Research Paper

Improving global terrestrial evapotranspiration estimation using support vector machine by integrating three process-based algorithms

Yunjun Yaoa,⁎, Shunlin Lianga, Xianglan Lib, Jiquan Chenc, Shaomin Liud, Kun Jiaa, Xiaotong Zhanga, Zhiqiang Xiaoa, Thomas Grünwaldj, Christian Bernhoferj, Olivier Roupsardk,l

a State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing Science and Engineering, 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 CGCEO/Geography, Michigan State University, East Lansing, MI 48823, USA
d State Key Laboratory of Earth Surface Processes and Resource Ecology, School of Natural Resources, Faculty of Geographical Science, Beijing Normal University, Beijing, 100875, China
e Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109, USA
f Numerical Terradynamic Simulation Group, Department of Ecosystem and Conservation Sciences, University of Montana, Missoula, MT 59812, USA
g Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey, NJ 08544, USA
h Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China
i University of Chinese Academy of Sciences, Beijing 100049, China
j Technische Universität Dresden, Institute of Hydrology and Meteorology, Chair of Meteorology, 01062 Dresden, Germany
k CIRAD, UMR Eco & Sol (Ecologie Fonctionnelle & Biogéochimie des Solis et des Agro-écosystèmes), 34060 Montpellier, France
l CATIE (Tropical Agricultural Centre for Research and Higher Education), 7170 Turrialba, Costa Rica

A R T I C L E  I N F O

Keywords:
Terrestrial evapotranspiration
Machine learning methods
Bayesian model averaging method
Plant functional type

A B S T R A C T

Terrestrial evapotranspiration (ET) for each plant functional type (PFT) is a key variable for linking the energy, water and carbon cycles of the atmosphere, hydrosphere and biosphere. Process-based algorithms have been widely used to estimate global terrestrial ET, yet each ET individual algorithm has exhibited large uncertainties. In this study, the support vector machine (SVM) method was introduced to improve global terrestrial ET estimation by integrating three process-based ET algorithms: MOD16, PT-JPL and SEMI-PM. At 200 FLUXNET flux tower sites, we evaluated the performance of the SVM method and others, including the Bayesian model averaging (BMA) method and the general regression neural networks (GRNNs) method together with three process-based ET algorithms. We found that the SVM method was superior to all other methods we evaluated. The validation results showed that compared with the individual algorithms, the SVM method driven by tower-specific (Modern Era Retrospective Analysis for Research and Applications, MERRA) meteorological data reduced the root mean square error (RMSE) by approximately 0.20 (0.15) mm/day for most forest sites and 0.30 (0.20) mm/day for most crop and grass sites and improved the squared correlation coefficient (R²) by approximately 0.10 (0.08) (95% confidence) for most flux tower sites. The water balance of basins and the global terrestrial ET calculation analysis also demonstrated that the regional and global estimates of the SVM-merged ET were reliable. The SVM method provides a powerful tool for improving global ET estimation to characterize the long-term spatiotemporal variations of the global terrestrial water budget.

1. Introduction

Evapotranspiration (ET), the sum of evaporation from the Earth’s surface and transpiration from plants into the atmosphere, is an important variable linking the global terrestrial water, carbon and energy exchanges (Allen et al., 1998; Liang et al., 2010; Wang and Dickinson, 2012). In general, ET returns approximately 60% of precipitation onto the Earth’s surface back to the atmosphere (Korzoun et al., 1978) and thereby conveys terrestrial water availability at the global scale (Mu et al., 2011; Yao et al., 2015). An accurate
estimation of terrestrial ET is crucial to understand the linkages between the terrestrial water budget and climate change. However, regional ET is inherently difficult to measure because of the heterogeneity in the landscape and the large number of complex controlling biophysical processes, such as available energy, plant biophysics and soil moisture (Friedl, 1996; Mu et al., 2007; National Research Council, 2007; Jiménez et al., 2011).

Remote sensing provides us broad spatial coverage and regular temporal sampling of biophysical parameters (e.g. vegetation indices, VI, albedo, leaf area index, LAI, fraction of absorbed photosynthetically active radiation, FPAR, land surface temperature, LST, and plant functional types, PFTs) (Liang et al., 2013; Los et al., 2000; Yao et al., 2013) for estimating regional ET. Over the past several years, many satellite-based methods were designed and developed to estimate regional ET, including (1) physically-based algorithms (Allen et al., 2007; Bastiaanssen et al., 1998; Fisher et al., 2008; Kustas and Daughtry, 1990; Mu et al., 2007; Norman et al., 1995; Priestley and Taylor, 1972); (2) data assimilation (DA) methods (Pipunic et al., 2008; Xu et al., 2011a,b) and (3) empirical/semi-empirical algorithms (Jackson et al., 1977; Wang et al., 2007; Wang and Liang, 2008; Wang et al., 2010a,b; Yao et al., 2015). Traditional physically-based algorithms, such as Surface Energy Balance System (SEBS) (Su, 2002), the Surface Energy Balance Algorithm for Land (SEBAL) algorithm (Bastiaanssen et al., 1998), the Two-Source ET model coupled with Atmosphere-Land Exchange Inverse (ALEXI) model (Anderson et al., 1997), the Moderate Resolution Imaging Spectroradiometer (MODIS) LAI-based Penman-Monteith (PM) equation (Mu et al., 2007; Mu et al., 2011) and Priestley-Taylor (PT) algorithm (Priestley and Taylor, 1972; Fisher et al., 2008), model the dynamics of ET process based on surface energy balance (SEB) equation and the Monin-Obukhov Similarity Theory (MOST) driven by satellite and meteorological observations (Wang and Dickinson, 2012). However, their simulation results may differ substantially due to the large errors from too many input variables and uncertainty that exists in the structures of the models. Although DA methods assimilate satellite-based parameters (e.g., LAI, LST) into biophysical or land surface models (LSMs) to improve ET estimation (Pipunic et al., 2008; Xu et al., 2011a,b), a longstanding limitation associated with DA methods is that the ET simulation accuracy has been mainly affected by the accuracy of satellite-based input variables.

Empirical/semi-empirical algorithms have been developed by relating ground-measured ET to satellite-based vegetation parameters and other key meteorological variables (Wang et al., 2007). As specific empirical algorithms, data-driven methods, including artificial neural network (ANN) (Lu and Zhuang, 2010), support vector machine (SVM) (Shrestha and Shukla, 2015; Yang, 2006) and model tree ensembles (MTE) (Jung et al., 2010) estimate ET by building relationships between input variables and outputs (ET) using training datasets. These methods are sound in theory and provide accurate estimates of ET as long as enough training datasets are representative of all the behaviors found in the systems. However, they still show substantial differences in partitioning ET for different regions and biomes due to the limited training data at certain sites. Moreover, large data requirements for data-driven methods can reduce their computational efficiency for generating satellite-based ET products.

Multi-model ensemble approaches have been successfully used to improve global terrestrial ET estimation. Former studies have indicated that a simple model averaging method (SMA) or Bayesian model averaging (BMA) method is superior to single model for predicting terrestrial latent heat flux (LE) and surface longwave radiation (Chen et al., 2015; Wu et al., 2012; Yao et al., 2014). For example, Yao et al. (2014) used the BMA method to merge five process-based LE algorithms and effectively improved the skills of the algorithms. Wu et al. (2012) also found that the BMA method has the highest accuracy than individual algorithms to combine eight land surface long-wave radiation algorithms. These multi-model ensemble approaches obtain more accurate estimates of the surface energy budget based on the linear combination of each single model by gathering useful information from multiple models to produce ensemble predictions. In theory, multi-model ensemble approaches based on a nonlinear combination of each single model, such as machine learning techniques, performs better than those based on a linear combination of each single model (e.g. BMA method) for predicting hydrologic and biophysical variables (Duan and Phillips, 2010; Sheffield and Wood, 2008). However, there is a lack of similar studies on predicting global terrestrial ET using machine learning methods for merging multi-models.

In this paper, to reduce uncertainties in global ET estimation using the individual process-based ET algorithms, we used the classical machine learning method, the SVM method, to improve global terrestrial ET estimation by merging three process-based algorithms. In Yao et al., 2014 paper, five ET algorithms including two PM algorithms, two PT algorithms and one semi-empirical Penman algorithm were merged for ET estimation. However, numerous studies found the similar performance of above two PM or PT algorithms for most land cover types (Yao et al., 2014; Yuan et al., 2010). Therefore, in this study, we only selected one PM algorithm, one PT algorithm and one semi-empirical Penman algorithm for ET estimation. Our specific objectives are to: 1) assess the performance of the SVM method for merging three process-based ET algorithms based on a series of cross-validations using long-term FLUXNET eddy covariance (EC) observations from 2000 through 2009; 2) compare the SVM method with the BMA method, the general regression neural networks (GRNNs) method and the water balance (WB) equation at the site and basin scales; and 3) generate a global daily ET product during 2003–2005 with well-quantified accuracy based on MODIS data and Modern Era Retrospective Analysis for Research and Applications (MERRA) meteorological data.

2. Data and methods

2.1. Data source

2.1.1. Data at eddy covariance flux tower sites

The performances of the SVM method, the GRNNs method, the BMA method and three process-based ET algorithms were examined using ground-measured EC data. The data were collected at 200 EC flux tower sites located in Asia, Europe, Africa, Australia, South America and North America (Fig. 1). The data were collected from AsiaFlux, AmeriFlux, LathuileFlux, Arid/Semi-arid experimental observation synergy and integration, the Chinese Ecosystem Research Network (CERN) and some individual principal investigators (PIs) of the FLUXNET project. The EC flux tower sites included nine major biomes: evergreen broadleaf forests (EBF, 14 sites), evergreen needleleaf forests (ENF, 50 sites), deciduous broadleaf forests (DBF, 24 sites), deciduous needleleaf forests (DNF, 4 sites), mixed forests (MF, 10 sites), shrubland (SHR, 12 sites), savanna (SAW, 8 sites), croplands (CRO, 30 sites) and grasslands and other types (GRA, 48 sites). The data included half-hourly or hourly surface net radiation (Ra), solar radiation (Rs), soil heat flux (G), air temperature (Ta), vapor pressure (e), maximum temperature (Tmax), relative humidity (RH), wind speed (WS), sensible heat flux (H) and ET. Half-hour EC measurements were obtained from the raw data sampled at 10 Hz with the post-processing software EdiRe (University of Edinburgh, http://www.geos.ed.ac.uk/abs/research/micromet/EdiRe). When the number of N of half-hourly measurements exceeded 40 per day, the daily average Ra, Rs, G, Ta, e, Tmax, RH, WS, H and ET were the averages of the measurements. Thus, the total daily ET can be calculated as:

$$ET = \frac{1}{N} \sum_{i=1}^{N} ET \times 48$$

Where i is the ith half-hourly observation on each day. If N was less than 40, the daily measurements were set to a fill value. Otherwise, they
were indicated as missing. Similarly, the monthly data were aggregated from the daily data (Jia et al., 2012; Liu et al., 2011; Liu et al., 2013; Xu et al., 2013). Considering that the EC method suffers an energy imbalance problem, the measured ET was corrected based on the method proposed by Twine et al. (2000).

\[ ET_{\text{corr}} = (R_n - G)/(H_{\text{ori}} + ET_{\text{ori}}) \times ET_{\text{ori}} \]

(2)

where \( ET_{\text{corr}} \) is the corrected ET, and \( H_{\text{ori}} \) and \( ET_{\text{ori}} \) are the uncorrected \( H \) and ET, respectively.

2.1.2. Satellite and reanalysis data

To examine the performances of all ET algorithms for all flux tower sites, the daily \( R_n, R_s, T_{\text{air}}, e, R_H, \) and WS products with a spatial resolution of 1/12° × 1/24° from MERRA data provided by the National Aeronautics and Space Administration (NASA) were used in this study. Details of the MERRA dataset are available from NASA website (http://gmao.gsfc.nasa.gov/research/merra). We interpolated the daily MERRA data spatially to 1 km based on the bilinear method. Accordingly, the 8-day MODIS FPAR/LAI (MOD15A2) product (Myneni et al., 2002) and the 16 day MODIS NDVI (MOD13A2) product (Huete et al., 2002) at 1-km spatial resolution were used to drive all ET algorithms. The daily FPAR/LAI (NDVI) values were temporally interpolated from the 8-day (16-day) averages using linear interpolation. When the data were missing, we temporally filled the missing FPAR, LAI and NDVI with 1-km MODIS pixel based on the method described by Zhao et al. (2005), which exploits the closest reliable 16 day (8 day) values to replace the missing data.

To generate the global terrestrial ET product at a spatial resolution of 0.05° from 2003 to 2005, we interpolated the daily MERRA data spatially to 0.05° based on the bilinear method. We also used the Collection 5 MODIS NDVI (MOD13C1: CMG, 0.05°), Collection 4 MODIS land cover (MOD12C1: CMG, 0.05°) (Friedl et al., 2002) and the Collection 5 MODIS FPAR/LAI (MOD15A2, 1-km) to drive the three satellite-based ET algorithms. The 1-km LAI/FPAR was also aggregated into 0.05° gridded data using the bilinear method.

2.1.3. Data at global large basins

A total of 32 global large basins covering areas from \( 2.3 \times 10^5 \) to \( 6.0 \times 10^6 \) km\(^2\) were collected from Pan et al. (2012) (Fig. 1). Basin averaged monthly data, including precipitation (\( P \)) and streamflow (\( Q \)), were used and aggregated into annual data (2003–2005). The \( P \) and \( Q \) gridded products at a spatial resolution of 0.5° were generated based on a constrained Kalman filter technique that merged a number of global datasets including in situ observations, remote sensing retrievals, land surface model simulations and global reanalysis (Pan et al., 2012). In addition, the Gravity Recovery and Climate Experiment (GRACE) satellites datasets (Center for Space Research Release 4: CSR RL04) from 2003 to 2005 were also interpolated into 0.5° and used to obtain the water storage changes (TWSC) (Swenson and Wahr, 2002). At the basin scale, these gridded variables (\( P \) and \( Q \) and TWSC) products were all averaged to derive ET for the global ET algorithms assessment.

2.2. Three process-based ET algorithms

Three process-based ET algorithms were used in this study, and the algorithms are illustrated using their abbreviations in the figure legends, for example, the MODIS ET product algorithm is abbreviated as MOD16. Table 1 describes the three process-based algorithms in detail.

2.2.1. MODIS ET product algorithm

The MODIS ET product algorithm (MOD16) is an improved Penman-Monteith equation (Mu et al., 2011), which is based on a beta version (Mu et al., 2007) after being adapted by Cleugh et al. (2007):

\[ ET = \frac{\Delta R_n + \rho C_v (e_s - e) r_s}{\Delta + \gamma (1 + e_s / r_s)} \]

(3)

where \( e_s \) is saturated water vapor pressure, \( \Delta \) is the slope of the curve relating saturated water vapor pressure to temperature, \( \rho \) is the air density, \( C_v \) is the specific heat capacity of air, \( \gamma \) is the psychrometric constant, \( r_s \) is the aerodynamic resistance, and \( r_s \) is the surface resistance. The MOD16 ET algorithm is the modified beta version (Mu...
et al., 2007) by calculating ET as the sum of daytime and nighttime components; modifying vegetation cover fraction with FPAR derived from MOD15A2 product; modifying calculations of aerodynamic, boundary-layer, and canopy resistance and dividing the canopy and soil into wet and dry components, respectively (Mu et al., 2011). The total ET is the sum of interception evaporation (ETᵢ), saturated soil wet soil evaporation (ETₛᵢ) and unsaturated soil evaporation (ETᵤᵢ).

\[
ET = ETᵢ + ETₛᵢ + ETᵤᵢ
\]

(4)

\[
ETᵢ = \frac{\Delta Rₑ + \rho C_p(e_ᵢ - e_f) [1 - fᵦ]}{\Delta + \gamma (1 + fᵦ)}
\]

(5)

\[
ETₛᵢ = \frac{\Delta Rₑ + \rho C_p(e_ᵢ - e_f) [1 - fᵦ]}{\Delta + \gamma} \times \frac{Rₑ}{Rₑ + \rho C_p(e_ᵢ - e_f) [1 - fᵦ]}
\]

(6)

\[
ETᵤᵢ = \frac{\Delta Rₑ + \rho C_p(e_ᵢ - e_f) [1 - fᵦ]}{\Delta + \gamma} \times \frac{Rₑ}{Rₑ + \rho C_p(e_ᵢ - e_f) [1 - fᵦ]}
\]

(7)

where \( Rₑ \) is the net radiation to the canopy, \( Rₑ \) is the net radiation to the soil, \( fᵦ \) is the vegetation cover fraction, \( fᵦ \) is the relative surface wetness cover from the PT-JPL model (Fisher et al., 2008), \( VPD \) is the vapor pressure deficit, \( \beta \) is a constant (2000), \( rhrc \) is the aerodynamic resistance on the wet canopy surface, \( rcv \) is the wet canopy resistance, \( ra₀ \) is the total aerodynamic resistance to vapor transport, and \( ra₀ \) is the aerodynamic resistance at the soil surface. Further details of the MOD16 algorithm can be found in Mu et al. (2011).

### 2.2.2. Priestley-Taylor-Based ET algorithm

Starting with the Priestley and Taylor (1972) equation for potential ET, Fisher et al. (2008) developed the PT-JPL model by introducing both ecophysiological (FPAR and LAI) and atmospheric (RH and VPD) constraints without using any ground-based observed data to reduce potential ET to actual ET. The total ET is partitioned into three components, the soil evaporation (ETₛ), the canopy transpiration (ETᶜ) and the interception evaporation (ETᵢ).

\[
ET = ETᵢ + ETₛ + ETᶜ
\]

(9)

\[
ETᵢ = \frac{\Delta \rho C_p(eᵢ - e_f) [1 - fᵦ]}{\Delta + \gamma (1 + fᵦ)} \times \frac{Rₑ}{Rₑ + \rho C_p(eᵢ - e_f) [1 - fᵦ]}
\]

(10)

\[
ETₛ = \frac{\Delta \rho C_p(eᵢ - e_f) [1 - fᵦ]}{\Delta + \gamma} \times \frac{Rₑ}{Rₑ + \rho C_p(eᵢ - e_f) [1 - fᵦ]}
\]

(11)

\[
ETᶜ = \frac{\Delta \rho C_p(eᵢ - e_f) [1 - fᵦ]}{\Delta + \gamma} \times \frac{Rₑ}{Rₑ + \rho C_p(eᵢ - e_f) [1 - fᵦ]}
\]

(12)

where \( \alpha \) is the Priestley-Taylor (PT) coefficient for a wet surface condition (1.26), \( fᵦ \) is the soil moisture constraint, \( fᵦ \) is the plant temperature constraint, \( fᵦ \) is the green canopy fraction, \( fᵦ \) is the plant moisture constraint, \( F_{\text{FPAR}} \) is the fraction of PAR absorbed by green vegetation cover and \( F_{\text{PAR}} \) is the fraction of PAR intercepted by total vegetation cover, which is estimated with NDVI (Fisher et al., 2008). Details of the PT-JPL algorithm were fully described by Fisher et al. (2008).

### 2.2.3. Semi-empirical Penman algorithm

Based on the Penman (1948) equation, the Semi-empirical Penman ET algorithm (SEMI-PM) was developed by Wang et al. (2010a). This algorithm considers that the total ET is composed of two components, the energy control component (ETₑ) and the aerodynamic control component (ETₐ).

\[
ETₑ = \frac{\Delta \rho C_p(eᵢ - e_f) [1 - fᵦ]}{\Delta + \gamma (1 + fᵦ)} \times \frac{Rₑ}{Rₑ + \rho C_p(eᵢ - e_f) [1 - fᵦ]}
\]

(14)

\[
ETₐ = \frac{\gamma}{\Delta + \gamma} \times \frac{WS[\alphaᵢ + \alphaᵦNDVI + (1 - \frac{RH}{100})] \alphaᵦNDVI}{VPD}
\]

(15)

\[
ETₐ = \frac{\gamma}{\Delta + \gamma} \times \frac{WS[\alphaᵢ + \alphaᵦNDVI + (1 - \frac{RH}{100})] \alphaᵦNDVI}{VPD}
\]

(16)

The empirical coefficients were derived from observed data collected at 64 globally distributed flux tower sites. The algorithm considers different climate conditions and is simple to operate. The algorithm includes WS, which may play an important role in annual or decadal ET variability (McVicar et al., 2012; Wang et al., 2010a,b).

### 2.3. Support vector machine

The support vector machine (SVM) method was used in this study to merge the three satellite-based ET algorithms to estimate the global terrestrial ET. For SVM, linear models in the new feature space can be used to resolve the original nonlinear problem because a multi-dimensional input space is more likely to be linearly separable in a new feature space (Vapnik, 1995; Yang, 2006; Nurmemet et al., 2015). For a given training dataset \((xᵢ, yᵢ), 1 \leq i \leq n\), \(yᵢ\) is the input of the ET derived from each single ET algorithm, \(yᵢ\) is the target concept of the ground-measured ET, and \(n\) is the number of training examples. To obtain a functional dependency \(f(x)\) between the inputs \(x\) and the target \(y\) derived from the set of independent and identically distributed observations, the objective function for the SVM method (Vapnik, 1995) can be formulated as follows:

\[
f(x) = \langle w, x \rangle + b
\]

(17)

Minimize \( \frac{1}{2} \|w\|² + K \sum_{i=1}^{n} (\etaᵢ + \etaᵢ²) \)

(18)

Subject to \( \langle w, xᵢ \rangle + b - yᵢ \leq \epsilon + \etaᵢ \)

(19)

\( \etaᵢ, \etaᵢ² \geq 0, i = 1, ..., n \)

(21)

where \( x \) is the input vector, \( w \) is the weights vector norm, \( \langle w, x \rangle \) is the dot product of \( x \) and \( w \), \( \epsilon \) is a cost of errors, \( \epsilon \) is Vapnik’s insensitive loss function, and \( \etaᵢ \) denotes the predicted value to

---

**Table 1**

Summary of the six ET models and forcing variables. ET is the total evapotranspiration; ETᵢ is the canopy transpiration; ETₛᵢ is the soil evaporation; ETᵤᵢ is the interception evaporation; ETₑᵢ is the total evapotranspiration derived from the MOD16 algorithm; ETₑᵢ is the total evapotranspiration derived from the PT-JPL algorithm; and ETₑᵢ is the total evapotranspiration derived from the SEMI-PM algorithm.

<table>
<thead>
<tr>
<th>ID</th>
<th>ET algorithm</th>
<th>Forcing Inputs</th>
<th>Outs</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MODIS ET products algorithm (MOD16)</td>
<td>Rs, Ta, Tain, RH, FPAR, LAI, PFTs</td>
<td>ET₁, ET₂, ETᵢ, ETₑᵢ</td>
<td>Mu et al. (2011)</td>
</tr>
<tr>
<td>2</td>
<td>Priestley-Taylor ET algorithm of Jet Propulsion Laboratory (PT-JPL)</td>
<td>Rs, Ta, Tain, RH, FPAR, LAI, NDVI</td>
<td>ET₁, ET₂, ETᵢ, ETₑᵢ</td>
<td>Fisher et al. (2008)</td>
</tr>
<tr>
<td>3</td>
<td>Semi-empirical Penman ET algorithm (SEMI-PM)</td>
<td>Rs, Ta, RH, WS, NDVI</td>
<td>ETᵢ</td>
<td>Wang et al. (2010a)</td>
</tr>
<tr>
<td>4</td>
<td>Bayesian model averaging method (BMA)</td>
<td>ET₁, ET₂, ETᵢ</td>
<td>ET</td>
<td>Raftery et al. (2005)</td>
</tr>
<tr>
<td>5</td>
<td>General regression neural networks (GRNNs)</td>
<td>ET₁, ET₂, ETᵢ</td>
<td>ET</td>
<td>Specht (1991)</td>
</tr>
</tbody>
</table>
be above the true value by more than $\epsilon$, and $\eta^*$ to be below the true value by more than $\epsilon$. Fig. 2 illustrates the one-dimensional linear regression function with an $\epsilon$-insensitive band. Data points out of the $\epsilon$-insensitive band are called support vectors, and only support vectors contribute to the optimization solution (Yang, 2006; Shrestha and Shukla, 2015).

The optimization problem presented in Eqs. (18)–(21) can be solved based on the technique of Lagrange multipliers ($a$ and $a^*$) by the following equation:

\[
\text{Maximize } <a, a^*> = \sum_{i=1}^{n} \eta_i (a_i - a_i^*) - \epsilon \sum_{i=1}^{n} (a_i + a_i^*) - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (a_i - a_i^*) (a_j - a_j^*) <x_i, x_j>
\]

Subject to \( \sum_{i=1}^{n} (a_i^* - a_i) = 0, a_i, a_i^* \in [0, K] \)

(22)

Then, the approximating $f(x)$ function can be written as:

\[
f(x) = \sum_{i=1}^{n} (a_i^* - a_i) <x, x_i> + b
\]

(24)

The kernel function $u(x, x_i)$ is introduced to bring the training data into a high dimension feature space and Eq. (24) can be updated as:

\[
f(x) = \sum_{i=1}^{n} (a_i^* - a_i) u(x, x_i) + b
\]

(25)

We used the Radial basis function (RBF) kernel in this study because previous studies have shown that the RBF kernel performs better than other kernels (Dibike et al., 2001; Khalil et al., 2006). The RBF kernel function can be expressed as:

\[
u(x, x_i) = \exp(-\frac{1}{2\sigma^2} \|x - x_i\|^2)
\]

(26)

where $\sigma$ is a variance. Further details of the SVM method can be found in Vapnik (1995).

2.4. Other multi-model ensemble methods

2.4.1. Bayesian model averaging method

The Bayesian model averaging (BMA) method is an approach to combine the forecast densities predicted by different models, producing a new forecast probability density function (PDF) (Duan and Phillips, 2010; Raftery et al., 1995; Yao et al., 2016). According to the BMA method, the combined forecast PDF of a variable $y$ (ET in this study), given the independent predictions of $k$ models, \([A_1, A_2, \ldots, A_k]\), and the corresponding EC ET observation, $O$, can be expressed as:

\[
p(y | A_1, A_2, \ldots, A_k, O) = \sum_{i=1}^{k} p(A_i | O) p(y | A_i, O)
\]

(27)

Where $p(A_i | O)$ is the posterior distribution of $y$ for $A_i$, $p(y | A_i, O)$ is the predictive model likelihood being correct using the observations, $O$, and it can be considered as the weight ($C_i$) of model $A_i$. Thus, Eq. (25) can be written as:

\[
p(y | A_1, A_2, \ldots, A_k, O) = \sum_{i=1}^{k} C_i p(y | A_i, O)
\]

(28)

$C_i$ can be calculated using the maximum likelihood function, which has been acquired from the expectation maximization (EM) algorithm (Raftery et al., 2005). Further details of the EM algorithm and the BMA method can be found in Duan and Phillips (2010).

2.4.2. General regression neural networks

General regression neural networks (GRNNs) are the generalizations of radial basis function networks and probabilistic neural networks (Specht, 1991). The functional estimate of the GRNN method is calculated directly from the training data without iterative training. The basic structure of the GRNN method includes four layers: the input layer, the pattern layer, the summation layer and the output layer (Jia et al., 2015; Xiao et al., 2014). The input layer includes the input variables (ET estimated from each single algorithm) and the output layer provides the GRNN method estimated ET by merging the three algorithms. The kernel function of the GRNN method meets the Gaussian distribution and the fundamental formulation can be written as:

\[
Y'(X) = \frac{\sum_{i=1}^{n} Y_i \exp(-\frac{D_i^2}{2\sigma^2})}{\sum_{i=1}^{n} \exp(-\frac{D_i^2}{2\sigma^2})}
\]

(29)

\[D_i^2 = (X - X_i)^T (X - X_i)
\]

(30)

where $Y'(X)$ is the estimation corresponding to the input vectors $X, Y_i$ is the output vector corresponding to the $i$th training input vector $X_i$, $n$ is the number of samples, $D_i^2$ is the squared Euclidean distance between $X$ and $X_i$, and $\sigma$ refers to a smoothing parameter that controls the size of the receptive region. $\sigma$ affects the weights and accuracy of the GRNNs method for ET prediction. The holdout method was used to determine $\sigma$ by removing one sample from the training data and then constructing the GRNNs using all of the remaining training samples. The training processes were terminated once the minimum of the cost function of $\sigma$ was reached:

\[
f(\sigma) = \frac{1}{n} \sum_{i=1}^{n} (\bar{Y}(X_i) - Y_i)^2
\]

(31)

where $\bar{Y}(X_i)$ is the estimate corresponding to $X_i$ based on the GRNNs trained over all of the remaining training samples, except the $i$th sample. More details of the GRNNs method can be found in Specht (1991).

2.5. Evaluation methods

2.5.1. SVM experimental setup based on cross-validation

To merge three satellite-based ET algorithms, we trained the SVM method based on the ground-measured ET for period of 2000–2009 and the corresponding estimated ET using the individual algorithms. To remove the influence of the input variables with different absolute magnitudes, we scaled all of the input variables on the range of $-1$ to $1$.

We trained and tested the models as follows. Firstly, we selected the radial basis function (RBF) kernel because it determines the performance of machine learning methods and requires only one parameter ($\sigma$). Secondly, we initially set a coarse grid search for $K(2^{-1}, 2^0, \ldots, 2^4, \epsilon (2^{-5}, 2^{-4}, \ldots, 2^{-2})$ and $\sigma (2^{-3}, 2^{-2}, \ldots, 2^0)$, and further found the $K$, $\epsilon$ and $\sigma$ with the lowest mean cross-validation root mean squared error (RMSE). Based on the selected $K$, $\epsilon$ and $\sigma$, a final training of the SVM
for 2000–2009 EC data was performed. Thirdly, we trained and tested the performances of the SVM method using a fourfold cross-validation method. The training data sets were stratified into four folds, each containing ca. 25% of the data (Jung et al., 2011). Entire sites were assigned to each fold (Jung et al., 2011; Tramontana et al., 2016). SVM training is performed four times on three of the groups, with the remaining group reserved for testing and parameters with the lowest cross-validation errors are chosen. Moreover, we evaluated the performance of the SVM method by comparing the SVM results with the BMA method, the GRNNs method and the WB equation. Here, similar procedures were performed to design the GRNNs experimental setup. Finally, we trained the SVM method using all available data to merge the three satellite-based ET algorithms to generate global terrestrial ET product.

Fig. 3. a) Taaylor diagrams for the daily ET observations and ET estimates using the different algorithms driven by tower-specific meteorology at the 200 EC sites. The dotted circular lines connecting the X and Y axes represent the STD, the dotted radial lines are the correlation (R), and the green curves denote the RMSE with respect to the reference dataset. The simulated ET based on the SVM method, the GRNNs method and the BMA method and by merging three satellite-based ET algorithms for each of the four groups was independently validated using the samples of the remaining three groups (mm/d refers to mm per day). b) Same as Fig. 3a) but for the results driven by MERRA meteorology at the 200 EC sites. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
2.5.2. Taylor diagrams

Taylor diagrams were used to assess the performance of the SVM method, the GRNNs method, the BMA method and the individual ET algorithms (Taylor, 2001). A Taylor diagram is a polar-style graph that includes the standard deviation (STD) between the simulations and the observations, the correlation coefficient (R) and the centered RMSE. In a Taylor diagram, STD is the radial distance from the origin, R is characterized by the cosine of the azimuth angle, and RMSE refers to the radial distance from the observed point. In addition, the average bias and p values for the estimated ET and ground-measured ET were used to assess the simulation errors in the different ET algorithms.

2.5.3. The Akaike information criterion and the Bayesian information criterion

The Akaike information criterion (AIC) and the Bayesian information criterion (BIC) were also used to evaluate the performance of the SVM method, the GRNNs method, the BMA method and the individual ET algorithms. The AIC is a measure of the quality of each model, relative to each of the other models for a given set of data (Akaike, 1974; Loehlin, 1992) and the AIC value of the model can be expressed as:

\[ AIC = -2 \ln L + 2c \]  

(32)

Where \( L \) is the maximum value of the likelihood function for the model.
and \( c \) is the number of free parameters in the model. The model with the smallest AIC is the best performance. The BIC is also an indicator for assessing model performance, but it takes into account the number of data points, \( n \) (Schwarz, 1978). The BIC is formally defined as:

\[
BIC = -2 \ln L + c \ln(n) \tag{33}
\]

The model with lowest BIC values is preferred. Thus, the good performance of different algorithms in this study is normally based on the low AIC and BIC values.

2.5.4. Water balance equation

The SVM-merged \( ET \) estimation over the basin and regional scale was evaluated based on the water balance equation. \( ET \) can be calculated based on the precipitation (\( P \)), the streamflow (\( Q \)) and the water storage changes (TWSC) within a water-closed basin.

\[
ET = P - Q - TWSC \tag{34}
\]

Of the four water budget components, \( P \) and \( Q \) can be acquired from the multiple datasets that were produced by Pan et al. (2012), and TWSC can be acquired from the GRACE data. Thus, terrestrial ET can be inferred using Eq. (34) within the 32 global large basins.

2.5.5. Contribution of each individual algorithm on merged \( ET \)

To test the contribution of each individual algorithm on SVM-merged \( ET \), we removed one of the individual algorithms and replicated the cross-validation training process. The mean cross-validation RMSE and the squared correlation coefficient \( R^2 \) from the cross-validation training process were quantitatively used to evaluate the contribution of each individual algorithm.

3. Results

3.1. The performance of the SVM method at the site scale

Fig. 3a) and b) show the Taylor diagrams for the daily \( ET \) observations and \( ET \) estimates using the different algorithms driven by tower-specific (defined as “ground-measured”) meteorology and MERRA meteorology at the 200 EC sites, respectively. Figs. 3 and 4 showed that the six algorithms exhibited substantial differences for each \( PFT \). For the ME, DNF and DBF sites, the SVM method driven by tower-specific (MERRA) meteorology behaved better than the MOD16 algorithm, the PT-JPL algorithm, the SMEI-PM algorithm, the BMA method and the GRNNs method, with an \( R^2 \) of greater than 0.78 (0.68), \( p < 0.01 \), a low bias ranging from −0.01 to 0.01 (−0.02−0.02) mm/day and smaller RMSEs of less than 0.70 (0.80) mm/day. Similarly, for the ENF and EBF sites using the SVM method driven by tower-specific (MERRA) meteorology, the RMSE of the estimated \( ET \) versus ground observations was approximately 0.66 (0.93) mm/day and the \( R^2 \) is approximately less than 0.75 (0.61) \( p < 0.01 \), but it still presented better performance than the BMA method, the GRNNs method and the individual algorithms. For all of the crop sites, the estimated \( ET \) using the SVM method for tower-specific (MERRA) meteorology inputs still exhibited the lowest RMSE of 0.81 (1.08) mm/day, and the highest \( R^2 \) of 0.74 (0.56) at the 99% level of confidence, compared with the BMA method, the GRNNs method and the individual algorithms. Almost all three individual algorithms showed the poor performance at the crop sites and so did the three merged estimates. Therefore, a poor model performance of the SVM method was also found at these crop sites. For the other \( PFTs \) (GRA, SAW and SHR) sites, the average RMSE was much lower and the average \( R^2 \) was slightly higher for the SVM method compared with the other five algorithms. As another machine learning method, the GRNNs method was superior to the BMA method and the individual algorithms for all \( PFTs \), but it still had lower performance with lower \( R^2 \) and higher RMSE than the SVM method. For all of the \( PFTs \), the SVM method was superior to the GRNNs and the BMA methods. Overall, compared with the individual algorithms, the RMSE of the SVM method driven by tower-specific (MERRA) meteorology decreased the RMSE by approximately 0.20 (0.15) mm/day for most forest sites and approximately 0.30 (0.20) mm/day for most crop and grass sites and increased the \( R^2 \) by more than 0.10 (0.08) (95% confidence) for most flux tower sites.

Fig. 5 demonstrated the AIC and BIC values calculated from six algorithms. It is clear that the SVM method driven by tower-specific (MERRA) meteorology gave the lowest AIC and BIC values for different \( PFTs \) when compared to those obtained from other five models. However, the AIC and BIC values of the SVM method are slightly lower than those of the GRNNs method. The GRNNs method provided the second best accuracies. Therefore, the SVM method provides a better representation of the \( ET \) data of the globally distributed eddy covariance tower sites used in this study than other five models.

Fig. 6 shows the SVM exhibited most features of measured \( ET \) seasonality in the ground-measured test data for different \( PFTs \). In comparison to the BMA method, the GRNNs method and the individual algorithms, the SVM method produced seasonal \( ET \) variations that were closest to the ground-measured \( ET \). The bias of the estimated \( ET \) based on the SVM method varied from −0.04 to 0.03 mm/day, the \( R^2 \) varied from 0.73 to 0.83, and the RMSE varied from 0.41 to 0.80 mm/day.

Fig. 7 shows the frequency distributions of the predictive errors in all six algorithms driven by tower-specific and MERRA meteorology, respectively. The errors distributions of the SVM-merged \( ET \) estimates are more closely centered on zero and the SVM method decreased the substantial positive and negative biases. Therefore, the SVM strategy can capture the \( ET \) variance and has good model performance.

To improve global terrestrial \( ET \) estimation using the SVM method, all of the data collected at the 200 flux tower sites were used as training data to determine the nonlinear combinations of the three satellite-based algorithms. Figs. 8 and 9 present the scatter plots between the monthly observed \( ET \) at all of the 200 flux tower sites and the \( ET \) estimates for the six algorithms driven by tower-specific and MERRA meteorology, respectively. The results show that the SVM method has the best performance, with the highest \( R^2 \) (0.90 and 0.80) \( p < 0.01 \) and the lowest RMSE (11.15 mm/month and 14.71 mm/month) compared with the other five algorithms. Previous substantial studies also illustrated that the SVM method, trained with hydro-climatic inputs, yields better \( ET \) estimates than do neural networks and other methods in a series of cross-validation experiments (Yang, 2006; Shrestha and Shukla, 2015). Therefore, the improved accuracy of the SVM method by merging the three satellite-based algorithms makes it useful for estimating the regional terrestrial \( ET \).

3.2. Evaluation of the SVM-merged \( ET \) at the basin level

We compared the estimated global \( ET \) using six algorithms driven by MERRA meteorology with the inferred \( ET \) from basin-scale water balance calculations for 32 major global basins (Fig. 10). In comparison to the BMA method, the GRNNs method and the individual algorithms, the SVM method still had the best performance with the lowest RMSE of 90.38 mm and the highest \( R^2 \) of 0.89 \( p < 0.01 \) over the 32 watersheds. Large differences between the SVM-merged \( ET \) and the inferred \( ET \) occurred in some of the high latitude basins, such as the Pechora, Yukon and Ural basins. The mean difference in those basins was approximately 110 mm/year. This discrepancy may be partially attributable to the few \( ET \) observations, which reduced the accuracy of the SVM-merged \( ET \). Pan et al. (2012) showed that the global terrestrial water budget (\( P \) and \( Q \)) determined by merging a number of global datasets has a higher accuracy compared with that based on the individual datasets, but there are still small biases in some regions. Therefore, the biases of \( P \), \( Q \) and TWSC from different data sets can also result in errors in the inferred \( ET \), which will contribute to SVM-merged \( ET \) and inferred \( ET \) differences in those regions. Although there were large differences between the SVM-merged and inferred \( ET \) in some of the basins, the good agreement based on a verification of the water
3.3. Contribution of each individual algorithm on SVM-based ET variations

Removing the SEMI-PM algorithm driven by tower-specific meteorology reduced the largest performance of SVM in cross-validation error analysis for DNF, ENF and MF PFTs (Fig. 11). $R^2$ decreased by approximately 0.10 and RMSEs increased by approximately 0.08 mm/day. Removal of the MOD16 algorithm caused the secondary performance reduction for all above three PFTs, leading to decreased $R^2$ of approximately 0.05 and increased RMSEs of 0.06 mm/day. Removing the PT-JPL algorithm yielded comparatively minor changes with the $R^2$ reduced by about 0.02 and RMSEs rose by 0.02 mm/day. In contrast, the largest performance reduction for other PFTs was to remove the PT-JPL algorithm: the RMSEs increased by more than 0.09 mm/day and the $R^2$ reduced by 0.12. While removal of the MOD16 algorithm resulted in small performance reduction for other PFTs. Therefore, the SEMI-PM algorithm captured most of the ET variations for DNF, ENF and MF PFTs, while the PT-JPL algorithm has the highest contribution to SVM-merged ET for other PFTs. Although our input each individual algorithm ranking was based on the tower-specific meteorology, similar conclusions can be drawn when using the MERRA meteorology as inputs.

3.4. SVM-merged global terrestrial ET patterns

We applied the SVM method, the GRNNs method, the BMA method and the individual algorithms with the MERRA meteorology and MODIS product to estimate annual ET globally at a 0.05° spatial resolution from 2003 to 2005. Over the 2003–2005 study period, average annual ET from the SVM method has the smallest values of 85 mm/yr in cold and arid regions, intermediate values of 321 mm/yr in the temperate regions, and highest values of 1279 mm/yr over the tropical and subtropical forests of the Congo basin in central Africa, the Amazon basins in South America and the Indonesia rain forests in Southeast Asia (Fig. 12). Compared with the MOD16 algorithm, the PT-JPL algorithm and the BMA method, the SVM method yields lower annual global terrestrial ET in rain forests regions (Indonesia, Amazon and Congo) and higher ET in arid and semi-arid regions (Fig. 13). However, there are opposite spatial differences between the SVM method and the other two methods (GRNNs and SEMI-PM).

The global terrestrial average annual ET based on the SVM method was 471.7 mm/yr, which was lower than the ET values that were based on PT-JPL (508.8 mm/yr), SEMI-PM (517.2 mm/yr), BMA (486.1 mm/yr) and GRNNs (475.9 mm/yr), and higher than the ET values that were based on MOD16 (433.7 mm/yr). The average annual ET for CRO, GRA, SAW, DNF, ENF, DBF and MF was 485 mm/yr, 244 mm/yr, 185 mm/yr, 244 mm/yr, 185 mm/yr, 589 mm/yr and 381 mm/yr, respectively. The seasonal patterns of ET averaged from 2003 through 2005 based on the balance approach for most of the basins demonstrates that the SVM method was reliable.
SVM method driven by the MERRA meteorology and MODIS product illustrated obviously seasonality for most PFTs (Fig. 14). However, there is no seasonality for EBF and SAW with high ET values around the whole year.

4. Discussion

4.1. The performance of the SVM method

By merging three process-based ET algorithms, the SVM method not only preserved the partial dynamic information of ET process, but yielded the global terrestrial ET with high accuracy. We found that the
SVM method successfully improved the ET estimate accuracy by 10–20% and 5–10% compared with the individual models and other ensemble methods (BMA and GRNNs), respectively. The SVM method performed well and explained more than 81% of the ET variability for the DBF, DNF and GRA flux tower sites. Previous studies have shown that the vegetation leaf, moisture and chlorophyll content of these

Fig. 6. Examples of the 8-day ET average as measured and estimated using the different tower-driven algorithms for the different PFTs.
biomes display obviously seasonal variations (Mu et al., 2007; Yao et al., 2015; Yebra et al., 2013). LAI and NDVI derived from remote sensing reflect the seasonal changes of vegetation information and based on these vegetation parameters, the individual algorithms have successfully captured the seasonal cycle of those biomes, which will improve ET estimation because the performance of the SVM method relies on the accuracy of the individual algorithms. In contrast to the deciduous forests and grassland cover types, the evergreen forests, including ENF and EBF, had less evident seasonal variations. Therefore, the weak variations in the satellite-based vegetation signals abated the ability of the individual algorithms and the SVM method to calculate ET (Eugster et al., 2000; Huete et al., 2002; Wang and Dickinson, 2012). In addition, for the irrigated CRO flux tower sites, the SVM method presented the poorer local performances for the ET ($R^2 = 0.51$, bias $= 0.91$ mm/day and $RMSE = 1.22$ mm/day) estimates with MERRA meteorology inputs. In contrast, the SVM method presented the better local performances for the ET ($R^2 = 0.60$, bias $= 0.20$ mm/day and $RMSE = 0.98$ mm/day) estimates. This may be attributable to the fact that the SVM method failed to simulate irrigation practice because the three satellite-based algorithms only use RH and VPD to infer soil moisture stress for model parameterization (Fishner et al., 2008; Mu et al., 2011; Wang et al., 2010a). Beyond these irrigated crop sites, the SVM method significantly improved the performance.

The relative contributions of each individual algorithm to SVM-merged ET vary for different PFTs. The SEMI-PM algorithm has the largest contribution for DNF, ENF and MF land cover types and the PT-JPL algorithm has largest contribution for other land cover types, which are generally consistent with the BMA-derived weights for the three process-based ET algorithms (Fig. 15). The study of Yang (2006) indicated that SVM outperformed other techniques (e.g. neural networks and multiple regressions) and the contribution of input variable may change with different PFTs and spatial resolution. Yao et al. (2014) also reported that SEMI-PM latent heat flux estimates had large contribution to BMA-merged ET for most land cover types because it closely matched the BMA latent heat flux estimate.

4.2. SVM-merged global terrestrial ET estimation

The SVM method for merging the three process-based ET demonstrated its reliability for estimating global terrestrial annual ET. Considering that we used the GRACE satellite data to compare the SVM-based ET and the GRACE data are available from March 2002, we generated SVM-based ET product during period of 2003–2005 in this study. Importantly, the SVM-merged annual global terrestrial ET (excluding Greenland and Antarctica) was 471.7 mm/yr from 2003 through 2005, which was comparable to other estimates. For instance, Wang and Dickinson (2012) reported that global average ET derived from surface water budget varied from 1.2 mm/d to 1.5 mm/d with an average of 1.3 ± 0.1 mm/d. Mueller et al. (2013) inferred that the estimates of globally averaged ET from satellite observation, reanalysis data and land surface model simulations were between 0.83 mm/d and 1.45 mm/d. The SVM ensemble results were similar to those results. However, spatial differences between the SVM-merged ET and other ET estimates are much greater than those for the global average values. This discrepancy may have been caused by the differences in the algorithm structures of the SVM and GRNNs methods.

Although the superior performance of the SVM method demonstrates that the use of the SVM method for merging different ET algorithms can effectively characterize the spatial distribution of ET, the SVM method underestimates monthly ET when the measurements exceed 120 mm per month. Similarly, SVM-merged averaged ET over the tropical and sub-tropical forests is 1279 mm/yr, which was lower than the results of other estimates. For instance, Bruinjzeel (1990) reported that annual ET ranges from 1310 to 1500 mm in humid tropical forests. Frank and Inouye (1994) used 25 year climate records to calculate annual ET at 10 sites and found annual ET of 1363 ± 77 mm/yr for wet tropical forest. Perhaps few training samples available for tropical forest attribute to the underestimate ET.

4.3. Uncertainty in SVM-merged ET estimate

Validation results indicate that uncertainty in SVM-merged daily ET estimate (with respect to FLUXNET) was found to range between 21 and 47%. We attribute the reasons for uncertainty in global terrestrial ET product to factors such as the corresponding errors in the tower EC observations, MERRA meteorology and satellite-based vegetation parameters (e.g., LAI and FPAR) and the spatial scale mismatch among the different data sources. Firstly, the energy balance closure of the EC observation was generally approximately 30% due to complexities in the wind patterns and to footprint variability (Foken, 2008; Twine et al., 2000; Wilson et al., 2002; Zhang et al., 2010). Although the EC data were corrected, they still had an error of approximately 5–20% (Foken, 2008), which would have reduced the accuracy of the algorithms used for the ET estimation. Secondly, many studies have demonstrated that there are large errors in the MERRA meteorology and MERRA data tend to underestimate $R_a$ at high values when compared with ground measurements (Rienecker et al., 2011; Zhao et al., 2006). This indicates that the biases in MERRA meteorological data can introduce substantial uncertainties into the ET estimates, and it is necessary to minimize those biases to improve the quality of the ET product. Thirdly, the accuracy of the MODIS LAI, FPAR and land cover types can also influence the accuracy of the ET estimates. Recent studies...
have revealed errors in MODIS LAI and FPAR when compared with ground measurements (Serbin et al., 2013). Similarly the accuracy of the MODIS Collection 5 Land Cover Type product is less than 75% globally (Hansen et al., 2000), which will lead to approximately 17% errors in SVM-merged ET estimate. Thus, these inaccurate MODIS products will also reduce the accuracy of ET estimates. The individual ET algorithms, such as MOD16, have large errors due to the biases of the MERRA and MODIS products (Mu et al., 2011; Velpuri et al., 2013). Mu et al. (2011) reported uncertainties in MOD16 ET product up to 20% on individual station-based FLUXNET validation. Finally, the spatial scale mismatch among the different data sources may have introduced errors in the ET estimation. The spatial resolution of the gridded data including the MERRA and MODIS products, was no less than 1-km, which was greater than the footprint for field measurements, which have spatial resolutions of several meters (Baldocchi, 2008). Such coarse MERRA and MODIS products may not adequately capture sub-grid scale meteorological and vegetation signals at these sites, especially in areas with complex land surfaces.

The performance of the SVM-merged ET estimates was not only validated at the site scale but was also evaluated at the basin scale using water balance approach. Basin scale validation results indicated uncertainties up to 21% of the annual estimates for SVM-merged ET. The accuracy of the inferred ET using water balance approach could be also affected by the sources of error in P, Q and TWSC. Pan et al. (2012) reported that about 10% relative error in both P and Q will persist at 15 per 10^6 km^2 despite increasing gauge density. The error in TWSC caused by the different methods for GRACE estimation (Swenson and Wahr, 2002) can lead to 14% error in calculated ET using water balance approach. In addition, the spatial resolution of MODIS products was approximately 5 km in size and was much finer than the resolution

---

**Fig. 8.** The scatter plots between the monthly observed ET at all 200 flux tower sites and the ET estimates for the six algorithms driven by tower-specific meteorology (mm/Mon refers to mm per month).
(more than 50 km) of other gridded products including MERRA, GRA-CE, fused P and Q datasets. Although all gridded products were interpolated into 5 km, error propagation through calculations, including threshold filtering, averaging, interpolation, and data fusion affected the uncertainty of the comparison of the SVM-merged ET and inferred ET based on water balance approach. Even if all the errors could be eliminated from a model and even if observational uncertainties could be reduced to zero, the modeled and observed estimates cannot be expected to be identical (Taylor, 2001). Therefore, the choice of a reasonable dataset should be made carefully depending on the requirements of the study.

4.4. Limitations and recommendations for future research

Although the SVM highlights global rather than local optima and leads to better performance compared with other machine learning methods, such as the GRNNs method, which ensures local optimization (Shrestha and Shukla, 2015; Specht, 1991; Vapnik, 1995; Verrelst et al., 2015; Yang, 2006), it faces three known limitations. Firstly, it requires a relatively long processing time (about 47.3 s for 1000 samples) to train a model. Secondly, it behaves relatively unpredictable when used with input ground-measured ET deviating from those presented during the training stage (Shrestha and Shukla, 2015; Verrelst et al., 2012). Finally, regardless of the performance outcome, however, we do not know that any of machine learning methods possess the useful information to directly deliver additional confidence ET maps. Confidence ET maps should be evaluated and validated using other ground-measured ET data from other PFTs EC sites.

To make the training samples more globally applicable, it is urgent to add samples from other PFTs (e.g. snow and ice). However, there are
few EC data available for these specific PFTs. During the past decades, there are many semi-empirical and physical methods for estimating the sublimation of snow and ice (Kuzmin, 1953; Hu and Jia, 2015). The advantage of these methods is that they do not require training samples to estimate the sublimation of snow and ice, though the accuracy of these methods may not be the highest. Future research will consider the development of machine learning methods when coupled with these semi-empirical and physical methods to improve the global terrestrial ET at more different PFTs.

5. Conclusions

We used the SVM method to merge three satellite-based ET algorithms (MOD16, PT-JPL and SEMI-PM) for global terrestrial ET estimation across multiple biomes. The inputs of each algorithm included tower-specific meteorology collected from 200 global flux tower sites, MERRA meteorology and MODIS products. Compared to the BMA method, the GRNNs method and the individual algorithms, the SVM methods had the best performance for each vegetation type and can be effectively applied to estimate global terrestrial ET.

The performance of the SVM method was examined at 200 FLUXNET EC flux towers based on a fourfold cross-validation method for each PFT. The SVM method enhanced ET estimates by merging the three satellite-based ET algorithms driven by tower-specific (MERRA) meteorology, decreasing the tower-specific RMSE of the daily ET by approximately 0.20 (0.15) mm/day for most of the forest sites and by
approximately 0.39 (0.20) mm/day for most of the crop and grass sites. The SVM-merged ET estimates captured the magnitudes of the ET measurements better than the BMA method, the GRNNs method and the individual algorithms. The regional water balance analysis also demonstrated that the regional estimates of the ensemble ET were reliable.

The SVM method improved annual ET estimates by merging the three satellite-based ET algorithms driven by MERRA meteorology and MODIS products. The mean annual SVM-merged ET over the global terrestrial ecosystem during 2003–2005 was 471.7 mm/yr, which was closer to the observations than that produced by the algorithms

Fig. 11. Impact of removing one of the three algorithms on the predicting performance ($R^2$ and RMSE) of SVM on ET. The results shown are the average from a fourfold cross-validation on the training data.

Fig. 12. The map of mean annual global terrestrial ET from 2003 through 2005 at a spatial resolution of 0.05° using different algorithms driven by MERRA meteorology.
individually. More importantly, the SVM-merged ET will provide critical information for the characterization of global terrestrial water and energy cycles as well as regional drought assessment.

Acknowledgements

We would like to thank Dr. Tongren Xu and Dr. Ziwei Xu from Beijing Normal University, China, and Prof. Guangsheng Zhou from the Institute of Botany, CAS, and Dr. Yan Li and Dr. Ran Liu from Xinjiang Institute of Ecology and Geography, CAS, and Prof. Guoyi Zhou and Dr. Yuelin Li from South China Botanic Garden, CAS, and Prof. Bin Zhao from Fudan University, China, for providing ground-measured data. 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, Carbitaly, CarboMont,ChinaFlux, Fluxnet-Canada (supported by CFCAS, NSERC, BIOCAP, Environment Canada, and NRCAn), GreenGrass, KoFlux, LBA, NECC, OzFlux, TCOS-Siberia, USCCC. We acknowledge the financial support to the eddy covariance data harmonization provided by CarboEuropeIP, FAO-GTOS-TCO, iLEAPS, Max Planck Institute for Biogeochemistry, National Science Foundation, University of Tuscia, Université Laval, Environment Canada and US Department of Energy and the database development and technical support from Berkeley Water Center, Lawrence Berkeley National Laboratory, Microsoft Research eScience, Oak Ridge National Laboratory, University of California-Berkeley and the University of Virginia. Other ground-measured data were obtained from the 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). GRACE product was downloaded online (http://grace.jpl.nasa.gov/). MODIS LAI/FPAR, NDVI and land cover satellite products were obtained online (http://reverb.echo.nasa.gov/reverb).

Fig. 13. Spatial differences in the average annual global terrestrial ET (2003–2005) between the SVM method and other models.
This work was partially supported by the Natural Science Fund of China (No. 41671331) and the National Key Research and Development Program of China (No.2016YFA0600102 and No. 2016YFB0501404).

References


