar is an 8-day composite LE. Dark gray color means no data (0 value).

1. The bias for MERRA [61] might lead to substantial bias for the two individual LE algorithms and hence to the EOF ensembles. Study showed that MERRA surface solar radiation, which was used as an input of LE models, had an average bias error of +20.2 W/m$^2$ on monthly and annual scales from American FLUXNET sites [62], resulting in an overestimation of LE. Zib et al. [63] also reported an annual mean bias of 3.9 W/m$^2$ at two Baseline Surface Radiation Network (BSRN) sites for MERRA surface solar radiation; Wang and Zeng [64] found an overestimation of up to 40 W/m$^2$ for MERRA surface solar radiation.

2. Scale mismatch between coarse resolution of input data and the field measurements footprint may results in substantial bias in EOF fusion producing. Wolde et al. [65] analyzed the different pixel resolution of remote sensing inputs, and showed that variation in ET flux between corn and soybean field could not be effectively distinguished when the input was of the order of 1000 m [66] found that coarse NCEP/NCAR reanalysis meteorology (NNR) data can introduce bias to match the local tower footprint in some regions.

3. By integrating the different LE algorithms, bias might be introduced during the EOF fusion process. The reason it that, when applying the EOF fusion method, EOF reconstruction scheme [42] does not distinguish between good or degraded quality pixel values. Consequently, bias is introduced, since all pixels in the image are included in the reconstruction during the spatio-temporal fusion process.

## 5. Conclusions

We proposed a data merging method based on EOF analysis and applied this method to integrate two satellite-derived LE products (MOD16 and PT-JPL). We also compared the proposed EOF method with simple SA fusion method. Ground-measured LE data in 2005 from 22 EC sites, incorporating eight major terrestrial biomes (CRO, DBF, EBF, ENF, GRA, MF, SAW and SHR), were used for validation, and demonstrated that the proposed method was suitable for terrestrial LE mapping.

MOD16 and observed data correlated well for EBF, MF, SAW and SHR biomes, producing higher R$^2$, although somewhat larger RMSE and high bias. For CRO, GRA, DBF and EOF sites, PT-JPL produced lower bias with lower RMSE. Although SA fusion method provided acceptable results compared with MOD16 and PT-JPL, the proposed EOF algorithm showed notable improvement by combining the advantages of MOD16 and PT-JPL and had a relatively low bias and RMSE with high R$^2$ for all biomes. EOF integrated images were superior to LE maps generated by the PT-JPL and MOD16 algorithms.

## Supporting Information

**S1 File. The EOF program and the validation result.** the program were contained in the S1 File. rar. The EOF program were derived in IDL(.pro). the validation results were in the Excel tables.
(RAR)

**S1 Table. Validation sites description.** The Characteristics of the the FLUXNET sites. Data was in the S1 Table.xlsx.
(XLSX)

## Acknowledgments

## Author Contributions

Conceived and designed the experiments: FF XL YY. Performed the experiments: FF XL. Analyzed the data: FF XL. Contributed reagents/materials/analysis tools: SL JC XZ KJ KP JHM. Wrote the paper: FF. Language editing: KP.

## References

1. Council N. Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond: The National Academies Press; 2007.

2. Wang KC, Dickinson RE. A Review of Global Terrestrial Evapotranspiration: Observation, Modeling, Climatology, and Climatic Variability. Rev Geophys. 2012; 50. PubMed PMID: WOS:000304588000001.

3. Meyer W, Smith D, Shell G. Estimating reference evaporation and crop evapotranspiration from weather data and crop coefficients. Technical Report CSIRO Land and Water, 1999 34/98.

4. Chen Y, Xia JZ, Liang SL, Feng JM, Fisher JB, Li X, et al. Comparison of satellite-based evapotranspiration models over terrestrial ecosystems in China. Remote Sens Environ. 2014; 140:279–93. doi: 10.1016/j.rse.2013.08.045. PubMed PMID: WOS:000329766200024.

5. Mu Q, Heinsch FA, Zhao M, Running SW. Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. Remote Sens Environ. 2007; 111(4):519–36. PubMed PMID: WOS:000251099300008.

6. Fisher JB, Tu KP, Baldocchi DD. Global estimates of the land-atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites. Remote Sens Environ. 2008; 112 (3):901–19. PubMed PMID: WOS:000254443700023.

7. Wang KC, Dickinson RE, Wild M, Liang SL. Evidence for decadal variation in global terrestrial evapotranspiration between 1982 and 2002: 1. Model development. J Geophys Res-Atmos. 2010; 115. PubMed PMID: WOS:000283546600003.

8. Wang KC, Dickinson RE, Wild M, Liang SL. Evidence for decadal variation in global terrestrial evapotranspiration between 1982 and 2002: 2. Results. J Geophys Res-Atmos. 2010; 115. PubMed PMID: WOS:000283546600005.

9. Yao YJ, Liang SL, Cheng J, Liu SM, Fisher JB, Zhang XD, et al. MODIS-driven estimation of terrestrial latent heat flux in China based on a modified Priestley-Taylor algorithm. Agr Forest Meteorol. 2013; 171:187–202. PubMed PMID: WOS:000316513000018.

10. Yuan WP, Liu SG, Yu GR, Bonnefond JM, Chen JQ, Davis K, et al. Global estimates of evapotranspiration and gross primary production based on MODIS and global meteorology data. Remote Sens Environ. 2010; 114(7):1416–31. PubMed PMID: WOS:000277878900008.

11. Mu Q, Zhao M, Running SW. Improvements to a MODIS global terrestrial evapotranspiration algorithm. Remote Sensing of Environment. 2011; 115(8):1781–800.

12. Montenegro A, Eby M, Mu QZ, Mulligan M, Weaver AJ, Wiebe EC, et al. The net carbon drawdown of small scale afforestation from satellite observations. Global Planet Change. 2009; 69(4):195–204. PubMed PMID: WOS:000272901500002.

13. Chen Y, Xia J, Liang S, Feng J, Fisher JB, Li X, et al. Comparison of satellite-based evapotranspiration models over terrestrial ecosystems in China. Remote Sens Environ. 2014; 140(0):279–93. http://dx.doi.org/10.1016/j.rse.2013.08.045.

14. Long D, Longuevergne L, Scanlon BR. Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites. Water Resour Res. 2014; 50(2):1131–51. doi: 10.1002/2013wr014581. PubMed PMID: WOS:000333563900022.

15. Fisher JB, Debiase T, Xu M, Qi Y, Fisher JB, Debiase T, et al. Evapotranspiration Methods Compared on a Sierra Nevada Forest Ecosystem. Environmental Modelling & Software. 2005; 20(6):783–96.

16. JiméNez C, Prigent C, Mueller B, Seneviratne SI, McCabe MF, Wood EF, et al. Global intercomparison of 12 land surface heat flux estimates. Journal of Geophysical Research: atmospheres. 2011; 116(D2):3–25.

17. Fisher JB, Whittaker RJ, Malhi Y. ET come home: potential evapotranspiration in geographical ecology. Global Ecology & Biogeography. 2011; 20(1):1–18.

18. Mueller B, Seneviratne SI, Jimenez C, Corti T, Hirschi M, Balsamo G, et al. Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations. Geophysical Research Letters. 2011; 38(6):422–33.

19. Vinukollu RK, Wood EF, Ferguson CR, Fisher JB. Global estimates of evapotranspiration for climate studies using multi-sensor remote sensing data: Evaluation of three process-based approaches. Remote Sens Environ. 2011; 115(3):801–23.

20. Ershadi A, Mccabe MF, Evans JP, Chaney NW, Wood EF. Multi-site evaluation of terrestrial evaporation models using FLUXNET data. Agricultural & Forest Meteorology. 2014; 187(8):46–61.

21. Ershadi A, McCabe MF, Evans JP, Walker JP. Effects of spatial aggregation on the multi-scale estimation of evapotranspiration. Remote Sens Environ. 2013; 131(0):51–62. http://dx.doi.org/10.1016/j.rse.2012.12.007.

22. Cammalleri C, Anderson MC, Gao F, Hain CR, Kustas WP. Mapping daily evapotranspiration at field scales over rainfed and irrigated agricultural areas using remote sensing data fusion. Agr Forest Meteorol. 2014; 186(0):1–11. http://dx.doi.org/10.1016/j.agrformet.2013.11.001.

23. Caparrini F, Castelli F, Entekhabi D. Mapping of land-atmosphere heat fluxes and surface parameters with remote sensing data. Bound-Lay Meteorol. 2003; 107(3):605–33. doi: 10.1023/A:1022821718791. PubMed PMID: WOS:000181481900005.

24. Duan QY, Ajami NK, Gao XG, Sorooshian S. Multi-model ensemble hydrologic prediction using Bayesian model averaging. Adv Water Resour. 2007; 30(5):1371–86. doi: 10.1016/j.advwatres.2006.11.014. PubMed PMID: WOS:000246092800025.

25. Qin J, Liang SL, Liu RG, Zhang H, Hu B. A weak-constraint-based data assimilation scheme for estimating surface turbulent fluxes. Ieee Geosci Remote S. 2007; 4(4):649–53. doi: 10.1109/Lgrs.2007.904004. PubMed PMID: WOS:000250389900031.

26. Xu TR, Bateni SM, Liang S, Entekhabi D, Mao KB. Estimation of surface turbulent heat fluxes via variational assimilation of sequences of land surface temperatures from Geostationary Operational Environmental Satellites. J Geophys Res-Atmos. 2014; 119(18):10780–98. PubMed PMID: WOS:000344052800011.

27. Xu TR, Liang SL, Liu SM. Estimating turbulent fluxes through assimilation of geostationary operational environmental satellites data using ensemble Kalman filter. J Geophys Res-Atmos. 2011; 116. PubMed PMID: WOS:000290622700003.

28. Yao YJ, Liang SL, Li XL, Hong Y, Fisher JB, Zhang NN, et al. Bayesian multimodel estimation of global terrestrial latent heat flux from eddy covariance, meteorological, and satellite observations. J Geophys Res-Atmos. 2014; 119(8):4521–45. PubMed PMID: WOS:000335809100007.

29. Mueller B, Hirschi M, Jimenez C, Ciais P, Dirmeyer PA, Dolman AJ, et al. Benchmark products for land evapotranspiration: LandFlux-EVAL multi-dataset synthesis. Hydrology & Earth System Sciences. 2013; 17(10):3707–20.

30. Wald L. Some terms of reference in data fusion. Ieee T Geosci Remote. 1999; 37(3):1190–3. PubMed PMID: WOS:000080352100002.

31. Hannachi A, Jolliffe IT, Stephenson DB. Empirical orthogonal functions and related techniques in atmospheric science: A review. International Journal of Climatology. 2007; 27(9):1119–52. doi: 10.1002/joc.1499

32. Richman MB. Principal Component Analysis in Meteorology and Oceanography—Preisendorfer,Rw. Nature. 1989; 339(6227):673-. PubMed PMID: WOS:A1989AD58200047.

33. Chen JM, Harr PA. Interpretation of Extended Empirical Orthogonal Function (Eeof) Analysis. Mon Weather Rev. 1993; 121(9):2631–6. PubMed PMID: WOS:A1993LU35300013.

34. Smith TM, Reynolds RW, Livezey RE, Stokes DC. Reconstruction of historical sea surface temperatures using empirical orthogonal functions. J Climate. 1996; 9(6):1403–20. PubMed PMID: WOS: A1996UY34000017.

35. Beckers JM, Rixen M. EOF calculations and data filling from incomplete oceanographic datasets. J Atmos Ocean Tech. 2003; 20(12):1839–56. PubMed PMID: WOS:000187668000011.

36. Alvera-Azcarate A, Barth A, Rixen M, Beckers JM. Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: application to the Adriatic Sea surface temperature. Ocean Model. 2005; 9(4):325–46. PubMed PMID: WOS:000228343700002.

37. Wang DD, Liang SL. Integrating MODIS and CYCLOPES Leaf Area Index Products Using Empirical Orthogonal Functions. Ieee T Geosci Remote. 2011; 49(5):1513–9. PubMed PMID: WOS:000289906200003.

38. Monteith JL. Evaporation and environment. Symposia of the Society for Experimental Biology. 1965; 19:205–34. PMID: 5321565

39. Priestley CHB, Taylor RJ. On the assessment of surface heat flux and evaporation using large-scale parameters. Mon Weather Rev. 1972; 100(100):81–92.

40. Fisher JB, Tu KP, Baldocchi DD. Global estimates of the land–atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites. Remote Sensing of Environment. 2008; 112(3):901–19.

41. Yao Y, Liang S, Zhao S, Zhang Y, Qin Q, Cheng J, et al. Validation and Application of the Modified Satellite-Based Priestley-Taylor Algorithm for Mapping?Terrestrial Evapotranspiration. Remote Sensing. 2014; 6(1):880–904.

42. Zhang B, Pinker RT, Stackhouse PW. An Empirical Orthogonal Function Iteration Approach for Obtaining Homogeneous Radiative Fluxes from Satellite Observations. Journal of Applied Meteorology & Climatology. 2007; 46(4):435–44.

43. Dirmeyer PA, Jin Y, Singh B, Yan X. Trends in Land-Atmosphere Interactions from CMIP5 Simulations. J Hydrometeor. 2013; 14(3):829–49.

44. Cleugh HA, Leuning R, Mu QZ, Running SW. Regional evaporation estimates from flux tower and MODIS satellite data. Remote Sens Environ. 2007; 106(3):285–304. PubMed PMID: WOS:000244352900002.

45. Priestley CHB, Taylor RJ. Assessment of Surface Heat-Flux and Evaporation Using Large-Scale Parameters. Mon Weather Rev. 1972; 100(2):81-+. PubMed PMID: WOS:A1972L907500001.

46. Liang SL, Zhao X, Liu SH, Yuan WP, Cheng X, Xiao ZQ, et al. A long-term Global LAnd Surface Satellite (GLASS) data-set for environmental studies. Int J Digit Earth. 2013; 6:5–33. PubMed PMID: WOS:000328243700002.

47. Running SW, Nemani RR, Heinsch FA, Zhao M. Improvements of the MODIS terrestrial gross and net primary production global data set. Remote Sens Environ. 2005; 95(2):164–76.

48. Wilson K, Goldstein A, Falge E, Aubinet M, Baldocchi D, Berbigier P, et al. Energy balance closure at FLUXNET sites. Agr Forest Meteorol. 2002; 113(1–4):223–43. PubMed PMID: WOS:000179188300012.

49. Liu SM, Xu ZW, Wang WZ, Jia ZZ, Zhu MJ, Bai J, et al. A comparison of eddy-covariance and large aperture scintillometer measurements with respect to the energy balance closure problem. Hydrol Earth Syst Sc. 2011; 15(4):1291–306. doi: 10.5194/hess-15-1291-2011. PubMed PMID: WOS:000290016400016.

50. Jia ZZ, Liu SM, Xu ZW, Chen YJ, Zhu MJ. Validation of remotely sensed evapotranspiration over the Hai River Basin, China. J Geophys Res-Atmos. 2012; 117. Artn D13113 doi: 10.1029/2011jd017037. PubMed PMID: WOS:000306463100001.

51. Xu ZW, Liu SM, Li X, Shi SJ, Wang JM, Zhu ZL, et al. Intercomparison of surface energy flux measurement systems used during the HiWATER-MUSOEXE. J Geophys Res-Atmos. 2013; 118(23):13140–57. doi: 10.1002/2013jd020260. PubMed PMID: WOS:000330266500034.

52. Twine TE, Kustas WP, Norman JM, Cook DR, Houser PR, Meyers TP, et al. Correcting eddy-covariance flux underestimates over a grassland. Agr Forest Meteorol. 2000; 103(3):279–300. PubMed PMID: WOS:000087530300004.

53. Kustas WP, Li F, Jackson TJ, Prueger JH, Macpherson JI, Wolde M. Effects of remote sensing pixel resolution on modeled energy flux variability of croplands in Iowa. Remote Sens Environ. 2004; 92 (4):535–47.

54. Chen Y, Xia J, Liang S, Feng J, Fisher JB, Li X, et al. Comparison of satellite-based evapotranspiration models over terrestrial ecosystems in China. Remote Sensing of Environment. 2014; 140(1):279–93.

55.  Ruhoff AL, Collischonn W, Paz AR, Aragao LEOC, Mu Q, Rocha HR, et al. Validation of the newly improved global evapotranspiration algorithm (MOD16) in two contrasting tropical land cover types. IAHS-AISH publication. 2012; 3(352):128–31.

56.  Ramoelo A, Majozi N, Mathieu R, Jovanovic N, Nickless A, Dzikiti S. Validation of Global Evapotranspiration Product (MOD16) using Flux Tower Data in the African Savanna, South Africa. Remote Sensing. 2014; 6(8):7406–23.

57.  Behrangi A, Sun W, Fisher KM, B. J. On the net surface water exchange rate estimated from remote-sensing observation and reanalysis. International Journal of Remote Sensing. 2014; 35(6):2170–85.

58.  Kim HW, Hwang K, Mu Q, Lee SO, Choi M. Validation of MODIS 16 Global Terrestrial Evapotranspiration Products in Various Climates and Land Cover Types in Asia. Ksce Journal of Civil Engineering. 2011; 16(2):229–38.

59.  Nemani RR, Keeling CD, Hirofumi H, Jolly WM, Piper SC, Tucker CJ, et al. Climate-Driven Increases in Global Terrestrial Net Primary Production from 1982 to 1999. Science. 2003; 300(5625):1560–3. PMID: 12791990

60.  Wang K, Liang S, editors. An Improved Method For Estimating Global Evapotranspiration Based On Satellite Determination Of Surface Net Radiation, Vegetation Index, Temperature, And Soil Moisture. Geoscience and Remote Sensing Symposium, 2008 IGARSS 2008 IEEE International; 2008.

61.  Badgley G, Fisher JB, Jiménez C, Tu KP, Vinukollu R. On uncertainty in global terrestrial evapotranspiration estimates from choice of input forcing datasets. Journal of Hydrometeorology. 2015;(2015: ).

62.  Zhao L, Lee X, Liu S. Correcting surface solar radiation of two data assimilation systems against FLUXNET observations in North America. Journal of Geophysical Research Atmospheres. 2013; 118(17):9552–64.

63.  Zib BJ, Dong X, Xi B, Kennedy A. Evaluation and Intercomparison of Cloud Fraction and Radiative Fluxes in Recent Reanalyses over the Arctic Using BSRN Surface Observations. J Climate. 2012; 25 (7):2291–305.

64.  Wang A, Zeng X. Evaluation of multireanalysis products with in situ observations over the Tibetan Plateau. Journal of Geophysical Research Atmospheres. 2012; 117(D5):214–21.

65.  Wolde M, Kustas WP, Li F, Prueger JH, Macpherson JI, Jackson TJ. Effects of remote sensing pixel resolution on modeled energy flux variability of croplands in Iowa. Remote Sens Environ. 2004; 92 (4):535–47.

66.  Zhang K, Kimball JS, Nemani RR, Running SW. A continuous satellite-derived global record of land surface evapotranspiration from 1983 to 2006. Water Resour Res. 2010; 46. Artn W09522 doi: 10. 1029/2009wr008800. PubMed PMID: WOS:000282044800002.