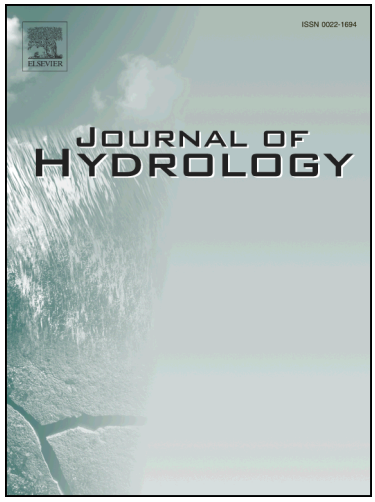
Estimation of high-resolution terrestrial evapotranspiration from Landsat data using a simple Taylor skill fusion method



Research papers

Estimation of high-resolution terrestrial evapotranspiration from Landsat data using a simple Taylor skill fusion method

Yunjun Yao, Shunlin Liang, Xianglan Li, Yuhu Zhang, Jiquan Chen, Kun Jia, Xiaotong Zhang, Joshua B. Fisher, Xuanyu Wang, Lilin Zhang, Jia Xu, Changliang Shao, Gabriela Posse, Yingnian Li, Vincenzo Magliulo, Andrej Varlagin, Eddy J. Moors, Julia Boike, Craig Macfarlane, Tomomichi Kato, Nina Buchmann, D.P. Billesbach, Jason Beringer, Sebastian Wolf, Shirley A. Papuga, Georg Wohlfahrt, Leonardo Montagnani, Jonas Ardö, Eugénie Paul-Limoges, Carmen Emmel, Lukas Johannes Hörtnagl, Torsten Sachs, Carsten Gruening, Beniamino Gioli, Ana López-Ballesteros, Rainer Steinbrecher, Bert Gielen

PII: S0022-1694(17)30539-5
DOI: http://dx.doi.org/10.1016/j.jhydrol.2017.08.013
Reference: HYDROL 22176

To appear in: *Journal of Hydrology*

Received Date: 7 January 2017
Revised Date: 20 July 2017
Accepted Date: 10 August 2017

Please cite this article as: Yao, Y., Liang, S., Li, X., Zhang, Y., Chen, J., Jia, K., Zhang, X., Fisher, J.B., Wang, X., Zhang, L., Xu, J., Shao, C., Posse, G., Li, Y., Magliulo, V., Varlagin, A., Moors, E.J., Boike, J., Macfarlane, C., Kato, T., Buchmann, N., Billesbach, D.P., Beringer, J., Wolf, S., Papuga, S.A., Wohlfahrt, G., Montagnani, L., Ardö, J., Paul-Limoges, E., Emmel, C., Hörtnagl, L.J., Sachs, T., Gruening, C., Gioli, B., López-Ballesteros, A., Steinbrecher, R., Gielen, B., Estimation of high-resolution terrestrial evapotranspiration from Landsat data using a simple Taylor skill fusion method, *Journal of Hydrology* (2017), doi: http://dx.doi.org/10.1016/j.jhydrol.2017.08.013

Estimation of high-resolution terrestrial evapotranspiration from Landsat data using a simple Taylor skill fusion method

Yunjun Yao[a, *], Shunlin Liang [a], Xianglan Li[b], Yuhu Zhang[c], Jiquan Chen[d], Kun Jia[a], Xiaotong Zhang[a], Joshua B. Fisher[e], Xuanyu Wang[a], Lilin Zhang[a], Jia Xu[a], Changliang Shao[d], Gabriela Posse[f], Yingnian Li[g], Vincenzo Magliulo[h], Andrej Varlagin[i], Eddy J. Moors[j], Julia Boike[k], Craig Macfarlane[l] Tomomichi Kato[m], Nina Buchmann[n], D.P. Billesbach[o], Jason Beringer[p], Sebastian Wolf[n], Shirley A. Papuga[q], Georg Wohlfahrt[r], Leonardo Montagnani[s], Jonas Ardö[t], Eugénie Paul-Limoges[n], Carmen Emmel[n], Lukas Johannes Hörtnagl[n], Torsten Sachs[u], Carsten Gruening[v], Beniamino Gioli[w], Ana López-Ballesteros[x], Rainer Steinbrecher[y], Bert Gielen[z]

a. State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing, 100875, China

b. College of Global Change and Earth System Science, Beijing Normal University, Beijing, 100875, China

c. College of Resource Environment and Tourism, Capital Normal University, Beijing, 100048, China

d. CGCEO/Geography, Michigan State University, East Lansing, MI 48823, USA

e. Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109, USA

f. Climate and Water Institute, Research Center of Natural Resources, National Institute of Agricultural Technology (CIRN- INTA), Hurlingham, Argentina

g. Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, 810001,China

h. CNR-Institute of Mediterranean Forest and Agricultural Systems, Via Patacca, 85, 80040-Ercolano (Napoli), Italy

i. A.N. Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, Moscow, 119071, Russia

j. Wageningen Environmental Research, Wageningen University and Research, Wageningen, The Netherlands

k. Alfred Wegener Institute for Polar and Marine Research, Telegrafenberg A43, 14473 Potsdam, Germany

l. CSIRO Land and Water, Floreat W.A. 6014, Australia

m. Research Faculty of Agriculture, Hokkaido University, Sapporo, 060-8589, Japan

n. Department of Environmental Systems Science, ETH Zurich, Zurich, Switzerland

o. Department of Biological Systems Engineering and School of Natural Resources, University of Nebraska, Lincoln, Nebraska 68583, USA

p. School of Earth and Environment, The University of Western Australia, Crawley, WA, 6020, Australia

q. School of Natural Resources and the Environment, University of Arizona, Tucson, Arizona, 85721, USA

r. Institute of Ecology, University of Innsbruck, Innsbruck 6020, Austria
s. Faculty of Science and Technology, Free University of Bolzano, Piazza Universita 5, 39100 Bolzano, Italy
t. Physical Geography and Ecosystem Science Lund University Sölvegatan 12, SE-223 62 Lund, Sweden
u. GFZ German Research Centre for Geosciences, Section Remote Sensing, 14473 Potsdam, Germany
v. European Commission, Joint Research Centre, Ispra, Italy
w. Institute of Biometeorology, National Research Council, Via Caproni 8, 50145 Firenze, Italy
x. Department of Ecology, Faculty of Sciences, University of Granada, Granada, 18071, Spain
y. Karlsruhe Institute of Technology (KIT), Institute of Meteorology and Climate Research (IMK-IFU), 82467 Garmisch-Partenkirchen, Germany
z. Centre of Excellence PLECO (Plant and Vegetation Ecology), Dept. of Biology, University of Antwerp, Belgium

Journal of Hydrology (Revised)

Reference Number: HYDROL23975R1

*Corresponding Author: Yunjun Yao

Beijing Normal University, Beijing, 100875, China

Email: boyyunjun@163.com

Tel: 86+ 10-5880-3002

July 20, 2017

Abstract

Estimation of high-resolution terrestrial evapotranspiration (*ET*) from Landsat data is important in many climatic, hydrologic, and agricultural applications, as it can help bridging the gap between existing coarse-resolution *ET* products and point-based field measurements. However, there is large uncertainty among existing *ET* products from Landsat that limit their application. This study presents a simple Taylor skill fusion (*STS*) method that merges five Landsat-based *ET* products and directly measured *ET* from eddy covariance (*EC*) to improve the global estimation of terrestrial *ET*. The *STS* method uses a weighted average of the individual *ET* products and weights are determined by their Taylor skill scores (*S*). The validation with site-scale measurements at 206 *EC* flux towers showed large differences and uncertainties among the five *ET* products. The merged *ET* product exhibited the best performance with a decrease in the averaged root-mean-square error (*RMSE*) by 2-5 W/m^2 when compared to the individual products. To evaluate the reliability of the *STS* method at the regional scale, the weights of the *STS* method for these five *ET* products were determined using *EC* ground-measurements. An example of regional *ET* mapping demonstrates that the *STS*-merged *ET* can effectively integrate the individual Landsat *ET* products. Our proposed method provides an improved high-resolution *ET* product for identifying agricultural crop water consumption and providing a diagnostic assessment for global land surface models.

Keywords: Terrestrial evapotranspiration; Eddy covariance; Fusion method; Landsat data; High-resolution products

1. Introduction

The latent heat of evapotranspiration (*ET*), which is the sum of the heat flux from the earth's surface to the atmosphere for soil evaporation, vegetation transpiration and evaporation of water intercepted by plant canopies, plays an important role in many geophysical applications (*e.g.*, climatic forecasting, crop yield forecasting and agricultural water resource management) (Jiménez et al. 2011; Kool et al. 2014; Liang et al. 2010; Wang and Dickinson, 2012; Zhang et al. 2009). *ET* exhibits strong heterogeneity across the land surface due to complex environmental controls and biophysical feedback processes (Kalma et al. 2008; Mallick et al. 2009; Yao et al. 2014; Yuan et al. 2010). Large-scale networks of direct biosphere-atmosphere measurements with the eddy covariance (*EC*) method have been widely used for site-scale studies. However, such local *ET* observations cannot represent *ET* at regional to global scales (Baldocchi et al. 2001; Choi et al. 2009; Kustas and Anderson, 2009; Liu et al. 2016; Xu et al., 2011; Xu et al., 2016;Yao et al. 2015).

Remote sensing has provided us with an effective way to obtain spatially and temporally continuous *ET* data at a regional scale. Currently, there are various moderate spatial resolution satellite-based *ET* products available, such as the Moderate Resolution Imaging Spectroradiometer (*MODIS*) product (*MOD16*), which has 1 km and 8 day of spatial and temporal resolution, respectively (Mu et al. 2007; Mu et al. 2011) or the *EUMETSAT* Satellite Application Facility on Land Surface Analysis (*LSA-SAF*) product (*LSA-SAF MSG*) (Ghilain et al. 2011; Ghilain et al. 2012) with 5 km spatial resolution and daily temporal resolution. However, validation

results with direct measurements indicate that the *MOD16* and *LSA-SAF MSG ET* products tend to consist of uncertainties at most *FLUXNET* flux tower sites (Chen et al. 2014; Hu et al. 2015; Kim et al. 2012). Other *ET* products (including reanalysis and data assimilation datasets), such as the European Centre for Medium-Range Weather Forecasts (*ECMWF*) *ERA*-40 reanalysis (Uppala et al. 2005) and the Global Land Data Assimilation System (*GLDAS*) datasets, have high temporal resolution (daily) but rather coarse spatial resolution ($\geqslant$0.5°) (Kumar et al. 2006; Rodell et al. 2004). This relative coarse spatial resolution of global *ET* products limits the representation of the heterogeneous terrestrial biosphere.

The Landsat multispectral data record from the Thematic Mapper (*TM*) and the Enhanced Thematic Mapper Plus (*ETM+*) are valuable data sources for producing *ET* product at rather high spatial resolution (~30m) (Ju and Roy, 2008). They can also bridge the gap between existing coarse-resolution *ET* products and point-based field measurements and be used to validate coarse-resolution data. Various Landsat-based *ET* algorithms, roughly classified as *VNIR* (visible and near infrared)-based remote sensing methods and *TIR* (thermal infrared)-based remote sensing methods, have been developed to estimate regional *ET* (Kalma et al. 2008; Li et al. 2009). *VNIR*-based remote sensing methods include some empirical/statistical models (Glenn et al. 2008; Jung et al. 2010; Wang et al. 2007;Wang and Liang, 2008) and Penman-Monteith (*PM*)/Priestley-Taylor (*PT*) methods (Fisher et al. 2008; Jin et al. 2011; Mu et al. 2007; Mu et al. 2011; Priestley and Taylor, 1972; Yao et al. 2013), which usually use remotely sensed normalized difference vegetation index (*NDVI*) or leaf area index

(*LAI*) to estimate *ET*. *TIR*-based remote sensing methods, such as the Mapping Evapotranspiration with Internalized Calibration (Allen et al. 2007), the two-source model (Anderson et al. 1997; Kustas and Norman, 1999; Norman et al. 1995), Surface Energy Balance System (Su, 2002), Surface Energy Balance Algorithm for Land (Bastiaanssen et al. 1998), and the surface temperature versus vegetation index triangle/trapezoid space (Jiang and Islam, 1999; Long and Singh, 2012; Tang et al. 2010; Zhang et al. 2005), calculate *ET* as the residual of surface energy balance (*SEB*) or *PT* method from *TIR*-derived land surface temperatures (*LST*). Although these methods provide reasonable *ET* estimates for Landsat data, they still have large uncertainties in regional *ET* simulations because of different model structures and environmental variables employed (Chen et al. 2014; Choi et al. 2009; Liaqat and Choi, 2015; Liu et al. 2011; Liu et al. 2013).

This issue has been partially resolved by several data fusion methods, such as Bayesian model averaging (*BMA*) and empirical orthogonal function (*EOF*), which merge multiple *ET* products to improve regional *ET* estimation (Feng et al. 2016; Yao et al. 2014; Yao et al. 2016; Zhu et al. 2016). For example, Yao et al. (2014) used the *BMA* method by merging five *ET* products to enhance daily *ET* estimates with smaller root mean square errors (*RMSEs*) than those of the individual products. Zhu et al. (2016) also documented that the *BMA* method by merging four *ET* models across north China has the advantage of generating more skillful and reliable predictions than the simple model averaging (*SMA*) scheme. Similarly, Feng et al. (2016) reported that the *EOF* fusion method was capable of integrating the two

satellite-based *ET* datasets with improved consistency and reduced uncertainties. However, the complex structures of these fusion methods, which affect their computational efficiency for calculating the weightings for individual datasets, can limit their wide application.

To reduce the complexity of the fusion method and to generate global *ET* products with high spatial resolution, in this study we developed a simple Taylor skill fusion (*STS*) method by merging five Landsat-based *ET* products produced by the individual algorithms and *FLUXNET* eddy covariance (*EC*) observations to improve terrestrial *ET* estimation. The objectives of this study are threefold: (1) to evaluate five Landsat-based *ET* datasets derived from five classic *ET* algorithms using global long-term *FLUXNET* measurements from 206 flux tower sites; (2) to apply and validate the *STS* method for five Landsat-based *ET* datasets to improve terrestrial *ET* estimation; and (3) give an example of mapping terrestrial *ET* using the *STS* method and Landsat data.

2. Data

2.1 Landsat-based ET products

We produced the individual Landsat-based *ET* products using five classic *ET* algorithms. We only used five traditional Landsat-based *ET* products derived from *VNIR*-based remote sensing methods in this article because there are some disadvantages when applying *TIR*-based remote sensing methods to *ET* estimations at the global scale (Hope et al. 2005; Su, 2002). The forcing data includes Landsat *NDVI* data with 30m spatial resolution and daily Modern Era Retrospective Analysis

for Research and Applications (*MERRA*) meteorological data with 1/2×2/3 degree spatial resolution. All coarse resolution *MERRA* data were spatially interpolated into 30m using the method described by Zhao et al. (2005). Theoretically, this spatial interpolation method improves the accuracy of meteorological data for each 30m pixel because it uses a cosine function and the four *MERRA* cells surrounding a given pixel to remove sharp changes from one side of a *MERRA* boundary to the other (Zhao et al., 2005). The individual *ET* products (Table 1) are briefly described below.

Table 1 HERE

2.1.1 RS-PM ET product

The *RS-PM ET* product was produced based on a revised remote sensing-based *PM* (*RS-PM*) algorithm (Monteith, 1965) modified from the Mu et al. (2007) algorithm (Appendix A). The input variables include daily surface net radiation (R_n), relative humidity (*RH*), air temperature (T_a), and vapor pressure (*e*) from *MERRA* data and *LAI* derived from Landsat *NDVI* data. The *RS-PM ET* product has 30m spatial resolution. The product has a 16-day temporal resolution inherited from the Landsat data in the same time and filled values with -9999 are assigned for cloudy pixels. The product is global in coverage spanning 1998-2010 and future years are produced periodically.

2.1.2 SW ET product

The Shuttleworth-Wallace dual-source (*SW*) model (Shuttleworth and Wallace, 1985) was adopted to generate the *SW ET* product. The *SW* model accounts

separately for the energy balance for vegetation and soil components of a soil-vegetation canopy unity (Hu and Jia, 2015; Sellers et al. 1992; Shuttleworth and Wallace, 1985) (Appendix A). The *SW ET* product requires R_n, *RH*, T_a, *e*, and wind speed (*WS*) from *MERRA* data and *LAI* derived from Landsat *NDVI* data. The *SW ET* product has global coverage during the period of 1995-2009 at 30m spatial resolution and 16-day temporal resolution. The *ET* values for cloudy pixels are also set as -9999.

2.1.3 PT-JPL ET product

The *PT-JPL ET* product was produced using a novel *PT* algorithm developed by Fisher et al. (2008). This algorithm (*PT-JPL*) considers the effects of both atmosphere and ecophysiology to derive constraints representing vegetation conductance without using any ground-based observed data (Appendix A). The input variables are R_n, *RH*, T_a, *e*, and vegetation parameters (*NDVI*, *LAI* and absorbed photosynthetically active radiation (*FPAR*)) derived from Landsat *NDVI* data. The *PT-JPL ET* product is available from 1998 to 2010 globally and with 16-day temporal and at 30m spatial resolution.

2.1.4 MS-PT ET product

The *MS-PT ET* product was generated based on the modified satellite-based *PT* (*MS-PT*) model developed by Yao et al. (2013) and this model uses the apparent thermal inertia (*ATI*) derived from diurnal air temperature range (*DT*) to parameterize surface soil moisture (*SM*) constraints (Appendix A). The *MS-PT ET* product only requires R_n, T_a, *DT* from *MERRA* data and *NDVI* from Landsat data as inputs. The

MS-PT ET product is generated at the same spatial resolution, temporal resolution, coverage period, and filling values as the *PT-JPL ET* product for the period of 1997-2009.

2.1.5 SIM ET product

A simple hybrid *ET* (*SIM*) formulation presented by Wang et al. (2008) is used in this study to generate the *SIM ET* product; further, the *SIM* algorithm partitions the R_n by introducing T_a, *NDVI*, *DT* and prior parameters (Appendix A). The *SIM ET* product is also available at 30m spatial and 16-day temporal resolution over the global land surface from 1998 to 2009. The data processing step includes *MERRA* interpolation to the target grid size of 30 m.

2.2 Eddy covariance data

Five Landsat-based *ET* products and the *STS* fusion method were evaluated and validated using a large data set of ground-measured flux data. The data from 206 *EC* flux tower sites were provided by AmeriFlux, ChinaFlux, LathuileFlux, AsiaFlux, Arid/Semi-arid experimental observation synergy network of China, Chinese ecosystem research network (*CERN*), Asian Automatic Weather Station Network (*ANN*), Swiss FluxNet and several individual principal investigators (*PIs*) of *FLUXNET* network. These *EC* flux tower sites are mainly located in Europe, North America, and Asia, with three sites in Australia, four sites in South America and three sites in Africa (Figure 1). The sites cover 10 global plant functional types (*PFTs*): deciduous broadleaf forest (*DBF*, 28 sites), deciduous needleleaf forest (*DNF*, 2 sites), evergreen broadleaf forest (*EBF*, 13 sites), evergreen needleleaf forest (*ENF*, 54 sites),

mixed forest (*MIF*, 10 sites), cropland (*CRO*, 31 sites), grassland (*GRA*, 36 sites), savanna (woody savanna and savanna) (*SAW*, 8 sites), shrubland (open, closed) (*SHR*, 13 sites) and wetland (*WET*, 11 sites). These *EC* flux tower sites were separated into two separate subsets for the merging algorithm calibration (103 sites) and validation (103 sites), each representing major global *PFTs*. These *EC* data covered the period of 2000-2009 (each tower to varying extent) and cover at least one growing season.

Figure 1 HERE

The half-hourly and hourly *EC* measurements (*ET*; sensible heat flux, *H*) were processed based on a method of gap filling and quality control that used both the covariance of surface fluxes with meteorological parameters and the temporal variations in surface fluxes (Reichstein et al. 2005). These turbulent fluxes were complemented by measurements of R_n and soil heat flux (*G*). If less than 30% of the total data are missing, the daily values for R_n, *ET*, *H* and *G* were calculated the averages of the ground-measurements. Else, the daily value was set as invalid value (-9999). Because turbulent *EC* measurements are susceptible to incomplete energy balance closure (Leuning et al. 2012), we corrected the daily *ET* using the method proposed by Twine et al. (2000).

$$ET_c = (R_n - G)/(H_u + ET_u) \times ET_u \quad (1)$$

Accordingly, ET_c is the corrected *ET*, and H_u and ET_u are the uncorrected *H* and *ET*, respectively.

3. Methods

3.1 Simple Taylor skill fusion method

A simple Taylor skill fusion (*STS*) method is developed to merge the five Landsat-based *ET* products into a single *ET* product. The *STS* fusion method uses a weighted average of all the Landsat-based *ET* products and the weights are determined by their Taylor skill scores (*S*) (Taylor, 2001). The *S* value of each *ET* product is calculated using the direct *EC* ground-measurements for reference. The weights for all *ET* products sum up to one and the weights are proportional to the *S* values of the five *ET* products. Thus, the weights can be expressed as:

$$W_i = S_i / \sum_{i=1}^{n} S_i \quad (2)$$

$$S_i = \frac{4(1+R_i)^4}{(\delta_i + 1/\delta_i)^2 (1+R_{\max})^4} \quad (3)$$

Where W_i is the weight for *ET* product *i*, S_i is the Taylor skill score of *ET* product *i* and *n* is the number of *ET* products (*n*=5 in this study). R_i is the correlation coefficient between the estimated *ET* for product *i* and the *EC* ground-measured *ET*. $R_{\max}$ is the maximum correlation coefficient that is set to 1.0 in this study. δ_i is the ratio of the standard deviation of the estimated *ET* for product *i* to the standard deviation of the corresponding *EC* ground-measured *ET*. *S* varies from zero (least skillful) to one (most skillful). The *STS* method ensures that the merged *ET* product has the maximal R^2 and minimal error variance.

3.2 Assessment methods

R^2, *S*, the root-mean-square error (*RMSE*) and the bias are used to assess the performance of the *STS* method and the individual *ET* products. R^2 measures the agreement between satellite-based estimated and ground-measured *ET*. *RMSE*

characterizes the closeness of the estimation and observations and is calculated as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(E_j - O_j)^2} \quad (4)$$

Where N is the number of samples, E_j is the estimated *ET* for sample *j* and O_j is the *EC* ground-measured *ET* for sample *j*. The bias is also a metric to evaluate the predictive skill and it reflects the difference between the average of the estimation and observations.

$$Bias = \frac{1}{N}\sum_{j=1}^{N}(E_j - O_j) \quad (5)$$

4. Results and discussion

4.1 Validation of Landsat-based ET products

To assess the accuracy of the five Landsat-based *ET* products, the estimated *ET* from five Landsat-based *ET* products were directly compared with *EC* ground-measurements at all 206 flux tower sites for different *PFTs*. At the site scale, large differences were found in the five Landsat-based *ET* products among different *PFTs* (Figure 2). For both *DBF* and *MIF* sites, the *RS-PM ET* product has the largest *S* (>0.74) and R^2 (>0.73, p<0.01) compared to the other four *ET* products, whereas the *MS-PT ET* product has the smallest *RMSE* (29.4 W/m^2 for *DBF* and 20.4W/m^2 for *MIF*) and bias (<8W/m^2). For both *ENF* and *WET* sites, the *SW ET* product exhibits highest *S* (0.48 for *ENF* and 0.62 and *WET*) and the smallest *RMSE* (30.1 W/m^2 and 22.3 W/m^2). For all of the *SAW* sites, both *RS-PM* and *SW ET* products have the highest accuracy with an average *S* of 0.68 and average *RMSE* of approximately 20 W/m^2. The *MS-PT ET* product exhibits the highest *S* (>0.57) with an *RMSE* of less

than 25.5W/m^2, and R^2 (>0.65) with a confidence level of $p < 0.05$ for both *EBF* and *CRO* sites. For *GRA* and *SHR* sites, the *SIM ET* product with the low *RMSE* of less than 26.3 W/m^2 presents the highest *S* (>0.59) and R^2 (>0.58, p<0.05) compared to others. Most *ET* products (excluding *SW ET* product) have the highest *S* and (more than 0.68) and R^2 (more than 0.66, p<0.05) with the lowest average *RMSEs*, less than 22 W/m^2 for all of the *DNF* sites, compared to those for the other land cover types. This may caused by a few samples for only 2 *DNF* sites artificially to highlight the good performance of these *ET* products. When we selected equal samples for all of the land cover types, most *ET* products still provide better fits to the flux tower observations for *DNF* sites. According to the *S* values, the accuracies of both *MS-PT* and *SIM ET* products are the highest for all *ENF* sites. Although the *PT-JPL ET* product does not result in the highest *S*, it still has good accuracy for the variety of vegetation types. Moreover, we found that none of the individual Landsat-based *ET* dataset provides the most accurate *ET* estimate for all land cover types.

Figure 2 HERE

Overall, the five Landsat-based *ET* products account for 52-57% of *ET* variability over all *EC* measurements (Figure 3). Both *MS-PT* and *SIM ET* products have the highest *S* of 0.59, followed by *RS-PM*, *SW* and *PT-JPL* with *S* values ranging from 0.55 to 0.58. This indicates that different algorithm parameterizations affect the accuracy of different *ET* products. The highest accuracy of both *MS-PT* and *SIM ET* products may be a result from the lower uncertainties in the required in the lower number of forcing data (Wang and Liang, 2008; Yao et al. 2013). In addition,

the algorithm for the *SIM ET* product is strongly related to the regression coefficients because it was calibrated over the Southern Great Plains (*SGP*) area of the United States. Therefore, the *SIM ET* product estimated from this algorithm provides a better fit to flux tower observations. In contrast, the *ET* products (*e.g. RS-PM* and *SW*) are produced using resistance-based methods. The surface resistance parameterization schemes of these methods are complex, which also affected the accuracy of these *ET* products (Ershadi et al. 2014; Zhu et al. 2016). All five Landsat-based products slightly underestimate *ET* compared to the measurements, which can be explained by the fact that the algorithms for producing these products are originally developed based on the *MODIS* data.

Figure 3 HERE

4.2 Merging of the Landsat-based ET products

4.2.1 Calibration against tower measurements

To merge five Landsat-based *ET* products according to the *STS* method, the data collected at the 103 merging algorithm calibration sites were considered as calibration data to determine weights for the individual *ET* products. Figure 4 presents the weights for the five Landsat-based *ET* products based on *EC* ground-measurements. Both *MS-PT* and *SIM ET* products have the highest weight of 20.5%, followed by *RS-PM* (20.2%) and *SW* (19.5%) *ET* products. In contrast, the weight for *PT-JPL ET* products is only 19.3%, indicating its contribution to the merged *ET* estimates is the smallest than for those of the other products. Although previous studies showed that the *PT-JPL* algorithm driven by the *MODIS* vegetation variables had the best

performance compared to the *PM* algorithm (Ershadi, et al. 2014; Yao et al. 2013), our result illustrates that the *PT-JPL* algorithm driven by the Landsat vegetation variables has the worst performance compared to other four algorithms (*RS-PM*, *SW*, *MS-PT* and *SIM*). This may be caused by the spectral reflectance difference between Landsat and *MODIS* data, which resulted in the difference between Landsat and *MODIS* vegetation variables (*e.g. NDVI*) (Jia, et al., 2012).

Figure 4 HERE

The statistical summaries of the *STS* method performance for the 103 calibration sites among different land cover types are plotted in Figure 5. One can notice that the *STS*-based *ET* estimation for different land cover types have higher *S* and lower *RMSEs* compared to the individual Landsat-based *ET* products at the site scale. For 27 *ENF* calibration sites, the *STS* method has a higher *S* of 0.47 and R^2 of 0.43 (p<0.05) and a lower *RMSE* of 27.8 W/m^2 than individual *ET* products, though it presents the worst performance than those at other land cover types. For 6 *WET* calibration sites, the *STS* method shows better performance than single *ET* products, with lower *RMSEs* of 22.4 W/m^2 and higher *S* of 0.68. Similarly, the *RMSEs* of the *STS*-based *ET* estimation versus *EC* ground-measurements at 70 other calibration sites are all less than 25W/m^2 and the *S* values are all more than 0.7. Figure 6 compares daily *ET* observations at all 103 calibration sites and *ET* estimates for the different *ET* products. The results show that the merged *ET* product has the best performance with the highest *S* of 0.66 and the lowest *RMSE* of 23.8W/m^2, compared to the other products. This indicates that a reasonable choice of fusion method is

feasible to improve the accuracy of Landsat-based *ET* estimation by combining *EC* ground-measurements and multiple *ET* products.

Figure 5 HERE

Figure 6 HERE

4.2.2 Validation against EC measurements

The performance of the *STS* method for the validation tower set is similar to the results from the merging algorithm calibration tower set (Figure 7), though the *S* and R^2 statistics for the partial validation set (*e.g. DBF*, *ENF*) are slightly smaller than the merging algorithm calibration set. The *RMSE* of the merged daily *ET* for different land cover types varies from 11.2 W/m^2 to 32.4W/m^2, the R^2 varies from 0.45 to 0.86, and the *S* is greater than 0.7 (excluding *ENF* and *WET*). The merged *ET* decreases the *RMSE* by ~2 W/m^2 for forests, cropland and grassland sites and ~3 W/m^2 for other sites, and increases the *S* by approximately 0.03 and increased the R^2 by more than 0.02 (p<0.05).

Figure 7 HERE

Figure 8 illustrates a time series for clear-sky daily ground-measurements and estimated *ET* from multiple datasets for different land cover types. In comparison to the single Landsat-based *ET* products, the merged *ET* based on the *STS* fusion method produces seasonal *ET* variations that are closest to the ground-observed values. Overall, the *S* (R^2) and *RMSE* of the merged *ET* were approximately 5% higher and 8% lower than those of the best Landsat-based *ET* product, respectively (Figure 9). The *STS* fusion method reduced the errors of the estimated *ET* by

adjusting the weights of the single products through incorporation of *EC* ground-measurements.

Figure 8 HERE

Figure 9 HERE

4.2.3 Implementation of merging the ET products

To merge the five Landsat-based *ET* products to generate a new *ET* dataset, we obtained the weights of the *STS* method using five *ET* products and all *EC* ground-measurements. Figure 10 demonstrates the weights of different Landsat-based *ET* products when merging the *ET*, which is similar to the weights from the merging algorithm calibration tower subset. One notices that the relative contributions differ considerably for five Landsat-based *ET* products. The greatest contributors to the merged *ET* are both *MS-PT* and *SIM*, contributing 20.5%, followed by *RS-PM* (20.1%), *SW* (19.8%) and *PT-JPL* (19.1%).

Figure 10 HERE

Figure 11 shows scatter plots of a comparison between daily merged and ground-measured *ET* using ground observation data at all 206 flux tower sites. The *RMSE* of the merged *ET* at the site scale is 25.1W/m^2, the bias is 5.4W/m^2, the *S* is 0.61 and the R^2 is 0.60 ($p<0.05$). In comparison to the other five Landsat-based *ET* products, the merged *ET* product yields the highest accuracy (Figure 3). Figure 12 illustrates that the error histograms for the single Landsat-based *ET* products are more biased compared to the *EC* observations, whereas the merged *ET* using the *STS* method is more centered around zero. A substantial number of previous studies

reported that the errors of the estimated *ET* from remotely sensed data is approximately 15-30% (Jung et al. 2010; Wang and Dickinson, 2012; Yao et al. 2014) and the overall error of the merged *ET* based on the *STS* method is approximately 10%. Therefore, the accuracy of the merged *ET* in this study can be applied to produce global terrestrial Landsat-based *ET* product.

Figure 11 HERE

Figure 12 HERE

4.3 A case study of mapping regional ET

To map the regional *ET* from the Landsat-based *ET* products, we selected an example from the Landsat data of a 1.4 by 1.2 km region (33.77°N-33.88°N and 117.94°E-118.09°E) that mainly included cropland to map daily *ET* (Figure 13). Figure 13 also shows the corresponding spatial patterns in *NDVI* for August 12, 2005, along with the associated frequency histogram. High vegetation cover fraction occurred on August 12 owing to rapid crop growth.

Figure 13 HERE

In Figure 14, the spatial pattern of *ET* from each product is illustrated along with a histogram showing the frequency distribution of values within the simulation domain. The maps of *ET* are strongly positively correlated with the *NDVI* (R^2 of more than 0.91), which may be explained by the fact that higher vegetation transpiration where there is a higher vegetation fractional cover. In terms of overall magnitude and spatial pattern, there are obvious differences among the multiple Landsat-based *ET* products. In general, the merged *ET* has an intermediate *ET* value with a histogram of spanning

a full range from 60 to 82 W/m^2, which is slightly smaller than the *RS-PM*, *PT-JPL* and *SW ET* products whereas slightly larger than both *MS-PT* and *SIM ET* products. The difference in spatial pattern of these *ET* products was mainly caused by the different physical structures of *ET* algorithms, such as the physical parameterizations of the *SW* algorithm, affecting its coupling with land surface and atmosphere (Dirmeyer et al. 2013).

Figure 14 HERE

4.4. Discussion

4.4.1 Uncertainties of the merged ET estimates

4.4.1.1 Input errors

The varied accuracies of the merged *ET* product were affected by the input errors of the *STS* fusion method, which refers to the errors from the individual Landsat-based *ET* products and *EC* ground-measurements. The individual Landsat-based *ET* products are estimated using the meteorological variables from *MERRA* data and vegetation parameters derived from Landsat data. Previous studies showed that no single reanalysis dataset is superior to others in terms of meteorological variables (T_a, *RH*, *e* and *WS*) to estimate land surface energy budgets (Shi and Liang, 2014; Wang and Zeng, 2012; Zhu et al. 2012). Recent studies revealed large bias for *MERRA* data when compared to ground-measurements (Rienecker et al. 2011; Zhao et al. 2006). Yao et al. (2015) found that daily R_n from *MERRA* tended to underestimate at high values compared to ground-measurements. In addition, there also exist large biases in the vegetation parameters (e.g. *LAI*) retrieved

from Landsat data (Ganguly et al. 2012). Eklundh et al. (2003) reported that Landsat data can only explain 50-80% of the variation in *LAI* for coniferous forests. Thus, the uncertainty from the individual and merged *ET* products could be inherited through errors from both *MERRA* and Landsat data inputs.

The errors of the *EC* ground-measurements determine the accuracy of the merged *ET* product because ground-measured *ET* is considered as the "true" value for calibrating the individual products. Although *EC* measurements are relatively accurate for *ET* acquisition, approximately 5-20% still exist (Foken, 2008). Moreover, there could be inaccuracies in interpreting their values owing to the energy imbalance in the *EC* method (Mahrt, 2010). Foken (2008) pointed out that the *EC* method can only capture small eddies and ignore large eddies in the lower boundary layer, which influence the energy imbalance. Although we used the method proposed by Twine et al. (2000) to correct the *ET*, currently no agreements or protocols have been reached for the causes and corrections of energy imbalance from eddy covariance measurement (Leuning et al. 2012; Wohlfahrt et al. 2009). These corrections still cause large errors of *EC* measurements (Finnigan et al. 2003; Twine et al. 2000). Thus, input errors of the *EC* measurements and error propagation through calculations, including *EC* data correction, gridded interpolation and different data fusion, all contribute to the uncertainties of the merged *ET* product.

4.4.1.2 Scaling effects

The error of the merged *ET* product introduced by the spatial mismatch between the flux tower site footprints and the individual Landsat-based *ET* pixel footprints is an

important issue. The footprint of the flux tower site is approximately several hundred meters while the spatial resolution of the individual Landsat-based *ET* products is only 30m (Baldocchi et al. 2008; McCabe and Wood, 2006). Directly using *EC* ground-measured *ET* as "true" value to merge the individual Landsat-based *ET* products would lead to large uncertainties in the merged *ET* estimates.

To investigate the impact of the resample scale of the individual Landsat-based *ET* products to the accuracy of the merged *ET*, we averaged the daily *ET* from different Landsat-based *ET* products by use of a 30-570m window. Compared with the original *ET* estimates at 30m, a substantial drop in *ET* estimation errors occurred when products were aggregated to slightly coarser resolutions (Figure 15). With an increase in window size, a rise in estimation errors appeared owing to the surface heterogeneity in *ET*. When spatial resolution arrived at 450m, which is much larger than that of *TM* (30m), the individual and merged *ET* products had the smallest *RMSE*. This supports the expectation that differences in the resample scales owing to the higher surface heterogeneity directly influences the accuracy of the merged *ET*.

Figure 15 HERE

4.4.1.3 Fusion method

In the case of high-level data fusion for terrestrial *ET*, the *STS* method constrains the error of the fused *ET* by introducing *EC* ground-measurements to adjust the weights for the individual Landsat-based *ET* products. To quantify the errors of the different fusion methods, experiments with the same inputs have been performed for different fusion models, including Multiple Linear Regression (*MLR*), Simple Model

Averaging (*SMA*), Bayesian Model Averaging (*BMA*) (Raftery et al. 1997), Supported Vector Machine (*SVM*) (Vapnik, 1999), Multivariate Adaptive Regression Splines model (*MARS*) (Friedman, 1991), Random Forest Regression (*RFR*) (Breiman, 2001) and the *STS* method. The results of the leave-one-out cross-validation illustrated that the largest absolute differences in *RMSE* and R^2 of *ET* between the *STS* method and other fusion methods are relatively low, by approximately 1.0 W/m^2 and 0.02 respectively (Table 2). Further, the *STS* produced comparable accuracy but reduced the complexity of the fusion algorithm to improve computational efficiency when compared with other advanced fusion methods, indicating that the *STS* method can effectively achieve the goal of *ET* products integration in this study.

Table 2 HERE

Although the *STS* method might have the statistical significance to a certain degree, it obviously lacks of physical mechanism. The *STS* method only considers the combinations of different algorithms or datasets and it does not improve the satellite-based retrieval algorithm itself in essence (Taylor, 2001; Yao et al. 2014). Therefore, the performance of the *STS* method is highly dependent on the weightings for individual datasets, which was calibrated using the data from a lot of flux tower sites. Our next step is to develop a novel physical-based fusion algorithm by combining the residual of surface energy balance method and the water balance equation to produce *ET* product for regional application.

4.4.2 Implications for agricultural water consumption and

global models assessment

Quantifying *ET* using Landsat data is critical for mapping regional-scale *ET* at relatively high spatial resolution, acknowledging agricultural and watershed water management (Anderson et al. 2008). The merged *ET* product in this study was estimated using Landsat *NDVI* without *LST*. *NDVI* change relatively slowly when compared with surface moisture conditions characterized in the *LST*, and sampling frequency may be less of an issue (Anderson et al. 2012). Thus, these *VNIR*-based methods to *ET* mapping have some practical advantages over *TIR*-based methods (Glenn et al. 2011). However, *NDVI* offers no information about bare soil evaporation after crops have senesced (Anderson et al. 2008). Fortunately, meteorological relative humidity (*RH*) and diurnal air temperature range (*DT*) can effectively characterize the soil water deficit (Fisher et al. 2008; Wang and Liang, 2008). Recent studies indicate *RH* was superior to other water stress metrics (including soil water content and *VPD*) in regional *ET* estimation (Yan and Shugart, 2010). In particular, for the period of the main winter wheat growing season, *ET* derived by *VNIR*-based methods demonstrated their reliability to characterize agricultural moisture conditions at a regional scale (Zhang et al. 2016). This suggests that the merged *ET* in this study can be used to identify climatically sensitive agricultural systems and provide a diagnostic assessment of agricultural crop water consumption.

Accurate estimates of global terrestrial *ET* will be important for understanding global energy, water and carbon cycles. However, current both coarse-resolution global *ET* products estimated by remote sensing or global climate models (*GCMs*)

have not been well validated owing to the sparse ground-measurements, complicated surface characteristics and the spatial mismatch between the flux tower site footprints and the coarse-resolution global *ET* products footprints (Anderson et al. 2012; Yebra et al. 2013; Yan et al. 2012). Like *TIR*-derived *ET* products, the merged *ET* product with high spatial resolution provides a reference dataset for evaluating and validating coarse-resolution global *ET* products (e.g., *GCMs*) because it provides a bridge between the tower flux footprint scale (several decades and hundreds meters) and the grid scale of coarse-resolution global *ET* products with several hundred kilometers. Further studies should focus on spatial upscaling of the merged *ET* product in this study to evaluate the different coarse-resolution global *ET* products.

5. Conclusions

We described a simple Taylor skill (*STS*) fusion method that merged five Landsat -based *ET* products produced by individual algorithms and *FLUXNET* eddy covariance (*EC*) observations for improving terrestrial *ET* estimation. These five Landsat -based *ET* products were also evaluated based on the globally distributed *FLUXNET EC* observations. We found that at the site scale, large differences were found in the five Landsat-based *ET* products among different plant functional types.

According to the *STS* method, weights for the individual *ET* products were calibrated based on the data collected at the 103 merging algorithm calibration sites and the results show that the merged *ET* product has the best performance compared to the individual products. The performance of the *STS* method for the validation tower subset was similar to the results from the merging algorithm calibration tower

subset, though the *S* and R^2 statistics for the partial validation set are slightly smaller than the merging algorithm calibration set.

The weights of the *STS* method using five *ET* products and all *EC* ground-measurements were used to map the regional *ET*. An example of regional ET mapping shows that the *STS*-merged *ET* provides valuable insights for agricultural *ET* estimation. Uncertainties of the *STS*-merged *ET* are also discussed. The merged *ET* product presented in this study provides the bridge between the tower flux footprint scale and the grid scale of coarse-resolution global *ET* products. However, the *STS* method obviously lacks of physical mechanism. Our next step is to develop a novel physical-based fusion algorithm by combining the residual of surface energy balance method and the water balance equation to produce *ET* product for regional application.

Acknowledgements

This work 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, Swiss FluxNet, 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 *GAME AAN* (http://aan.suiri.tsukuba.ac.jp/), the Arid/Semi-arid experimental observation synergy and integration of northern China (http://observation.tea.ac.cn/), and the water experiments of Environmental and Ecological Science Data Center for West China (http:// westdc.westgis.ac.cn/water). This work was partially supported by the Natural Science Fund of China (No. 41671331), the National Key Research and Development Program of China (No.2016YFA0600102) and the National Science Foundation Division of Earth Sciences Award #1255013 (S.A.P.). JBF contributed to this manuscript from the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration; he was supported, in part, by the NASA Science Utilization of the Soil Moisture Active-Passive Mission (SUSMAP) program.

Appendix A. Algorithms for Landsat-based ET products

Appendix A.1.RS-PM algorithm

The revised remote sensing-based *PM* (*RS-PM*) algorithm is developed based on the Mu et al. (2011) algorithm, which is revised from the *PM* equation and it can be computed as follows:

$$ET = \frac{\Delta(R_n - G) + \rho C_p VPD / r_a}{\Delta + \gamma(1 + r_s / r_a)} \quad \text{(A1)}$$

Where Δ is the slope of the saturation water vapor pressure curve (P_a/K); γ is the psychrometric constant (P_a/K); ρ is the density of the air (k/gm^3); VPD is the vapor pressure deficit (P_a); r_a is the aerodynamic resistance (s/m) and r_s is the surface resistance (s/m). r_a calculation is described in Mu et al. (2011). For calculating r_s, Mu et al. (2011) calculated the temperature and moisture constraints for stomatal conductance using different parameters among various ecosystem types. In this study, we revised the temperature constraint (m_T) with an optimum air temperature (T_{opt}) set at 25°C (Fisher et al. 2008; Yao et al. 2013; Yuan et al. 2010).

$$m_T = \exp[-(\frac{T_a - T_{opt}}{T_{opt}})^2] \quad \text{(A2)}$$

We also revised the moisture constraint (Mu et al. 2007) (m_{VPD}) by setting VPD_{close} and VPD_{open} as 650 Pa and 2900 Pa for all ecosystem types, respectively.

$$m_{VPD} = \begin{cases} 1.0 & VPD \le VPD_{open} \\ \dfrac{VPD_{close} - VPD}{VPD_{close} - VPD_{open}} & VPD_{open} < VPD < VPD_{close} \\ 0.1 & VPD \ge VPD_{close} \end{cases} \quad \text{(A3)}$$

Where *close* refers to nearly complete inhibition and *open* refers to no inhibition to transpiration. Yuan et al. (2010) also found that it is possible to set invariant model parameters across different ecosystem types to reduce the effects of misclassification of land cover types.

Appendix A.2.SW algorithm

The Shuttleworth-Wallace (*SW*) algorithm is designed by combing two *PM* models for soil evaporation and vegetation transpiration (Shuttleworth and Wallace,

1985). The *SW* algorithm can be written as:

$$ET = C_s PM_s + C_v PM_v \quad \text{(A4)}$$

$$PM_s = \frac{\Delta(R_n - G) + (\rho C_p VPD - \Delta r_{as} R_{nc})/(r_{aa} + r_{as})}{\Delta + \gamma[1 + r_{ss}/(r_{aa} + r_{as})]} \quad \text{(A5)}$$

$$PM_v = \frac{\Delta(R_n - G) + [\rho C_p VPD - \Delta r_{ac}(R_{ns} - G)]/(r_{aa} + r_{ac})}{\Delta + \gamma[1 + r_{sc}/(r_{aa} + r_{ac})]} \quad \text{(A6)}$$

$$C_s = \frac{1}{1 + [R_s R_a/(R_c(R_s + R_a))]} \quad \text{(A7)}$$

$$C_v = \frac{1}{1 + [R_c R_a/(R_s(R_c + R_a))]} \quad \text{(A8)}$$

$$R_a = (\Delta + \gamma) r_{aa} \quad \text{(A9)}$$

$$R_s = (\Delta + \gamma) r_{as} + r_{ss}\gamma \quad \text{(A10)}$$

$$R_c = (\Delta + \gamma) r_{ac} + r_{sc}\gamma \quad \text{(A11)}$$

Where C_s and C_v (dimensionless) are the surface resistance coefficients for soil and vegetation, respectively. PM_s and PM_v are variables related to describe evaporation from soil and transpiration from vegetation, respectively. R_{ns} and R_{nc} are R_n into soil and vegetation, respectively. r_{aa} is aerodynamic resistances from vegetation canopy height to reference height. r_{as} and r_{ac} are aerodynamic resistances from the soil surface to canopy height and leaf to canopy height, respectively. r_{ss} and r_{sc} are the surface resistance for soil and vegetation, respectively. In general, r_{ss} is calculated using a function of the top layer of soil moisture (Sellers et al. 1992). In this study, we used RH^{VPD} to replace soil moisture to calculate r_{ss} because RH^{VPD} can be used to characterize soil moisture constraints for soil evaporation (Fisher et al. 2008) and it can be expressed as:

$$r_{ss} = \exp(8.206 - 4.255RH^{VPD}) \quad \text{(A12)}$$

Appendix A.3.PT-JPL algorithm

Based on the Priestley-Taylor algorithm, Fisher et al. (2008) developed the *PT-JPL* algorithm by downscaling potential *ET* (*PET*) to actual *ET* and it is written as:

$$ET = ET_s + ET_c + ET_i \quad \text{(A13)}$$

$$ET_s = \alpha[f_{wet} + (1 - f_{wet}) f_{sm}] \frac{\Delta}{\Delta + \gamma} (R_{ns} - G) \quad \text{(A14)}$$

$$ET_c = \alpha(1 - f_{wet}) f_g f_T f_M \frac{\Delta}{\Delta + \gamma} R_{nc} \quad \text{(A15)}$$

$$ET_i = \alpha f_{wet} \frac{\Delta}{\Delta + \gamma} R_{nc} \quad \text{(A16)}$$

Where ET_s refers to soil evaporation, ET_c refers to vegetation transpiration and ET_i refers to the canopy interception evaporation. a is the *PT* coefficient (1.26). f_{wet} is the wet surface fraction (RH^4). f_{sm} is the soil moisture constraint (RH^{VPD}). f_g is the green canopy fraction(f_{APAR}/f_{IPAR}). f_T is the plant temperature constraint (m_T) and $f_M$is the plant moisture constraint ($f_{APAR}/f_{APARmax}$). f_{APAR} is the absorbed photosynthetically active radiation (*PAR*) and f_{IPAR} is the intercepted *PAR*.

Appendix A.4.MS-PT algorithm

The modified satellite-based *PT* (*MS-PT*) algorithm developed by Yao et al. (2013) estimates *ET* by calculating the sum of the unsaturated soil evaporation (ET_{ds}), the saturated wet soil surface evaporation (ET_{ws}), the canopy transpiration (ET_v), and the canopy interception evaporation (ET_{ic}). The total *ET* can be expressed as:

$$ET = ET_{ds} + ET_{ws} + ET_v + ET_{ic} \quad \text{(A17)}$$

$$ET_{ds} = \alpha(1 - f_{wet}) f_{sm} \frac{\Delta}{\Delta + \gamma}(R_{ns} - G) \quad \text{(A18)}$$

$$ET_{s} = \alpha f_{wet} \frac{\Delta}{\Delta + \gamma}(R_{ns} - G) \quad \text{(A19)}$$

$$ET_{v} = \alpha(1 - f_{wet}) f_{c} f_{T} \frac{\Delta}{\Delta + \gamma} R_{nc} \quad \text{(A20)}$$

$$ET_{ic} = \alpha f_{wet} \frac{\Delta}{\Delta + \gamma} R_{nc} \quad \text{(A21)}$$

$$f_{sm} = (\frac{1}{DT})^{DT / DT_{\max}} \quad \text{(A22)}$$

$$f_{wet} = f_{sm}^{4} \quad \text{(A23)}$$

$$f_{c} = \frac{NDVI - NDVI_{\min}}{NDVI_{\max} - NDVI_{\min}} \quad \text{(A24)}$$

Where DT_{max} is the maximum diurnal air temperature range (40 °C) and f_c is vegetation cover fraction. $NDVI_{min}$ and $NDVI_{max}$ were the minimum and maximum *NDVI* during the study period, set as constants of 0.05 and 0.95 (Zhang et al. 2009) in this algorithm, respectively.

Appendix A.5.SIM algorithm

The simple hybrid ET (*SIM*) formulation was developed by Wang et al. (2008) based on satellite determination of surface net radiation, vegetation index, temperature, and *DT* and the *SIM* algorithm can be written as:

$$ET = R_n(a_0 + a_1 NDVI + a_2 T_a - a_3 DT) \quad \text{(A25)}$$

Where a_0=0.1440, a_1=0.6495, a_2=0.0090 and a_3=0.0163. These coefficients were calibrated using the ground measurements at the Southern Great Plains (*SGP*) sites in the United States from January 2002 to May 2005. Considering the *SGP* sites cover the variety of land cover that includes grass, rangeland, pastures, crop fields, forests,

and mixed cover-including vegetation and bare soil-and that their locations also differ considerably from each other, it can be used to estimate global terrestrial *ET* (Wang et al. 2008).

References

Allen, R.G., Tasumi, M., Trezza, R., 2007. Satellite-based energy balance formapping evapotranspiration with internalized calibration (METRIC)-model. J. Irrig. Drain. Eng. 133, 380-394.

Anderson, M.C., Allen, R.G., Morse, A., Kustas, W.P., 2012. Use of Landsat thermal imagery in monitoring evapotranspiration and managing water resources. Remote Sens. Environ. 122, 50-65.

Anderson, M.C., Norman, J.M., Diak, G.R., Kustas, W.P., Mecikalski, J.R., 1997. A two-source time-integrated model for estimating surface fluxes using thermal infrared remote sensing. Remote Sens. Environ. 60, 195-216.

Anderson, M.C., Norman, J.M., Kustas, W.P., Houborg, R., Starks, P.J., Agam, N., 2008. A thermal-based remote sensing technique for routine mapping of land-surface carbon, water and energy fluxes from field to regional scales. Remote Sens. Environ. 112, 4227-4241.

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.

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis,K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., Paw U, K.T., Pilegaard, K., Schmid, H.P., Valentini, R., Verma, S., Vesala, T., Wilson, K., Wofsy, S., 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.

Bastiaanssen, W., Menenti, M., Feddes, R., Holtslag, A., 1998. A remote

sensingsurface energy balance algorithm for land (SEBAL). 1. Formulation. J. Hydrol.212, 198-212.

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5-32.

Chen, Y., Xia, J., Liang, S., Feng, J., Fisher, J., Li, X., Li, X., Liu, S., Ma, Z., Miyata, A., Mu, Q., Sun, L., Tang, J., Wang, K., Wen, J., Xue, Y., Yu, G., Zha, T., Zhang, L., Zhang, Q., Zhao, T., Zhao, L., Yuan, W., 2014. Comparison of satellite-based evapotranspiration models over terrestrial ecosystems in China. Remote Sens. Environ. 140, 279-293.

Choi, M., Kustas, W.P., Anderson, M.C., Allen, R.G., Li, F., Kjaersgaard, J.H., 2009. An intercomparison of three remote sensing-based surface energy balance algorithms over a corn and soybean production region (Iowa, U.S.) during SMACEX. Agric. For. Meteorol. 149, 2082-2097.

Dirmeyer, P.A., Jin, Y., Singh, B., Yan, X., 2013. Trends in land-atmosphere interactions from CMIP5 simulations. J. Hydrometeorol. 14, 829-849.

Eklundh, L., Hall, K., Eriksson, H., Ardö, J., Pilesjö, P., 2003. Investigating the use of Landsat thematic mapper data for estimation of forest leaf area index in southern Sweden. Can. J. Remote Sens. 29, 349-362.

Ershadi, A., McCabe, M., Evans, J., Chaney, N., Wood, E., 2014. Multi-site evaluation of terrestrial evaporation models using FLUXNET data. Agric. For. Meteorol. 187, 46-61.

Feng, F., Li, X., Yao, Y., Liang, S., Chen, J., Zhao, X., Jia, K., Pinter, K., McCaughey, J.H., 2016. An empirical orthogonal function-based algorithm for estimating terrestrial latent heat flux from eddy covariance, meteorological and satellite observations. Plos One, 11, e0160150.

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

Foken, T., 2008. The energy balance closure problem: An overview. Ecol. Appl. 18, 1351-1367.

Friedman, J.H., 1991. Multivariate adaptive regression splines. Ann. Stat. 19, 1-67.

Ganguly, S., Nemani, R., Zhang, G., Hashimoto, H., Milesi, C., Michaelis, A., Wang, W., Votava, P., Samanta, A., Melton, F., Dungan, J.L., Vermote, E., Gao, F., Knyazikhin, Y., Myneni, R., 2012. Generating global Leaf Area Index from Landsat: Algorithm formulation and demonstration. Remote Sens. Environ. 122, 185-202.

Ghilain, N., Arboleda, A., Gellens-Meulenberghs, F., 2011. Evapotranspiration modeling at large scale using near-real time MSG SEVIRI derived data. Hydrol. Earth Syst. Sci. 15, 771-786.

Ghilain, N., Arboleda, A., Sepulcre-Canto, G., Batelaan, O., Ardo, J., Gellens-Meulenberghs, F., 2012. Improving evapotranspiration in a land surface model using biophysical variables derived from MSG/SEVIRI satellite. Hydrol. Earth Syst. Sci. 16, 2567-2583.

Glenn, E.P., Doody, T., Guerschman, J., Huete, A., King, E., McVicar, T., Van Dijk, A.,Van Niel, T., Yebra, M., Zhang, Y., 2011. Actual evapotranspiration estimation by ground and remote sensing methods: The Australian experience. Hydrol. Process. 25, 4103-4116.

Glenn, E. P., Morino, K., Didan, K., Jordan, F., Carroll, K. C., Nagler, P. L., Hultine, K., Sheader, L., Waugh, J., 2008. Scaling sap flux measurements of grazed and ungrazed shrub communities with fine and coarse-resolution remote sensing. Ecohydrology 1, 316-329.

Hope, A. S., Engstrom, R., Stow, D. A. 2005. Relationship between AVHRR surface temperature and NDVI in arctic tundra ecosystems. Int. J. Remote Sens. 26, 1771-1776.

Hu, G., Jia, L., Menenti, M., 2015. Comparison of MOD16 and LSA-SAF MSG evapotranspiration products over Europe for 2011. Remote Sens. Environ. 156, 510-526.

Hu, G., Jia, L., 2015. Monitoring of evapotranspiration in a semi-arid inland river

basin by combining microwave and optical remote sensing observations. Remote Sens. 7, 3056-3087.

Jiang, L., Islam, S., 1999. A methodology for estimation of surface evapotranspiration over large areas using remote sensing observations. Geophys. Res. Lett. 26, 2773-2776.

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.

Jiménez, C., Prigent, C., Mueller, B., Seneviratne, S. I., McCabe, M. F., Wood, E. F., Rossow, W. B., Balsamo, G., Betts, A. K., Dirmeyer, P. A., Fisher, J. B., Jung, M., Kanamitsu, M., Reichle, R. H., Reichstein, M., Rodell, M., Sheffield, J., Tu, K., Wang, K., 2011. Global intercomparison of 12 land surface heat flux estimates. J. Geophys. Res. Atmos. 116, D02102.

Jin, Y., Randerson, J., Goulden, M., 2011. Continental-scale net radiation and evapotranspiration estimated using MODIS satellite observations. Remote Sens. Environ. 115, 2302-2319.

Ju, J.C., Roy, D.P., 2008. The availability of cloud-free Landsat ETM plus data over the conterminous United States and globally. Remote Sens. Environ. 112, 1196-1211.

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S., Sheffield, J., Goulden, M., Bonan, G., Cescatti, A., Chen, J., Richard, D., Johannes Dolman, A., Eugster, W., Gerten, D., Gianelle, D., Gobron, N., Heinke, J., Kimball, J., Law, B., Montagnani, L., Mu, Q., Mueller, B., Oleson, K., Papale, D., Richardson, A., Roupsard, O., Running, S., Tomelleri, E., Viovy, N., Weber, U., Williams, C., Wood, E., Zaehle, S., Zhang, K., 2010. Recent decline in the global land evapotranspiration trend due to limited moisture supply. Nature. 467, 951-954.

Kalma, J., McVicar, T., McCabe, M., 2008. Estimating land surface evaporation: A review of methods using remotely sensed surface temperature data accomplished. Surv. Geophys. 29, 421-469.

Kim, H., Hwang, K., Mu, Q., Lee, S., Choi, M., 2012. Validation of MODIS 16

global terrestrial evapotranspiration products in various climates and land cover types in Asia. KSCE J. Civ. Eng. 16, 229-238.

Kool, D., Agama, N., Lazarovitch, N., Heitman, J.L., Sauer, T.J., Ben-Gal, A., 2014. A review of approaches for evapotranspiration partitioning. Agric. For. Meteorol. 184, 56-70.

Kumar, S.V., Peters-Lidard, C.D., Tian, Y., Houser, P.R., Geiger, J., Olden, S., Lighty, L., Eastman, J.L., Doty, B., Dirmeyer, P., Adams, J., Mitchell, K., Wood, E.F., Sheffield, J., 2006. Land information system-an interoperable framework for high resolution land surface modeling. Environ Modell. Softw. 21, 1402-1415.

Kustas, W., Anderson, M., 2009. Advances in thermal infrared remote sensing for land surface modeling. Agric. For. Meteorol. 149, 2071-2081.

Kustas, W., Norman, J., 1999. Evaluation of soil and vegetation heat fluxpredictions using a simple two-source model with radiometric temperatures for partial canopy cover. Agric. For. Meteorol. 94, 13-29.

Leuning, R., van Gorsel, E., Massman, W.J., Isaac, P.R., 2012. Reflections on the surface energy imbalance problem. Agric. For. Meteorol.156, 65-74.

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.

Liaqat, U.W., Choi, M., 2015. Surface energy fluxes in the Northeast Asia ecosystem: SEBS and METRIC models using Landsat satellite images. Agric. For. Meteorol. 214-215, 60-79.

Liu, S., Xu, Z., Song, L., Zhao, Q., Ge, Y., Xu, T., Ma, Y., Zhu, Z., Jia, Z., Zhang, F., 2016. Upscaling evapotranspiration measurements from multi-site to the satellite pixel scale over heterogeneous land surfaces, Agric. For. Meteorol. 230-231, 97-113.

Liu, S., Xu, Z., Wang, W., Jia, Z., Zhu, M., Bai, J., Wang, J., 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., Xu, Z., Zhu, Z., Jia, Z., Zhu, M., 2013. Measurements of evapotranspiration

from eddy-covariance systems and large aperture scintillometers in the Hai River Basin, China. J. Hydrol. 487, 24-38.

Li, Z.-L., 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.

Long, D., Singh, V., 2012. A two-source trapezoid model for evapotranspiration (TTME) from satellite imagery. Remote Sens. Environ. 121, 370-388.

Mahrt, L., 2010. Computing turbulent fluxes near the surface: needed improvements. Agric. For. Meteorol. 150, 501-509.

Mallick, K., Bhattacharya, B.K., Rao, V.U.M., Reddy, D.R., Banerjee, S., Venhatesh, H., Pandey, V., Kar, G., Mukherjee, J., P. Vyas, S., Gadgil, A.S., Patel, N.K., 2009. Latent heat flux estimation in clear sky days over Indian agroecosystems using noontime satellite remote sensing data. Agric. For. Meteorol. 149, 1646-1665.

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.

Monteith, J., 1965. Evaporation and environment. Symp. Soc. Exp. Biol. 19, 205-224.

Mu, Q., Heinsch, F., Zhao, M., Running, S., 2007. Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. Remote Sens. Environ. 114, 519-536.

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

Norman, J.M., Kustas, W.P., Humes, K.S., 1995. Source approach for estimating soiland vegetation energy fluxes in observations of directional radiometric surfacetemperature. Agric. For. Meteorol. 77, 263-293.

Priestley, C. H. B., Taylor, R. J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. Mon. Weather Rev. 100, 81-92.

Raftery, A. E., Madigan, D., Hoeting, J. A., 1997. Bayesian model averaging for linear regression models. J. Am. Stat. Assoc. 92, 179-191.

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grünwald, T., Havránková, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J.-M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., Valentini, R., 2005. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: Review and improved algorithm. Global Change Biol. 11, 1424-1439.

Rienecker, M.M, Suarez, M.J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M.G., Schubert, S.D., Takacs, L., Kim, G.K., Bloom, S., Chen, J.Y., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R.D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C.R., Reichle, R., Robertson, F.R., Ruddick, A.G., Sienkiewicz, M., Woollen, J., 2011. MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. J. Climate. 24, 3624-3648.

Rodell, M., Houser, P.R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J.K., Walker, J.P., Lohmann, D., Toll, D., 2004. The global land data assimilation system. Bull. Am. Meteor. Soc. 85, 381-394.

Sellers, P.J., Heiser, M.D., Hal, F.G., 1992. Relations between surface conductance and spectral vegetation indices at intermediate (100m^2 to 15 km^2) length scales. J. Geophys. Res. Atmos. 97(D17), 19033-19059.

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.

Shuttleworth, W.J., Wallace, J.S., 1985. Evaporation from sparse crops-An energy combination theory. Q. J. R. Meteorol. Soc. 111, 839-855.

Su, Z., 2002. The surface energy balance system (SEBS) for estimation of turbulent heat fluxes. Hydrol. Earth Syst. Sci. 6, 85-99.

Tang, R., Li, Z., Tang, B., 2010. An application of the Ts-VI triangle method with

enhanced edges determination for evapotranspiration estimation from MODIS data in arid and semi-arid regions: Implementation and validation. Remote Sens. Environ. 114, 540-551.

Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res. 106, 7183-7192.

Twine, T.E., Kustas, W.P., Norman, J.M., Cook, D.R., Houser, P.R., Meyers, T.P., Prueger, J.J., Starks, P.J., Wesely, M.L., 2000. Correcting eddy-covariance flux underestimates over a grassland. Agric. For. Meteorol.103, 279-300.

Uppala, S.M., KÅllberg, P.M., Simmons, A.J., Andrae, U., Da Costa Bechtold, V., Fiorino, M., Gibson, J.K., Haseler, J., Hernandez, A., Kelly, G.A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R.P., Andersson, E., Arpe, K., Balmaseda, M.A., Beljaars, A.C.M., Van De Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B.J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., Mcnally, A.P., Mahfouf, J.-F., Morcrette, J.-J, Rayner, N.A., Saunders, R.W., Simon, P., Sterl, A., Trenberth, K.E., Untch, A., Vasiljevic, D., Viterbo, P., Woollen, J., 2005. The ERA-40 re-analysis. Q. J. Roy. Meteor. Soc. 131, 2961-3012.

Vapnik, V., 1999. The Nature of Statistical Learning Theory, Information Science and Statistics, Springer, New York.

Wang, A., Zeng, X., 2012. Evaluation of multireanalysis products with in situ observations over the Tibetan Plateau. J. Geophys. Res. 117, D05102.

Wang, K., Dickinson, R. E., 2012. A review of global terrestrial evapotranspiration: observation, modeling, climatology and climatic variability. Rev. Geophys. 50, RG2005.

Wang, K., Liang, S., 2008. An improved method for estimating global evapotranspiration based on satellite determination of surface net radiation, vegetation index, temperature, and soil moisture. J. Hydrometeorol. 9, 712-727.

Wang, K., Wang, P., Li, Z.Q., Cribb, M., Sparrow, M., 2007. A simple method to estimate actual evapotranspiration from a combination of net radiation, vegetation index, and temperature. J. Geophys. Res. Atmos. 112, D15107.

Wohlfahrt, G., Haslwanter, A., Hortnagl, L., Jasoni, R., Fenstermaker, L., Arnone,J., Hammerle, A., 2009. On the consequences of the energy imbalance for calculating surface conductance to water vapour. Agric. For. Meteorol. 149, 1556-1559.

Xu, T., Liu, S., Liang, S., Qin, J., 2011. Improving predictions of water and heat fluxes by assimilating MODIS land surface temperature products into common land model. J. Hydrometeorol. 12, 227-244.

Xu, T., Bateni, S.M., Margulis, S.A., Song, L., Liu, S., Xu, Z., 2016. Partitioning evapotranspiration into soil evaporation and canopy transpiration via a two-source variational data assimilation system. J. Hydrometeorol. 17, 2363-2370.

Yan, H., Shugart, H., 2010. An air relative-humidity-based evapotranspiration model from eddy covariance data. J. Geophys. Res. Atmos.115, D16106.

Yan, H., Wang, S. Q., Billesbach, D., Oechel, W., Zhang, J.H., Meyers, T., Martin, T.A., Matamala, R., Baldocchi, D., Bohrer, G., Dragoni, D., Scott, R., 2012. Global estimation of evapotranspiration using a leaf area index-based surface energy and water balance model. Remote Sens. Environ. 124, 581-595.

Yao, Y., Liang, S., Cheng, J., Liu, S., Fisher, J., Zhang, X., Jia, K., Zhao, X., Qin, Q., Zhao, B., Han, S., Zhou, G.S., Zhou, G.Y., Li. Y., Zhao, S., 2013. MODIS-driven estimation of terrestrial latent heat flux in China based on a modified Priestly-Taylor algorithm. Agric. For. Meteorol. 171-172, 187-202.

Yao, Y., Liang, S., Li, X., Chen, J., Wang, K., Jia, K., Cheng, J., Jiang, B., Fisher, J., Mu, Q., Grünwald, T., Bernhofer, C., Roupsard, O., 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.

Yao, Y., Liang, S., Li, X., Hong, Y., Fisher, J., Zhang, N., Chen, J., Cheng, J., Zhao, S., Zhang, X., Jiang, B., Sun, L., Jia, K., Wang, K., Chen, Y., Mu, Q., Feng, F., 2014. Bayesian multimodel estimation of global terrestrial latent heat flux from eddy covariance, meteorological, and satellite observations. J. Geophys. Res.

Atmos. 119, 4521-4545.

Yao, Y., Liang, S., Li, X., Liu, S., Chen, J., Zhang, X., Jia, K., Xie, X., Munier, S., Liu, M., Yu, J., Lindroth, A., Varlagin, A., Raschi, A., Noormets, A., Pio, C., Wohlfahrt, G., Sun, G., Domec, J.C., Montagnani, L., Lund, M., Eddy, M., Blanken, P.D., Grünwald, T., Wolf, S., Magliulo, V., 2016. Assessment and simulation of global terrestrial latent heat flux by synthesis of CMIP5 climate models and surface eddy covariance observations. Agric. For. Meteorol. 223, 151-167.

Yebra, M., Van Dijk, A., Leuning, R., Huete, A., Guerschman, J.P., 2013. Evaluation of optical remote sensing to estimate actual evapotranspiration and canopy conductance. Remote Sens. Environ. 129, 250-261.

Yuan, W., Liu, S., Yu, G., Bonnefond, J.-M., Chen, J., Davis, K., Desai, A.R., Goldstein, A.H., Gianelle, D., Rossi, F., Suyker, A.E., Verma, S.B., 2010. Global estimates of evapotranspiration and gross primary production based on MODIS and global meteorology data. Remote Sens. Environ. 114, 1416-1431.

Zhang, J., Mu, Q., Huang, J., 2016. Assessing the remotely sensed Drought Severity Index for agricultural drought monitoring and impact analysis in North China. Ecol. Indic. 63, 296-309.

Zhang, K., Kimball, J.S., Mu, Q., Jones, L.A., Goetz, S.J., Running, S.W., 2009. Satellite based analysis of northern ET trends and associated changes in the regional water balance from 1983 to 2005. J. Hydrol. 379, 92-110.

Zhang, R. H., Sun, X. M., Wang, W. M., Xu, J. P., Zhu, Z. L., Tian, J., 2005. An operational two-layer remote sensing model to estimate surface flux in regional scale: Physical background. Sci. China, Ser. D: Earth Sci. 48, 225-244.

Zhao, M., Heinsch, F. A., Nemani, R., Running, S. W., 2005. Improvements of the MODIS terrestrial gross and net primary production global data set. Remote Sens. Environ. 95, 164-176.

Zhao, M., Running, S., Nemani, R., 2006. Sensitivity of Moderate Resolution Imaging Spectroradiometer (MODIS) terrestrial primary production to the accuracy of meteorological reanalysis. J. Geophys. Res. Atmos. 111, G01002.

Zhu, G., Li, X., Zhang, K., Ding, Z., Han, T., Ma, J., Huang, C., He, J., Ma, T., 2016. Multi-model ensemble prediction of terrestrial evapotranspiration across north China using Bayesian model averaging. Hydrol. Process. 30, 2861-2879.

Zhu, X., Liu, Y., Wu, G., 2012. An assessment of summer sensible heat flux on the Tibetan Plateau from eight data sets. Sci. China, Ser. D: Earth Sci. 55, 779-786.

Table 1 Summary of the Landsat *ET* products generated in this study for 2000-2009.

ET products	***Spatial resolution***	***Temporal resolution***	***Time Period***	***Algorithms***	***Forcing Inputs of the ET products***	
					MERRA	***Landsat***
RS-PM	30m	16-day	1998-2010	Remote sensing-based Penman-Monteith algorithm	R_n, *RH*, T_a, *e*	*NDVI*
SW	30m	16-day	1995-2009	Shuttleworth-Wallace dual-source model	R_n, *RH*, T_a, *e*, *WS*	*NDVI*
PT-JPL	30m	16-day	1998-2010	Priestley-Taylor algorithm of Jet Propulsion Laboratory, Caltech	R_n, *RH*, T_a, *e*	*NDVI*
MS-PT	30m	16-day	1997-2009	Modified satellite-based Priestley-Taylor algorithm	R_n, T_a, *DT*	*NDVI*
SIM	30m	16-day	1998-2009	Simple hybrid algorithm	R_n, T_a, *DT*	*NDVI*

Table 2 Comparison of the cross validation results of daily *ET* from multiple Landsat-based *ET* products and fused products. *S* represents the Taylor skill scores.

ET Products	***RMSE***	R^2	***S***
STS	25.4	0.60	0.61
MLR	26.2	0.59	0.60
SMA	26.9	0.58	0.59
BMA	24.9	0.61	0.62
SVM	24.2	0.63	0.63
MARS	24.4	0.62	0.62
BFR	24.1	0.64	0.64
RS-PM	27.9	0.56	0.58
SW	29.6	0.55	0.57
PT-JPL	29.8	0.52	0.55
MS-PT	27.1	0.57	0.59
SIM	27.6	0.57	0.59

Figure captions

Figure 1. Locations of the flux tower sites used to merge *ET* algorithm calibration (103 sites) and validation (103 sites).

Figure 2. Bar graphs of the statistics (*RMSE*, *Bias*, R^2 and *S*) of the comparison between daily *ET* from multiple Landsat-based *ET* products (including merged *ET* product) and ground-measurements at all 206 flux tower sites for different land cover types. *DBF*: deciduous broadleaf forest, *DNF*: deciduous needleleaf forest, *EBF*: evergreen broadleaf forest, *ENF*: evergreen needleleaf forest, *MIF*: mixed forest, *CRO*: cropland, *GRA*: grassland, *SAW*: woody savanna and savanna, *SHR*: open and closed shrubland, and *WET*: wetland.

Figure 3. Scatterplots of the daily *ET* from multiple Landsat-based *ET* products and ground-measurements at all 206 flux tower sites.

Figure 4. Weights for five Landsat-based *ET* products at the 103 calibration tower sites.

Figure 5. Same as Figure 2 but for the 103 calibration tower sites. The merged *ET* was calculated using the weights for five Landsat-based *ET* products at the 103 calibration tower sites.

Figure 6. Scatterplots of the daily *ET* from multiple Landsat-based *ET* products (including merged *ET* product) and ground-measurements at the 103 calibration tower sites.

Figure 7. Same as Figure 2 but for the 103 validation tower sites for different land cover types. The merged *ET* was calculated using the weights for five Landsat-based *ET* products at the 103 calibration tower sites.

Figure 8. Example of a time series for daily *ET* as ground-measured and estimated using different Landsat-based *ET* products (including merged *ET* product) at ten validation sites. The merged *ET* was calculated using the weights for five Landsat-based *ET* products at the 103 calibration tower sites.

Figure 9. Same as Figure 6 but for the 103 validation tower sites. The merged *ET* was calculated using the weights for five Landsat-based *ET* products at the 103 calibration tower sites.

Figure 10. Weights for five Landsat-based *ET* products at all 206 flux tower sites.

Figure 11. Scatterplots of the daily *ET* from the merged *ET* products and ground-measurements at all 206 flux tower sites. The merged *ET* was calculated using the weights for five Landsat-based *ET* products at all 206 flux tower sites.

Figure 12. Error histograms for daily *ET* derived from five Landsat-based *ET* products, and the merged *ET* product for all 206 flux tower sites.

Figure 13. a) An example of a partial region of Landsat imagery with a false-color composite on August 12,2005; b) *NDVI* maps for August 12,2005, and c) frequency histograms for *NDVI* on August 12,2005.

Figure 14. Daily *ET* maps of a partial region shown in Figure 13 with frequency histograms from five Landsat-based *ET* products, and the merged *ET* product for August 12, 2005.

Figure 14. Continued.

Figure 15. Change of *RMSE* of estimating daily *ET* from five Landsat-based *ET*

products, and the merged *ET* product with spatial resolutions at the 103 validation tower sites. The merged *ET* was calculated using the weights for five Landsat-based *ET* products at the 103 calibration tower sites.

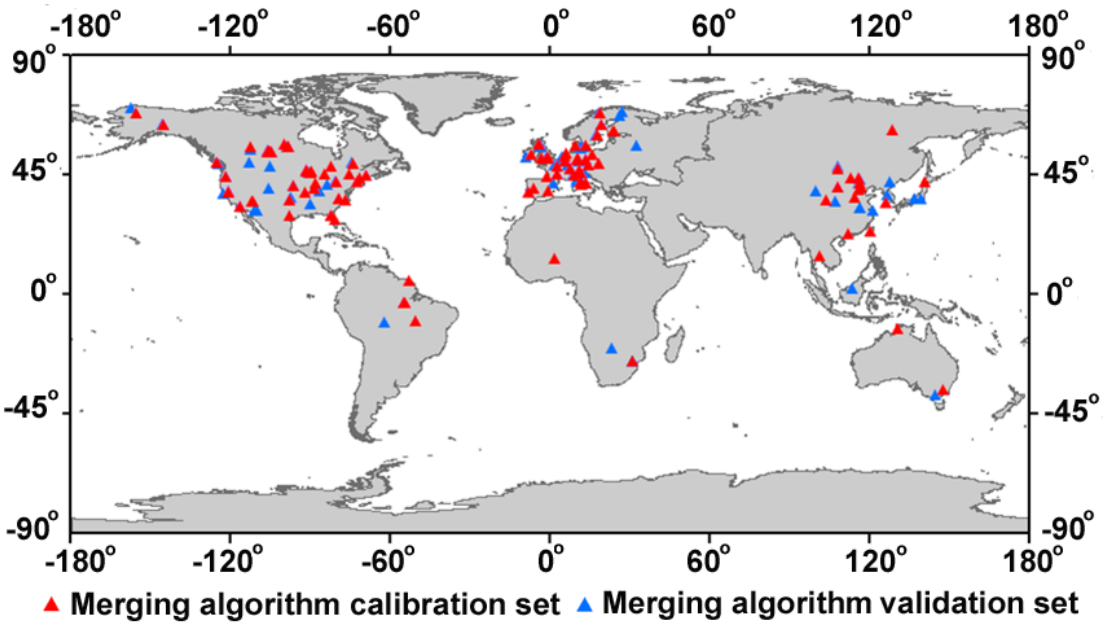
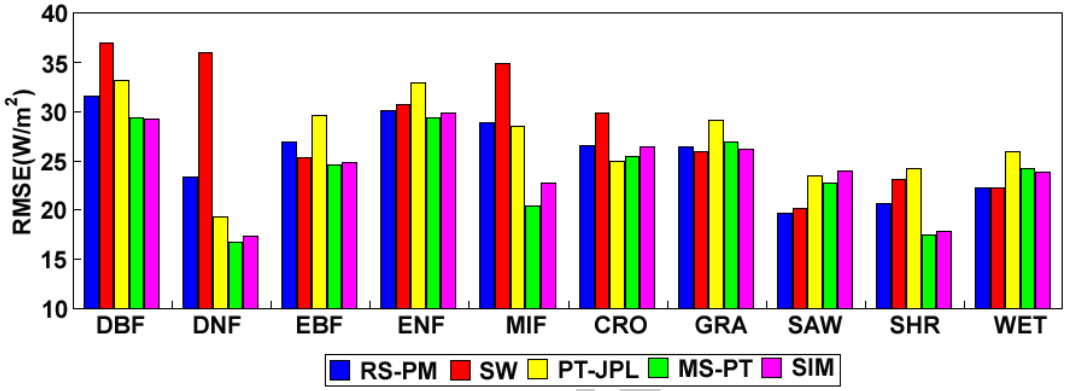
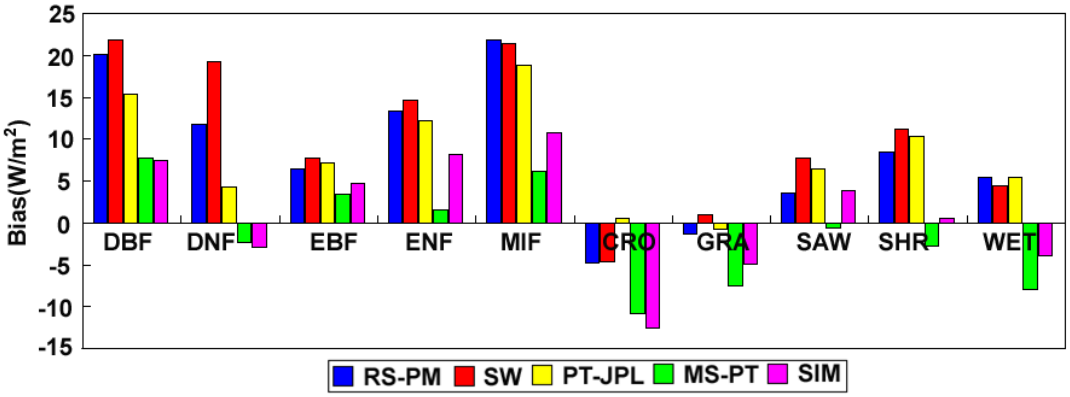
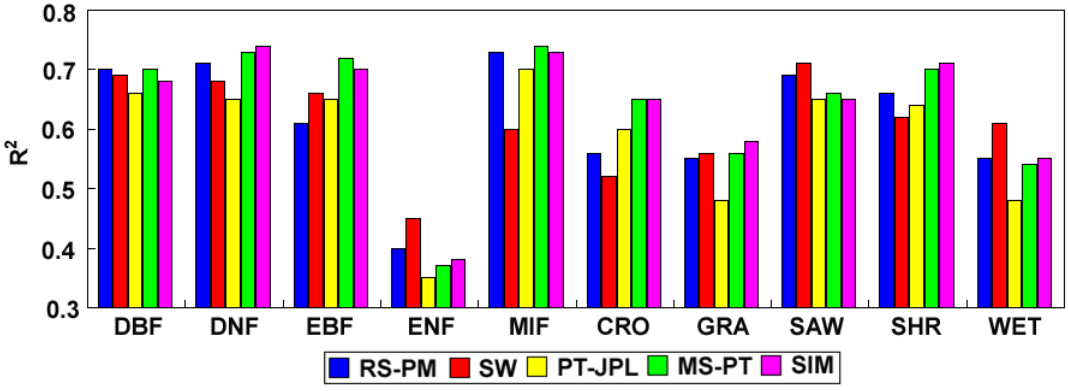


Figure 1. Locations of the flux tower sites used to merge *ET* algorithm calibration (103 sites) and validation (103 sites).







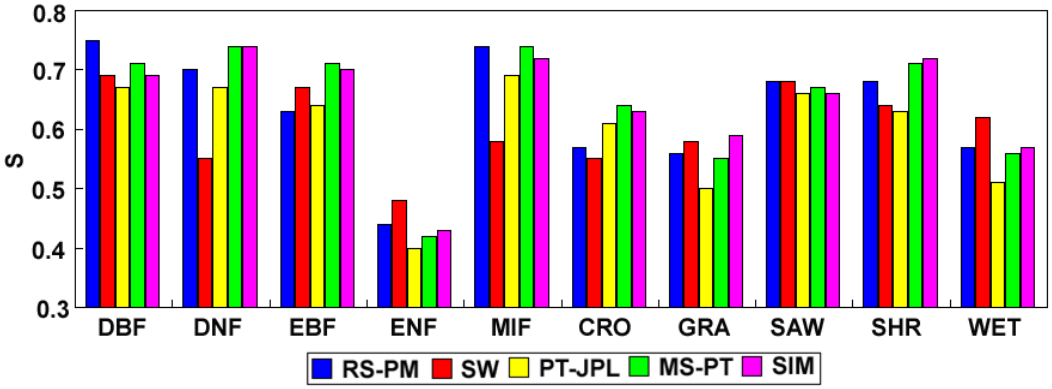
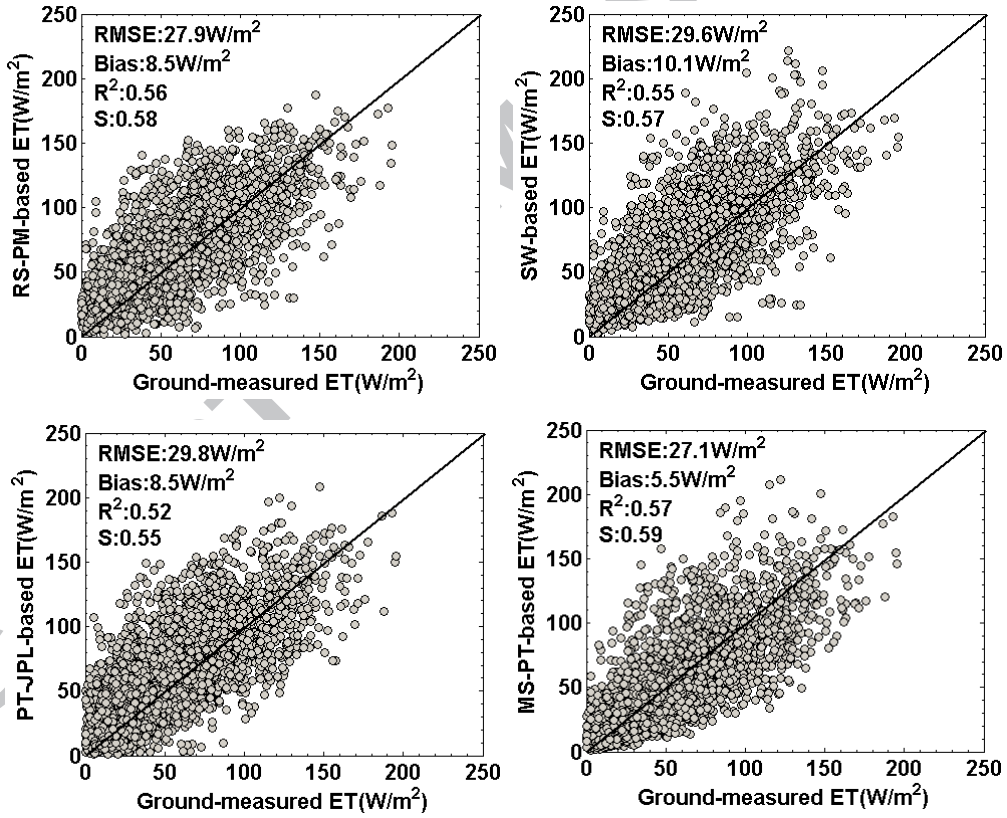


Figure 2. Bar graphs of the statistics (*RMSE*, *Bias*, R^2 and *S*) of the comparison between daily *ET* from multiple Landsat-based *ET* products (including merged *ET* product) and ground-measurements at all 206 flux tower sites for different land cover types. *DBF*: deciduous broadleaf forest, *DNF*: deciduous needleleaf forest, *EBF*: evergreen broadleaf forest, *ENF*: evergreen needleleaf forest, *MIF*: mixed forest, *CRO*: cropland, *GRA*: grassland, *SAW*: woody savanna and savanna, *SHR*: open and closed shrubland, and *WET*: wetland.



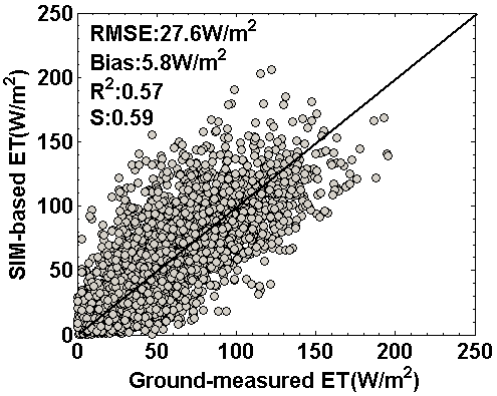


Figure 3. Scatterplots of the daily *ET* from multiple Landsat-based *ET* products and ground-measurements at all 206 flux tower sites.

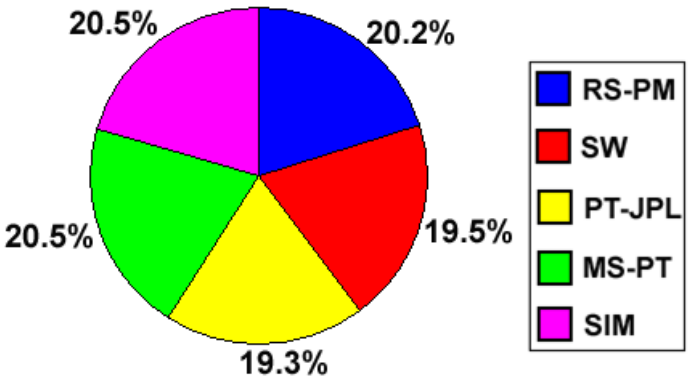
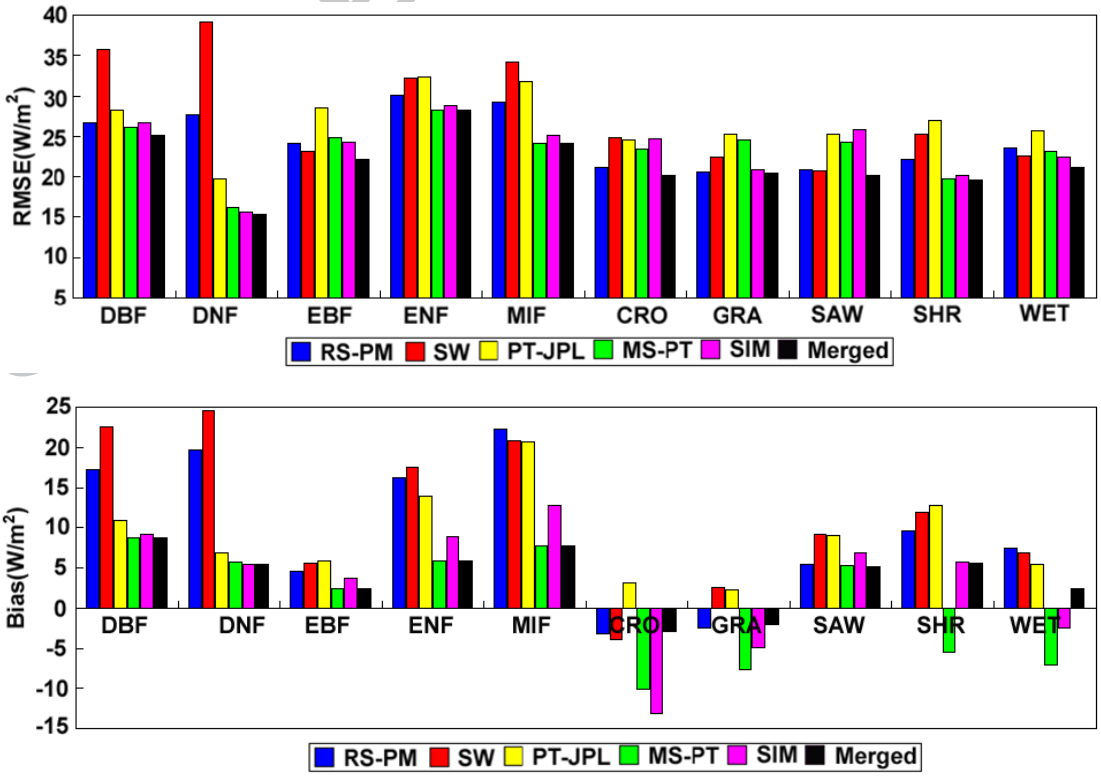


Figure 4. Weights for five Landsat-based *ET* products at the 103 calibration tower sites.



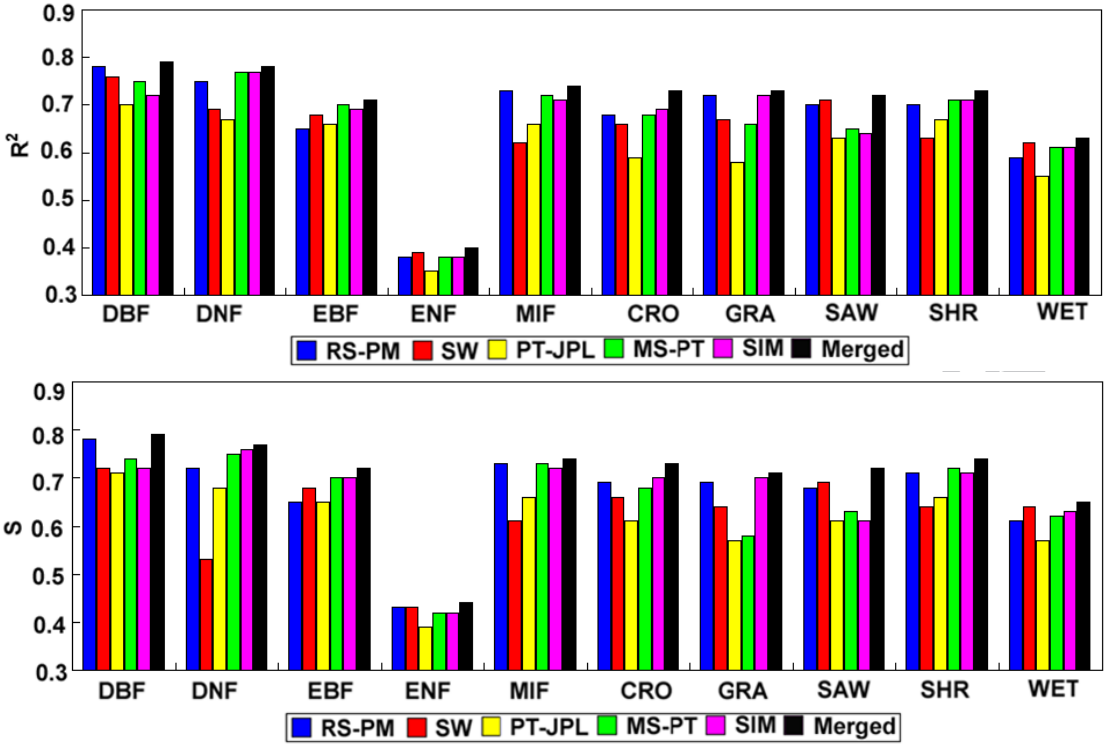
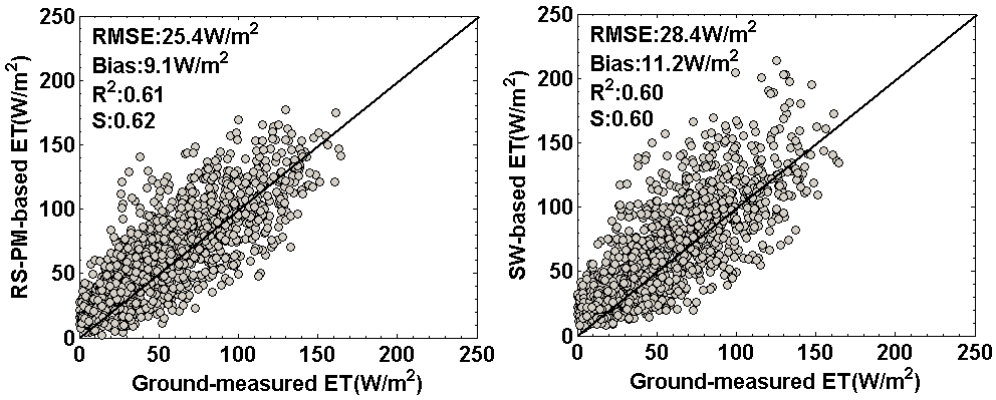


Figure 5. Same as Figure 2 but for the 103 calibration tower sites. The merged *ET* was calculated using the weights for five Landsat-based *ET* products at the 103 calibration tower sites.



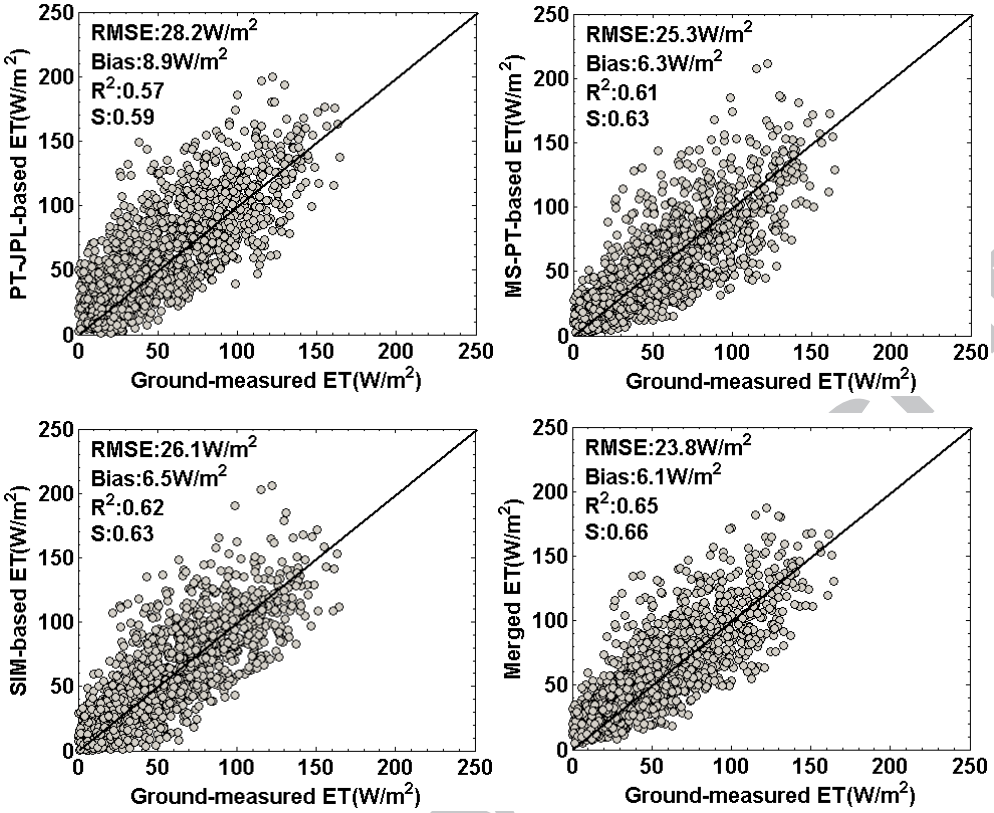
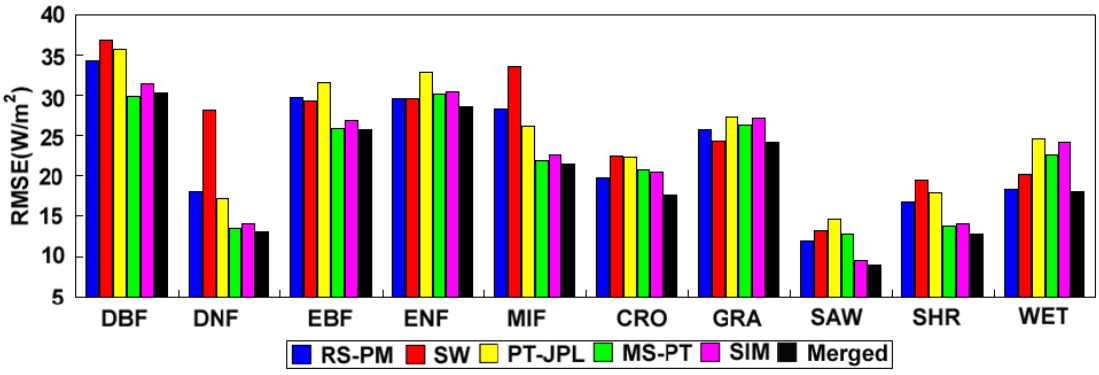


Figure 6. Scatterplots of the daily *ET* from multiple Landsat-based *ET* products (including merged *ET* product) and ground-measurements at the 103 calibration tower sites.



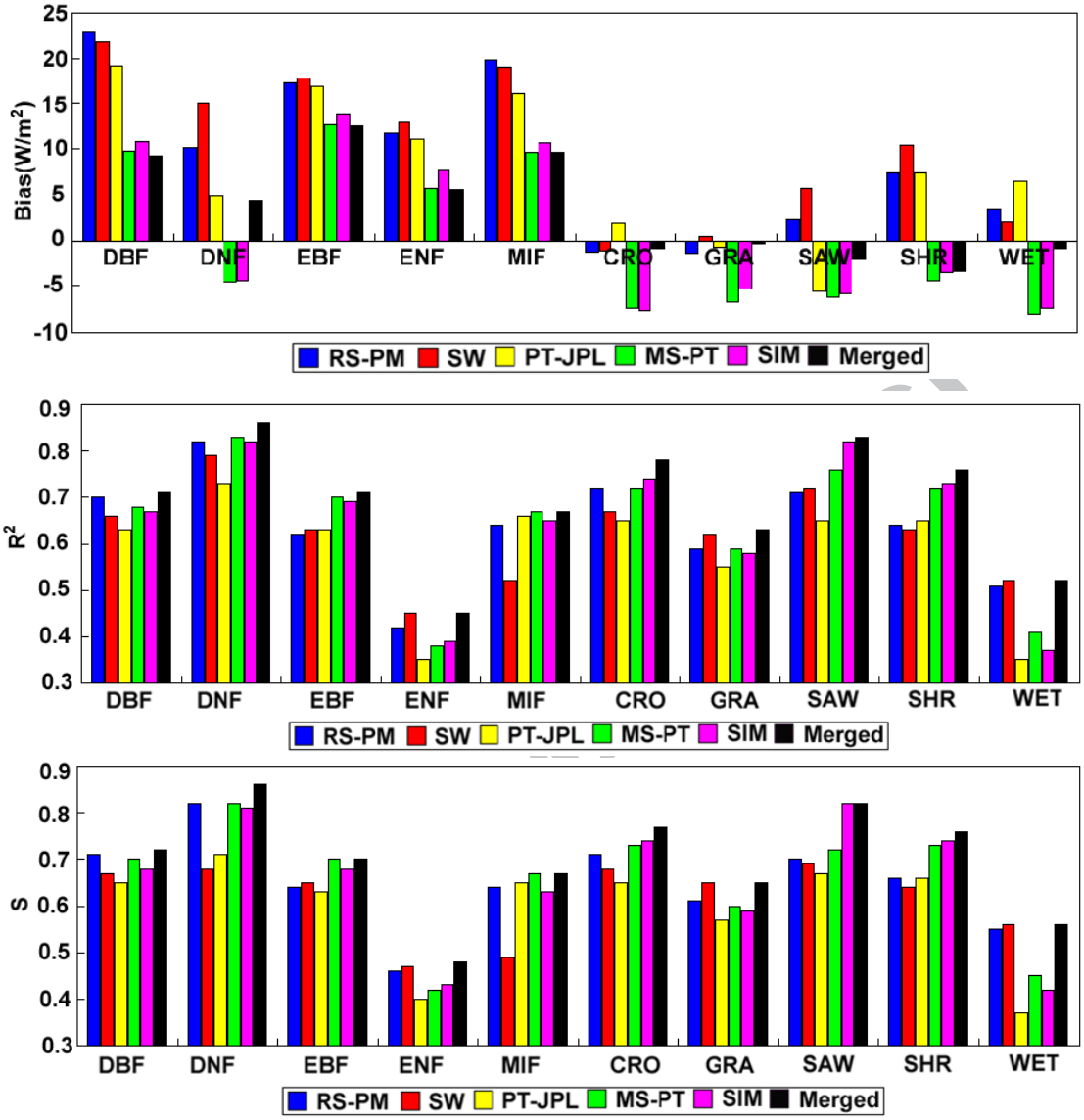
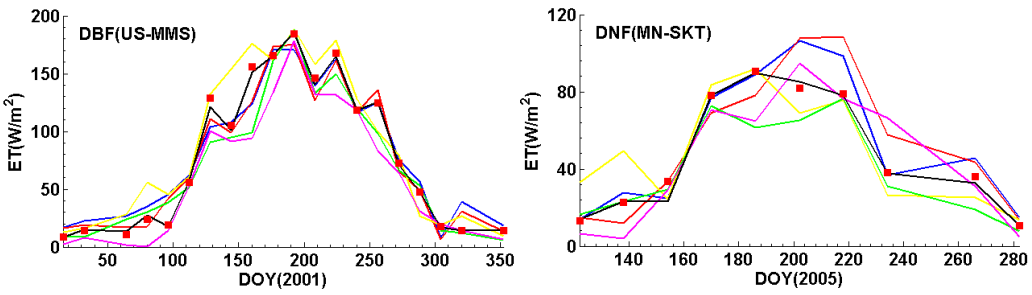


Figure 7. Same as Figure 2 but for the 103 validation tower sites for different land cover types. The merged *ET* was calculated using the weights for five Landsat-based *ET* products at the 103 calibration tower sites.



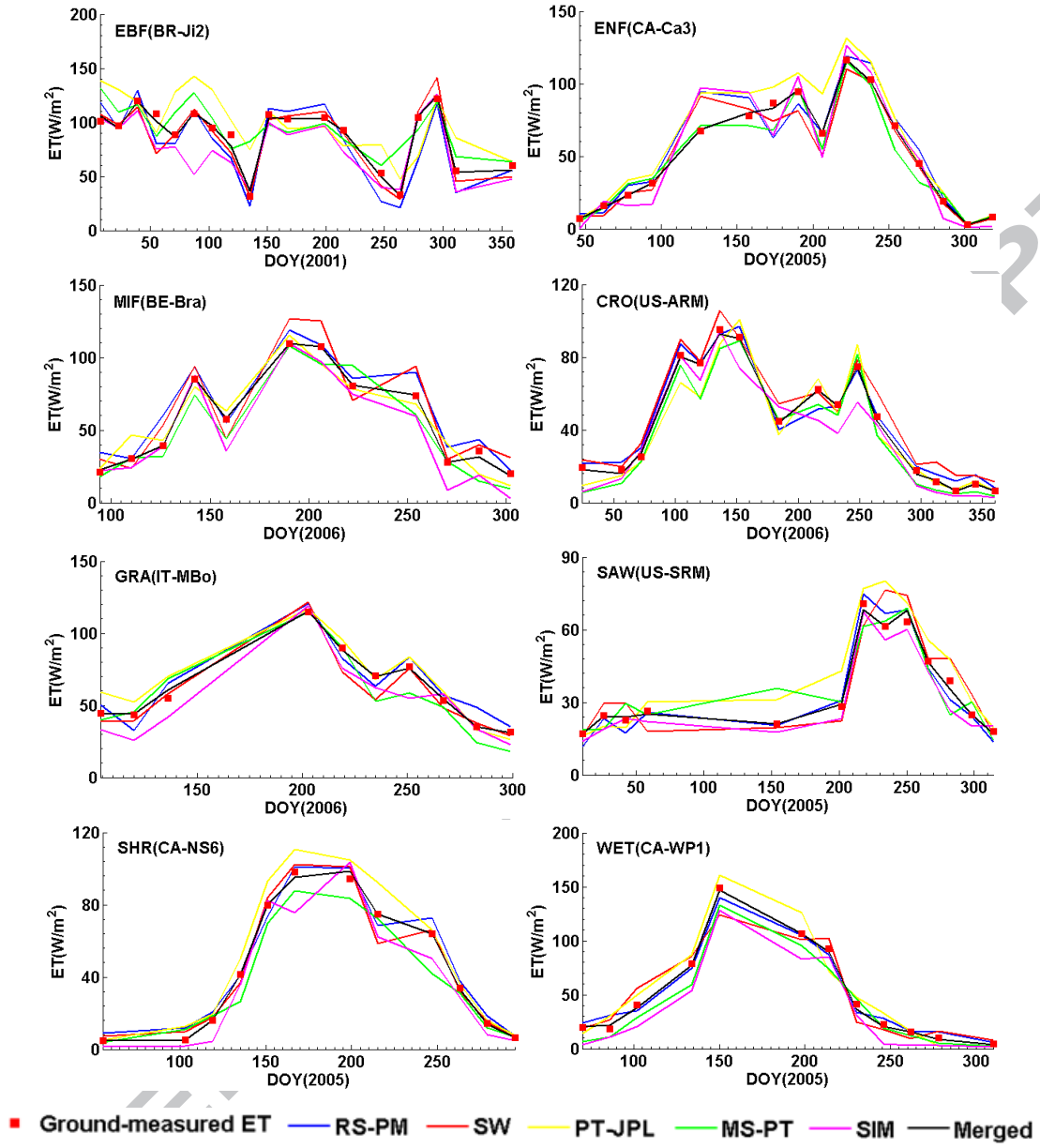


Figure 8. Example of a time series for daily *ET* as ground-measured and estimated using different Landsat-based *ET* products (including merged *ET* product) at ten validation sites. The merged *ET* was calculated using the weights for five Landsat-based *ET* products at the 103 calibration tower sites.

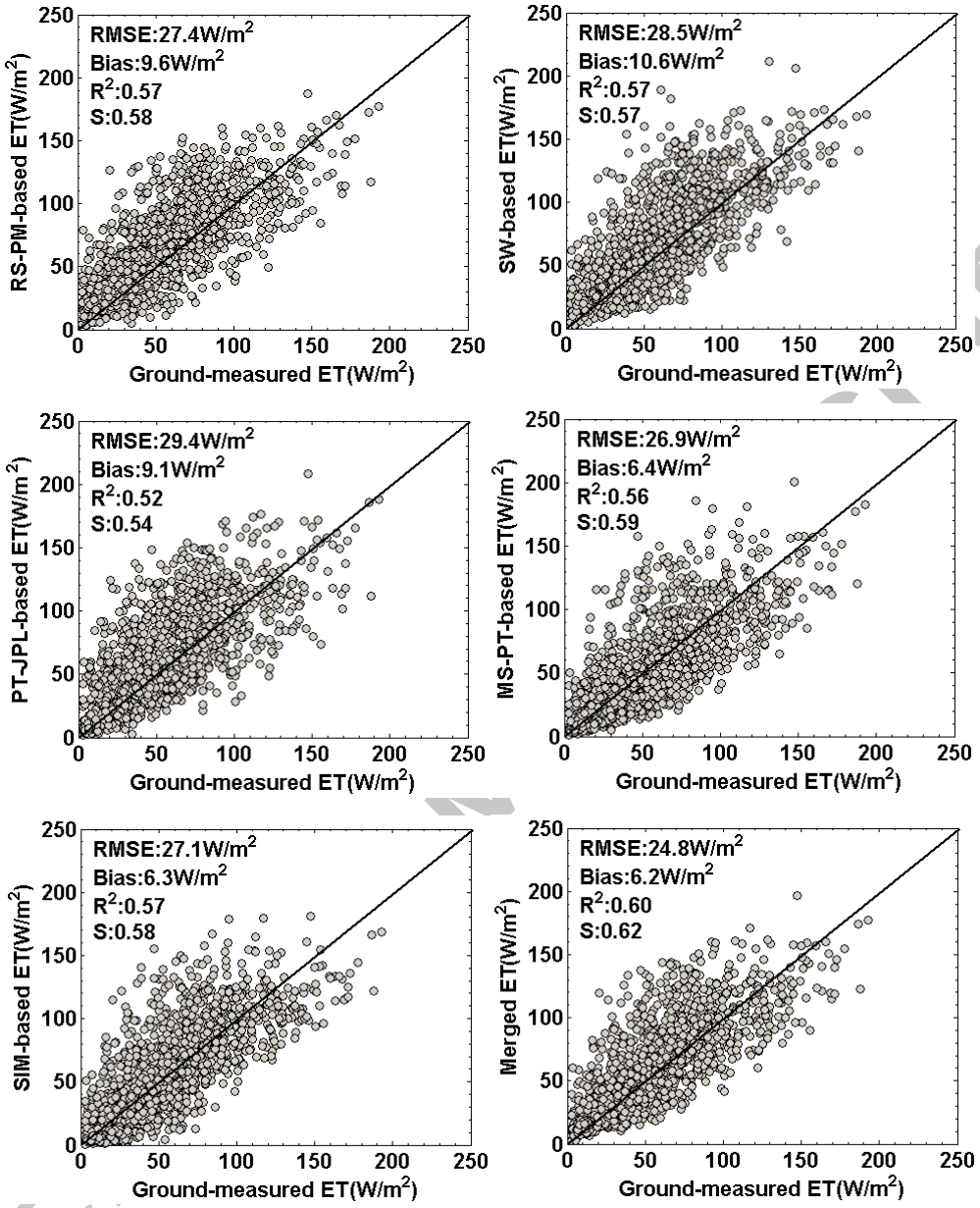


Figure 9. Same as Figure 6 but for the 103 validation tower sites. The merged *ET* was calculated using the weights for five Landsat-based *ET* products at the 103 calibration tower sites.

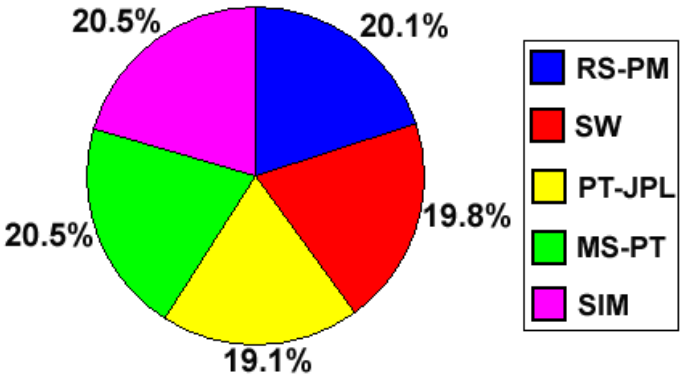


Figure 10. Weights for five Landsat-based *ET* products at all 206 flux tower sites.

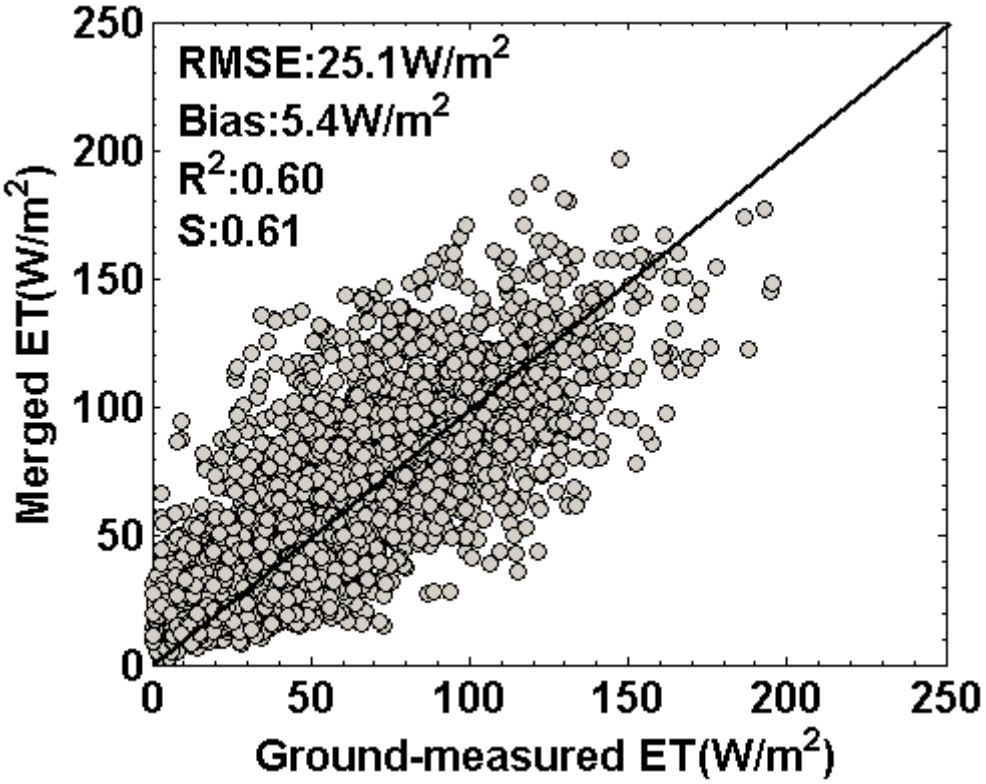


Figure 11. Scatterplots of the daily *ET* from the merged *ET* products and ground-measurements at all 206 flux tower sites. The merged *ET* was calculated using the weights for five Landsat-based *ET* products at all 206 flux tower sites.

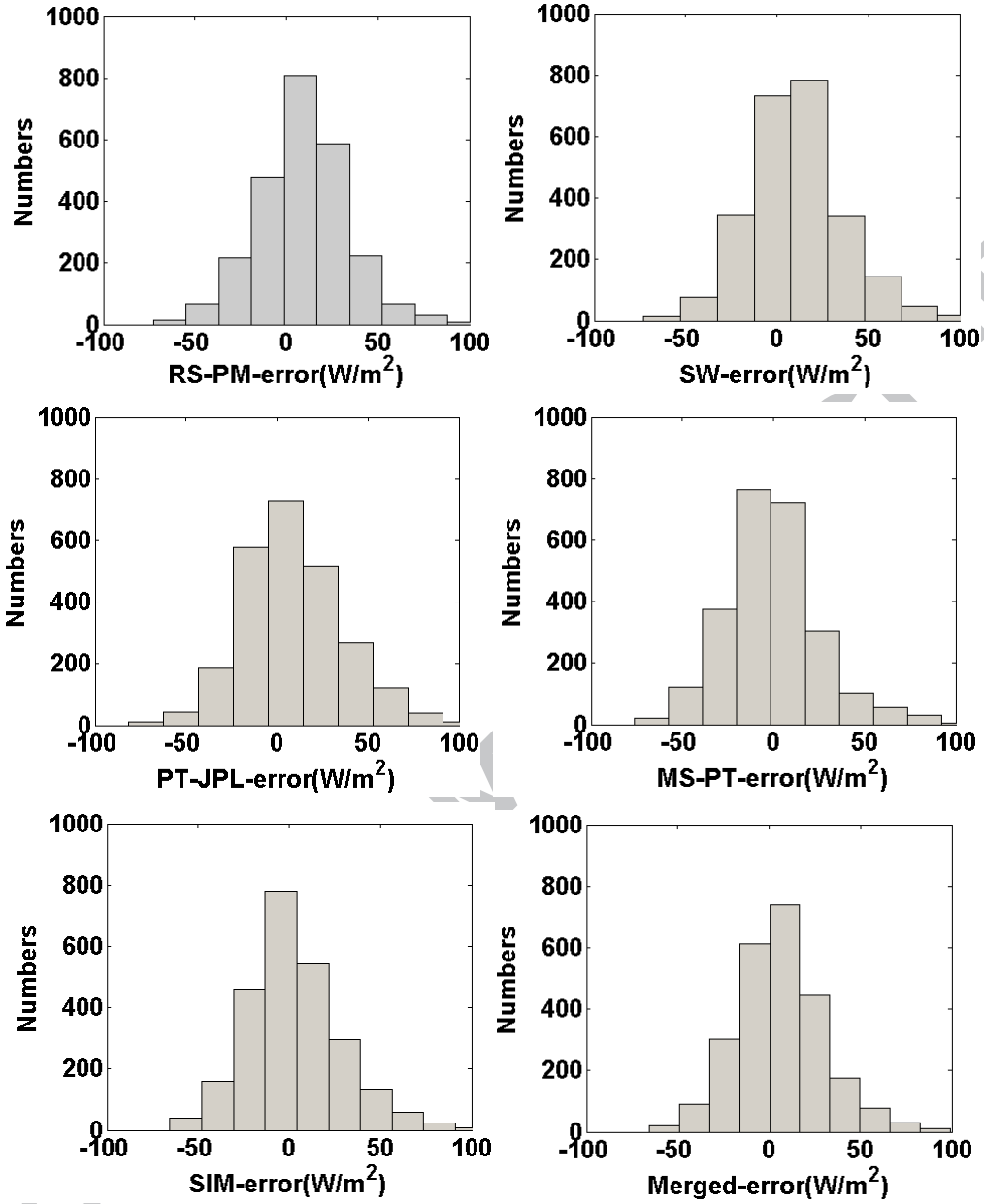


Figure 12. Error histograms for daily *ET* derived from five Landsat-based *ET* products, and the merged *ET* product for all 206 flux tower sites.

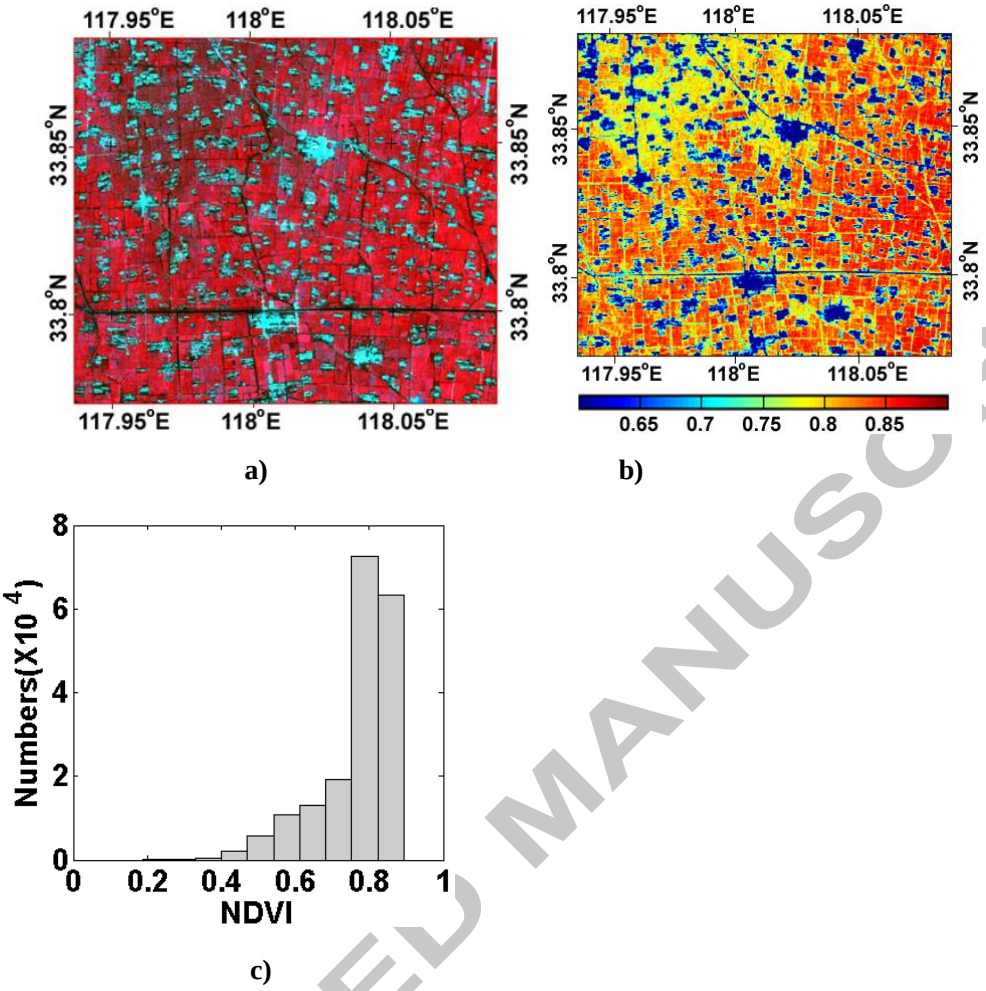


Figure 13. a) An example of a partial region of Landsat imagery with a false-color composite on August 12,2005; b) *NDVI* maps for August 12,2005, and c) frequency histograms for *NDVI* on August 12,2005.

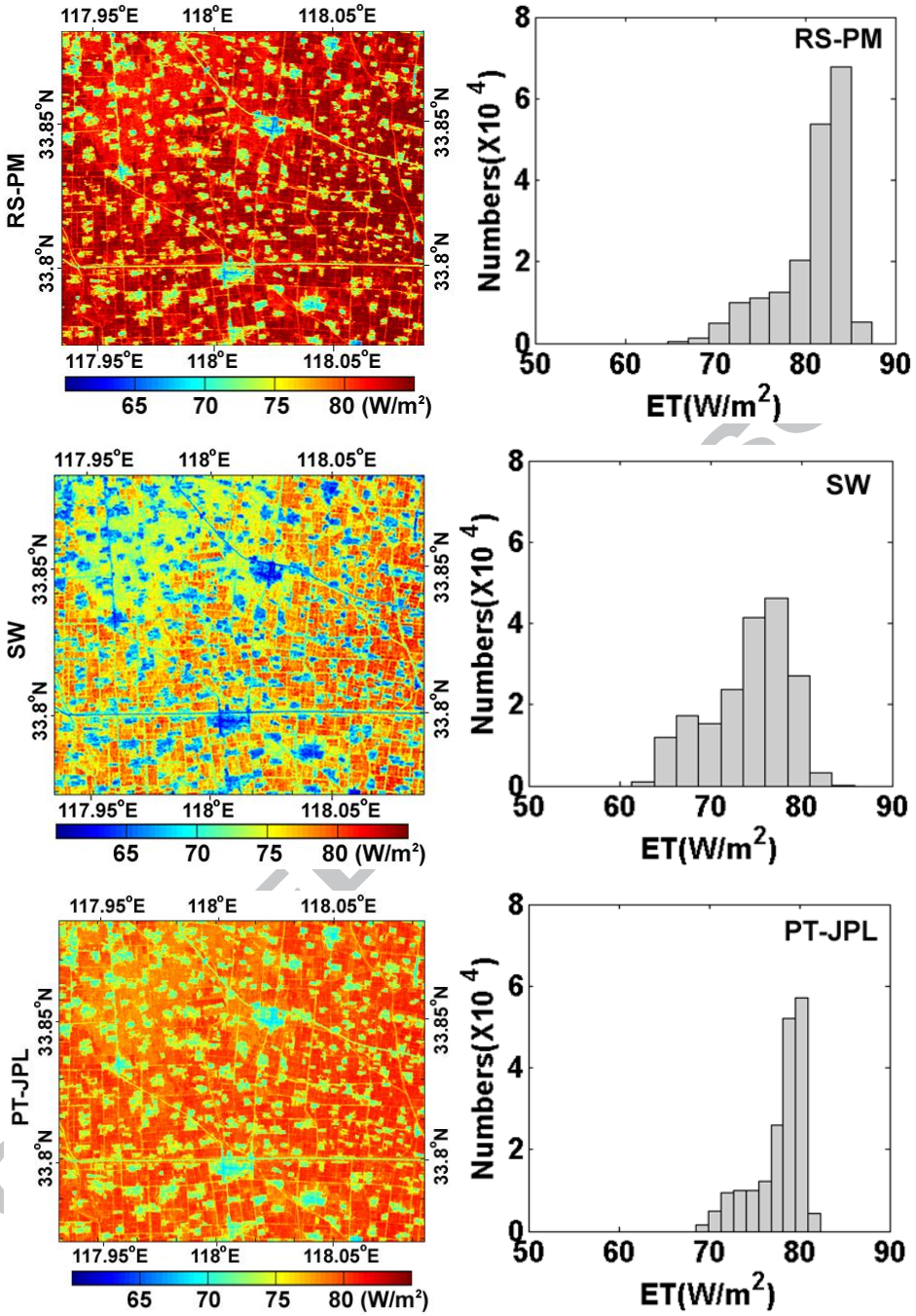


Figure 14. Daily *ET* maps of a partial region shown in Figure 13 with frequency histograms from five Landsat-based *ET* products, and the merged *ET* product for August 12, 2005.

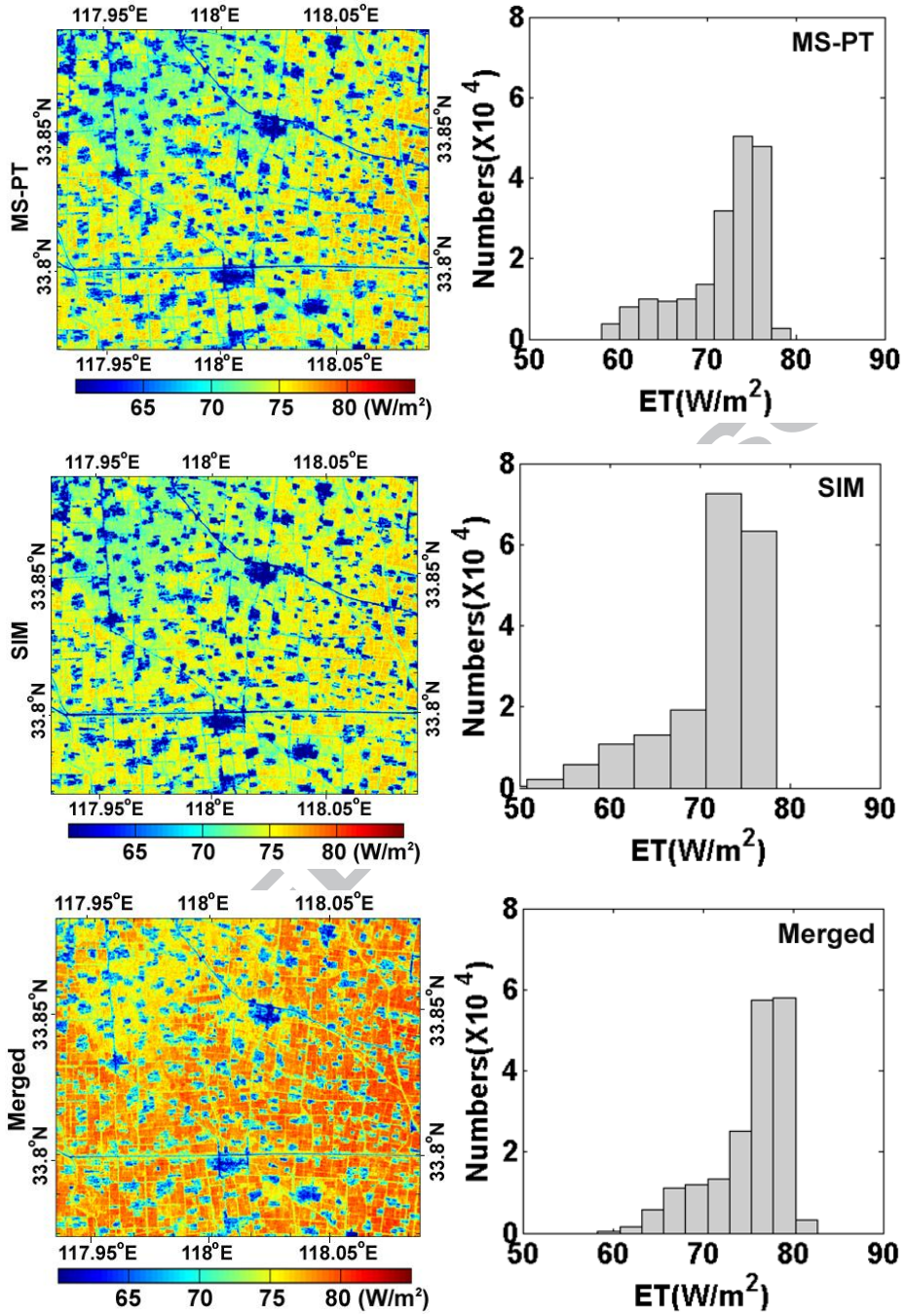


Figure 14. Continued.

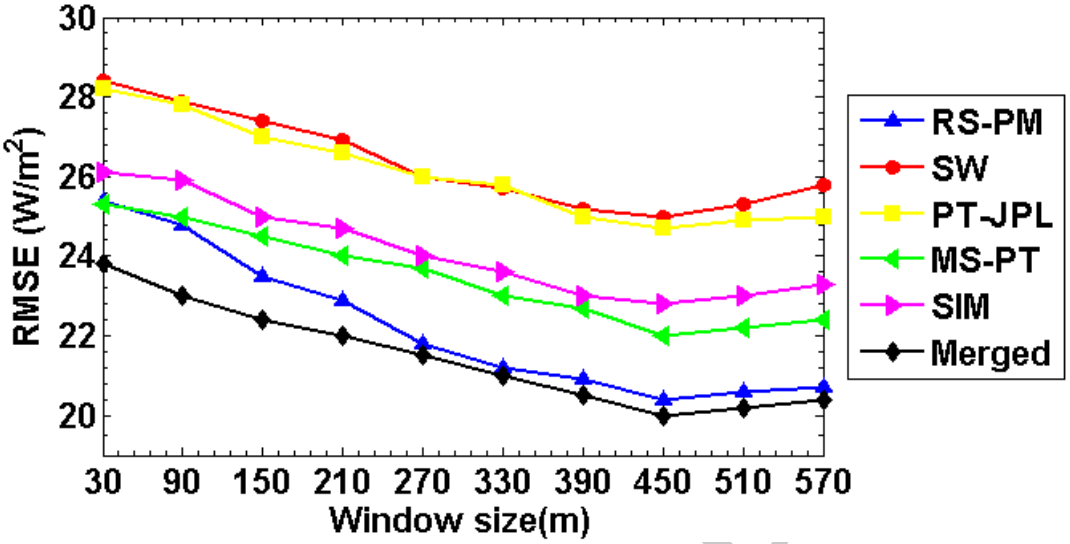


Figure 15. Change of *RMSE* of estimating daily *ET* from five Landsat-based *ET* products, and the merged *ET* product with spatial resolutions at the 103 validation tower sites. The merged *ET* was calculated using the weights for five Landsat-based *ET* products at the 103 calibration tower sites.

Highlights

1. None of the individual Landsat-based ET dataset provides the most accurate ET.

2. A simple Taylor skill fusion method was developed to merge different ET products.

3. This model produced higher accuracy when compared with the individual products.