Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1029/2018JD028447

Special Section:
Water-soil-air-plant-human nexus: Modeling and observing complex land-surface systems at river basin scale

Evaluating Different Machine Learning Methods for Upscaling Evapotranspiration from Flux Towers to the Regional Scale

Tongren Xu[1], Zhixia Guo[1], Shaomin Liu[1], Xinlei He[1], Yangfanyu Meng[1], Ziwei Xu[1], Youlong Xia[2], Jingfeng Xiao[3], Yuan Zhang[1], Yanfei Ma[4], and Lisheng Song[5]

[1]State Key Laboratory of Earth Surface Processes and Resource Ecology, School of Natural Resources, Faculty of Geographical Science, Beijing Normal University, Beijing, China, [2]I. M. Systems Group at Environmental Modeling Center (EMC), National Centers for Environmental Prediction (NCEP), College Park, MD, USA, [3]Earth Systems Research Center, Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, Durham, NH, USA, [4]Department of Geography, Handan College, Handan, China, [5]School of Geographical Sciences, Southwest University, Chongqing, China

Abstract Evapotranspiration (ET) is a vital variable for land-atmosphere interactions that links surface energy balance, water, and carbon cycles. The in situ techniques can measure ET accurately but the observations have limited spatial and temporal coverage. Modeling approaches have been used to estimate ET at broad spatial and temporal scales, while accurately simulating ET at regional scales remains a major challenge. In this study, we upscale ET from eddy covariance flux tower sites to the regional scale with machine learning algorithms. Five machine learning algorithms are employed for ET upscaling including artificial neural network, Cubist, deep belief network, random forest, and support vector machine. The machine learning methods are trained and tested at 36 flux towers sites (65 site years) across the Heihe River Basin and are then applied to estimate ET for each grid cell (1 km × 1 km) within the watershed and for each day over the period 2012–2016. The artificial neural network, Cubist, random forest, and support vector machine algorithms have almost identical performance in estimating ET and have slightly lower root-mean-square error than deep belief network at the site scale. The random forest algorithm has slightly lower relative uncertainty at the regional scale than other methods based on three-cornered hat method. Additionally, the machine learning methods perform better over densely vegetated conditions than barren land or sparsely vegetated conditions. The regional ET generated from the machine learning approaches captured the spatial and temporal patterns of ET at the regional scale.

Key Points:
- Evapotranspiration (ET) was upscaled from 36 flux tower sites to the regional scale with five machine learning methods
- The five machine learning methods (ANN, Cubist, DBN, RF, and SVM) had almost identical performances in estimating ET
- The upscaled ET product captured the spatial and temporal patterns of ET at the regional scale

Correspondence to:
S. Liu,
smliu@bnu.edu.cn

Citation:
Xu, T., Guo, Z., Liu, S., He, X., Meng, Y., Xu, Z., et al. (2018). Evaluating different machine learning methods for upscaling evapotranspiration from flux towers to the regional scale. *Journal of Geophysical Research: Atmospheres*, *123*, 8674–8690. https://doi.org/10.1029/2018JD028447

Received 7 FEB 2018
Accepted 30 JUL 2018
Accepted article online 5 AUG 2018
Published online 27 AUG 2018

1. Introduction

Evapotranspiration (ET) is the sum of evaporation and plant transpiration from the Earth's surface to the atmosphere. ET is a vital climate variable in land surface processes as it uniquely links the surface energy balance (latent heat flux), water cycle (evaporation), and carbon cycle (transpiration-photosynthesis trade-off; Fisher et al., 2017). The quantification of ET for regional, continents, or the globe can improve our understanding of the water, heat, and carbon interactions and the feedback to the climate, which is important for water resource management and global change research.

Over the last two decades, several different types of approaches have been developed to estimate ET at the regional scale, such as remote sensing data-based methods (Ma et al., 2018; Song et al., 2016; Yao et al., 2014), land surface models (Dai & Coauthors, 2003; Niu et al., 2011), variational data assimilation methods (Bateni et al., 2014; Bateni & Entekhabi, 2012; Bateni et al., 2013; Xu et al., 2018), and land data assimilation systems (Xia et al., 2012; Xu et al., 2011). These methods can estimate ET at regional or global scales. However, daily ET estimates based on these algorithms showed substantial differences (Long et al., 2014), and the relative errors ranged from 14% to 44% compared with ground measurements (Kalma et al., 2008; Velpuri et al., 2013; Yao et al., 2013).

The ET products based on modeling approaches need to be examined extensively because they have large uncertainties for regional application. Ground-based observation networks (e.g., FLUXNET) and intensive experiments (e.g., TERENO and HiWATER) have been implemented to acquire ET accurately over a

variety of vegetative and climate conditions (Baldocchi & Ryu, 2001; Bogena et al., 2006; Li et al., 2013). These observations can be used to quantify ET at the ecosystem level for different vegetation types. However, they are typically sparse at both spatial and temporal scales. The representativeness of these measurements is limited to the flux tower footprint with the longitudinal dimension from several hundred meters to several kilometers (Liu et al., 2016). To quantify ET from the land surface to the atmosphere and to assess ET products at the regional scale, we need to upscale ET from flux towers to regional scales. Satellite sensors detecting two-dimensional information of land surface become an effective way to upscale ET from flux towers to large areas.

There are mainly four types of methods to upscale ground-measured ET to large scales using satellite remote sensing data. The first type of methods links ET observations from flux towers with vegetation index (e.g., normalized difference vegetation index, enhanced vegetation index), land surface temperature observations, and meteorological parameters (e.g., net radiation, air temperature, precipitation; Fang et al., 2016; Sun et al., 2011; Wang & Liang, 2008; Wang et al., 2007). The predictive equations are usually developed empirically, and the parameters in these equations are estimated with the least squares method. The second type of methods is so-called "geostatistical methods" based on kriging theoretical framework (Ge et al., 2015; Hu et al., 2015) or Bayesian theory framework (Gao et al., 2014; Qin et al., 2013). The third type of methods upscaled turbulent fluxes using semitheoretical models (Liu et al., 2016) or running theoretical models (Heinemann & Kerschgens, 2005). The fourth type of method is based on machine learning techniques (Bodesheim et al., 2018; Jung et al., 2011; Lu & Zhuang, 2010; Metzger et al., 2013; Wang et al., 2017; Xiao, Ollinger, et al., 2014; Xu et al., 2017a, 2017b; Yang et al., 2006).

Machine learning approaches have been increasingly used to estimate ET at regional scales. For example, Yang et al. (2006) estimated eight-day-averaged ET in United States by using flux tower observations from AmeriFlux network, three remote sensing variables (LST, enhanced vegetation index, and land cover), and surface shortwave radiation with support vector machine (SVM) approach. Lu and Zhuang (2010) produced a daily ET product in United States using remotely sensed data, meteorological, and flux tower observations with the artificial neural networks (ANN) technique. Jung et al. (2011) obtained monthly sensible and latent heat fluxes globally using FLUXNET observations, meteorological, and vegetation statues data with model tree ensemble method. Bodesheim et al. (2018) upscaled half-hourly FLUXNET observations and generated sensible and latent heat fluxes globally with random forest tree method. Xiao, Ollinger, et al. (2014) upscaled ET measurements from flux towers to the continental scale for North America over the period 2000–2012 using a piecewise regression approach (Xiao et al., 2008) based on Cubist. Wang et al. (2017) estimated daily latent heat flux for different plant functional types across North America using three machine learning algorithms: ANN, SVM, and multivariate adaptive regression spline. Metzger et al. (2013) and Xu et al. (2017a, 2017b) upscaled sensible and latent heat fluxes to fine spatiotemporal resolution with an environmental response function and boosted regression tree method. All of these studies trained the machine learning models with flux observations and other ground-measured variables related to ET or sensible and latent heat fluxes. The trained models were then applied to produce ET over continental or global scales with remote sensing and meteorological inputs. Although the machine learning techniques have been used to upscale ET from flux observations to large scales, some commonly used machine learning methods have not been tested and compared systematically. Moreover, the predicted regional ET estimates were not examined with independent ground observations over large scales (e.g., ET measurements from large aperture scintillometers). In addition, the relative uncertainty of these approaches has not been well evaluated over the regional scale.

In this study, five commonly used machine learning algorithms were employed to upscale tower-based ET observations to the regional scale. These machine learning methods include artificial neural network (ANN), Cubist, deep belief network (DBN), random forest (RF), and support vector machines (SVM). ET observations from 36 flux tower sites (65 site years) obtained by eddy covariance (EC) instruments across the Heihe River Basin in northwestern China were used to test the selected machine learning methods at the site scale. The five machine learning methods were assessed extensively with independent ET observations from eight groups of large aperture scintillometer (LAS) systems. Then, the three-cornered hat (TCH) method was employed to assess the relative uncertainty of each method at the regional scale. Finally, the optimal machine learning method was used to generate daily ET (defined as ETMap) during the period 2012–2016 over the whole watershed.

2. Methodology and Data

2.1. Machine Learning Algorithms

In this study, five machine learning algorithms including artificial neural network (ANN), Cubist, deep belief network (DBN), random forest (RF), and support vector machines (SVM) are trained and tested over the Heihe River Basin. In the machine learning methods, ET is the target variable, and variables that can account for the variations in ET are explanatory variables (shown in section 3).

Artificial neural network (ANN) is an algorithm that mimics the behavior of neural network of the human brain. The ANN algorithm is composed of the input layer, the hidden layer, and the output layer. Within an ANN algorithm, the data are put into the input layer to train the model, the weights are obtained in the hidden layer, and the prediction results are generated in the output layer. The weight expressed in the hidden layer is the connection strength between the hidden units without actual physical meanings (Kumar et al., 2002). An ANN algorithm can process information through adjusting relationship between a large number of nodes in the hidden layers according to the complexity of the system. ANN has the self-adaptive, self-organizing, and self-learning ability, which can express the complex nonlinear relationship between the input and output variables.

Cubist is a powerful tool based on modified regression tree theory. Cubist can generate a set of rule-based predictive models that balance the need for accurate prediction against the requirements of intelligibility (RuleQuest, 2008). Cubist models are expressed as collections of rules, where each rule has an associated multivariate linear model. Whenever a case matches a rule's conditions, the associated model is used to calculate the predicted value. Cubist models generally give better results than those produced by simple techniques such as multivariate linear regression, while also being easier to understand than neural networks. Cubist was designed to analyze big data containing millions of records with high speed and ease of use.

Deep belief network (DBN) is a deep learning method proposed by Hinton et al. (2006). DBN can learn more useful features and generate more accurate predictions through constructing a machine learning model with multiple hidden layers and massive training data. A DBN model is a deep network stack in series by a number of restricted Boltzmann machines (RBM). In the training processes, RBM is trained from low layer to high layer to generate suboptimal model initial parameters. Then, machine learning algorithms are employed to fine-tune the network so that the model converges to the optimal value.

Random forest (RF) algorithm can generate independent regression trees through randomly selecting training samples automatically (Breiman, 2001). Each independent regression tree is generated using samples selected by bootstrap sampling method. The final predictions are determined by averaging the outputs, after fixing individual trees in entity. The RF can handle the high-dimensional data and strong nonlinear problems. Since the correlation between each independent tree is low, the RF method can avoid falling into overfitting problems in practical implementation.

Support vector machine (SVM) is a machine learning method based on the principle of structural risk minimization, and it can handle nonlinear regression relationships (Vapnik, 1998). Generally, the original problem is expressed in a finite-dimensional space that is nonlinear. Thus, SVM can project the data sets of original finite-dimensional space into the much higher-dimensional space with kernel function to achieve linear regression, and obtain the global optimal solution by solving the convex two-order programming problem. SVM can be used for both classification and regression analysis. The commonly used kernel functions include polynomial kernel function, Gauss kernel function, and radial basis kernel function. The radial basis kernel function is used in this study as it outperforms others in previous studies (Wang et al., 2017).

2.2. Uncertainty Evaluation

The uncertainty of upscaled regional ET from five machine learning methods is evaluated with three-cornered hat method (TCH). The original TCH method was developed based on the hypothesis of uncorrelated variables (Gray & Allan, 1974). Tavella and Premoli (1994) proposed the generalized TCH algorithm that lifted the too restrictive hypothesis of uncorrelated variables. The generalized TCH method can be used to estimate uncertainties and relative uncertainty of the ET data sets from different models at the regional scale without any prior knowledge. The detailed information for generalized TCH method are described as follows.

The time series of the available ET data sets can be stored as $\{X_i\}_{i=1,2,\ldots,N}$. The subscript i denotes i^{th} ET data set and N is the total number of products to be evaluated. X_i is assumed to be comprised by two components: the true value (X_t) and the error (ε_i),

$$X_i = X_t + \varepsilon_i, \quad \forall_i = 1, 2, \ldots, N \tag{1}$$

In order to obtain the uncertainty of each ET data set (ε_i), we need to know the true value of each data set (X_t). However, it is hard to get the true value of ET data set in reality. The TCH method defined the differences between ET data sets (X_i) and reference ET data set (X_R) as

$$\mathbf{Y}_{i,M} = X_i - X_R = \varepsilon_i - \varepsilon_R \quad \forall_{i=1,2,\ldots,N-1} \tag{2}$$

where $\mathbf{Y}$ is stored in an $M \times (N-1)$ matrix and M is the time sample. The reference ET data set (X_R) can be arbitrary chosen in either of X_i. The covariance matrix of $\mathbf{Y}$ can be obtained by $\mathbf{S} = cov(\mathbf{Y})$. The unknown $N \times N$ covariance matrix of the individual noise $\mathbf{R}$ is related to $\mathbf{S}$ as

$$\mathbf{S} = J \cdot \mathbf{R} \cdot J^T \quad \text{with} \quad J = \left[\mathbf{Z} - a^T\right] \tag{3}$$

where $\mathbf{Z}$ is $(N-1) \times (N-1)$ identity matrix and a is $[1\ 1 \cdots 1]_{(1 \times (N-1))}$. Since the number of unknown elements is larger than the number of equations, equation (3) is ill posed that cannot be solved. The remaining free elements require a reasonable way to get the unique value. Galindo and Palacio (1999) have proposed the constrained minimization problem using the Kuhn-Tucker theorem.

Finally, the $\mathbf{R}$ matrix is obtained through the above process. The uncertainty of the time series $\{X_i\}_{i=1,2,\ldots,N}$ is the square root of the diagonal elements in the $\mathbf{R}$ matrix, and is stored as $\{\sigma_i\}_{i=1,2,\ldots,N}$. The relative uncertainty is defined as the ratio of σ_i to the mean value of X_i.

2.3. Study Area and Data Sets

The Heihe River Basin (HRB) is the second-largest endorheic river basin located in the arid and semiarid regions of northwest China. HRB covers an area of approximately 1.43×10^6 km^2. The upstream of HRB is characterized by mountainous areas with relatively high precipitation and is mainly covered by grassland (alpine meadow) and evergreen needleleaf forest (Qinghai spruce, etc.). The Desert-Oasis landscape exists in the relatively dry midstream and downstream areas. The dominant vegetation type is irrigated cropland in the downstream area and shrub-forest (*Tamarix*, *Populus euphratica*) in the lower downstream area (Figure 1a).

A prototype hydrometeorology observatory network was set up in the "Watershed Allied Telemetry Experimental Research" (WATER) experiment over HRB from 2008 to 2011 (Li et al., 2009). In the WATER experiment, three long-term eddy covariance (EC) systems, three automatic weather stations, and one set of large aperture scintillometer were established in the upstream and midstream areas. A comprehensive hydrometeorological observatory network was established in the "Heihe Watershed Allied Telemetry Experimental Research" (HiWATER) experiment in 2013. The network includes three superstations and 18 ordinary stations that cover the upstream, midstream, and downstream areas of HRB (Li et al., 2013). Moreover, a thematic experiment (Multi-Scale Observation Experiment on Evapotranspiration over Heterogeneous Land Surfaces (HiWATER-MUSOEXE)) was conducted in the midstream area to better understand the ET process under the heterogeneous surface from May to September 2012 (Li et al., 2017; Liu et al., 2016). Two nested matrices were set up in the HiWATER-MUSOEXE: one large experimental area (30 km × 30 km) and one kernel experimental area (5.5 km × 5.5 km). The large experimental area contained one superstation and four regular stations. The kernel experimental area contained 17 stations and four sets of LASs (Figure 1c). Except for WATER and HiWATER experiments, other hydrometeorological observatory experiments were also conducted at Hulugou (Chen et al., 2014), Linze (Ji et al., 2011) and Jinta sites (Wen et al., 2012).

Turbulent heat fluxes from a total of 36 flux tower sites (65 site years) were collected in HRB. Table 1 summarized 19 long term flux tower sites, and Figure 1a shows the site locations. Sites 1–5 are located in upstream of HRB, sites 6–13 are located in the midstream area, and sites 14–19 are located in the downstream area. Moreover, turbulent heat flux from 17 flux tower sites (15 cropland sites, 1 orchard site, and 1 build-up

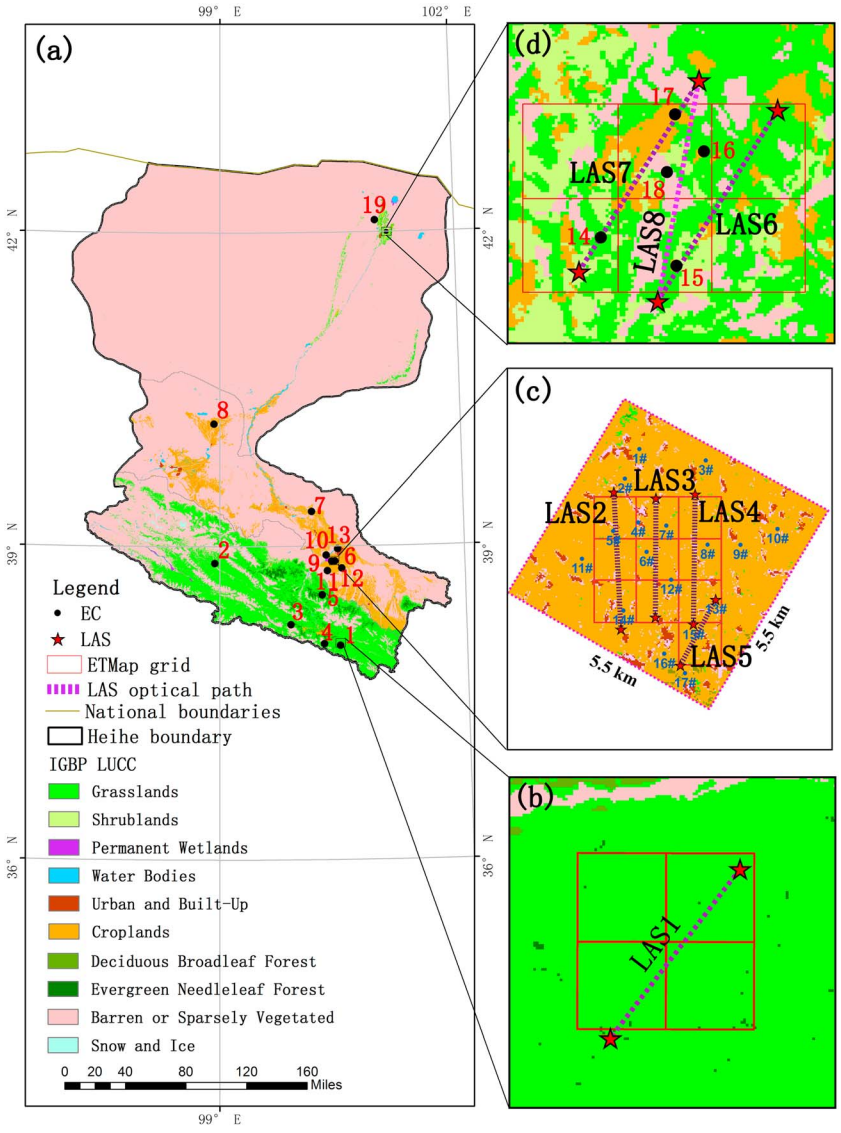

Figure 1. Land cover and flux tower site locations of the Heihe River Basin.

site) was collected during HiWATER-MUSOEXE within a 5.5 km × 5.5 km area from May to September 2012 (site locations can be found in Figure 1c). The half-hourly sensible and latent heat fluxes measured by EC instruments were collected from 36 flux tower sites (65 site years) in HRB. The half-hourly meteorological data (e.g., wind speed, air temperature, relatively humidity, solar radiation, precipitation) were also obtained with automatic weather station mounted on flux towers. Additionally, sensible heat flux was measured via eight groups of LAS instruments in HRB. The LASs measure the averaged sensible heat flux along the optical path shown in Figure 1. The LAS locations can be found in Figure 1, and the detailed information of each LAS instrument was summarized in Table 2. LAS1 is installed in the upstream area of HRB near the Arou site, LAS2–LAS5 are installed in the midstream area near the Daman site, and LAS6–LAS8 are installed in the downstream area near the Sidaoqiao site.

The daily ET from EC instruments was calculated from half-hourly latent heat flux (*LE*) measurements. Meanwhile, the Bowen ratio closure method was used to force the energy to balance when processing EC data (Twine et al., 2000). The daily *LE* from the LAS instruments was generated as the residual of the surface energy balance equation ($LE = R_n - G - H$) during the vegetation growth season. The net radiation was measured by a four-component radiometer, and the ground heat flux was measured by a ground heat flux plate. The detailed data processing procedure can be found in Liu et al. (2011, 2013) and Xu et al. (2013).

Table 1
Summary of Eddy Covariance Flux Tower Sites in the Heihe River Basin

No.	Site name	Duration	Land cover	Longitude (E)	Latitude (N)	Elevation (m)
1	Arou	6/2008–12/2016	Grassland	100.4643	38.0473	3033
2	Dashalong	8/2013–12/2016	Grassland	98.9406	38.84	3739
3	Hulugou	9/2011–12/2016	Grassland	99.8667	38.25	3232
4	Yakou	1/2015–12/2016	Grassland	100.2421	38.0142	4147
5	Guantan	1/2008–12/2011	ENF	100.2503	38.5333	2835
6	Yingke	1/2008–12/2011	Cropland	100.4103	38.8571	1519
7	Linze	4/2013–12/2014	Cropland	100.1408	39.3281	1399
8	Jinta	6/2008–8/2008	Cropland	98.9287	40.1722	1252
9	Daman	9/2012–12/2016	Cropland	100.3722	38.8555	1556
10	Bajitan Gobi	6/2012–4/2015	Barren land	100.3042	38.915	1562
11	Huazhaizi desert steppe	6/2012–12/2016	Barren land	100.3186	38.7652	1731
12	Shenshawo sandy desert	6/2012–4/2015	Barren land	100.4933	38.7892	1694
13	Zhangye wetland	6/2012–12/2016	Wetland	100.4464	38.9751	1460
14	*Populus euphratica*	7/2013–12/2015	DBF	101.1239	41.9932	876
15	Mixed forest	7/2013–12/2016	DBF	101.1335	41.9903	874
16	Sidaoqiao	7/2013–12/2016	Shrub	101.1374	42.0012	873
17	Cropland	7/2013–10/2015	Cropland	101.1338	42.0048	875
18	Barren land	7/2013–3/2016	Barren land	101.1326	41.9993	878
19	Desert	5/2015–12/2016	Barren land	100.9872	42.1137	1054

Note. ENF and DBF represent evergreen needleleaf forest and deciduous broadleaf forest

Leaf area index (LAI) data products were obtained from the Global LAnd Surface Satellite (GLASS) product (Xiao, Liang, et al., 2014; http://glass-product.bnu.edu.cn). The fractional vegetation cover (FVC) can be calculated via FVC = 1 − exp (−0.5 LAI) (Anderson et al., 1997; Norman et al., 1995; Xu et al., 2014). The land cover data were provided by Zhong et al. (2014). The regional meteorological variables (air temperature, relative humidity, precipitation, and solar radiation) were produced by the Weather Research and Forecasting model (Pan et al., 2012). The detailed information of these data is shown in Table 3. The land cover data were aggregated from 30 m × 30 m to 1 km × 1 km with the majority resampling method. The regional meteorological variables were resampled from 5 km × 5 km to 1 km × 1 km with bilinear interpolation method. The LAI data were interpolated from 8 days to daily; the air temperature, relative humidity, and solar radiation were averaged from hourly to daily; and the precipitation was accumulated from hourly to 30-day time step.

3. Experiment Setup

ET is affected by atmosphere factors including energy, water, and land surface vegetation cover conditions. Thus, air temperature (T_a), relative humidity (RH), solar radiation (R_s), precipitation (P), and leaf area index (LAI) were employed to predict daily ET over HRB. For the model training, all these variables and EC-derived ET from 36 sites over HRB were input to the five machine learning algorithms. For model application, regional daily ET (with spatial resolution 1 km × 1 km) was generated by using trained model combined with regional explanatory variables in Table 3.

Table 2
Summary of Large Aperture Scintillometer Stations in the Heihe River Basin

Site name	Instrument type, manufactures	Duration	Optical path length (m)	Height (m)
LAS1	BLS450, Scintec Germany	03/2008–12/2016	2390	9.5
LAS2	BLS900, Scintec Germany	06/2012–09/2012	3256	33.45
LAS3	BLS900, Scintec Germany	06/2012–09/2012	2841	33.45
LAS4	BLS900, Scintec Germany	06/2012–09/2012	3111	33.45
LAS5	BLS450, Scintec Germany	06/2012–12/2016	1854	22.45
LAS6	BLS900, Scintec Germany	07/2013–04/2015	2390	25.5
LAS7	BLS900, Scintec Germany	09/2013–04/2015	2380	25.5
LAS8	BLS900,Scintec Germany	04/2015–12/2016	2350	25.5

Table 3
Summary of Regional Data Collected in the Heihe River Basin

Products	Spatial resolution	Temporal resolution	References
Land cover	30 m × 30 m	Monthly	Zhong et al. (2014)
Leaf area index	1 km × 1 km	8 days	Xiao, Liang, et al. (2014)
Air temperature	5 km × 5 km	Hourly	Pan et al. (2012)
Relative humidity	5 km × 5 km	Hourly	
Precipitation	5 km × 5 km	Hourly	
Solar radiation	5 km × 5 km	Hourly	

Since the five type of input training data are not at the same order of magnitude, a normalization is needed to eliminate the difference in the order of magnitude among each data. Otherwise, the machine learning algorithms may be affected heavily by the input variables with huge magnitude. The following equation is adopted to normalize all the data range [−1, 1]:

$$N_i = \frac{X_i - X_{\text{avg}}}{X_{\max} - X_{\min}} \tag{4}$$

where N_i is the normalized data; X_i is the original data; and X_{avg}, $X_{\max}$, and $X_{\min}$ are the average, maximum, and minimum values of original data, respectively.

ET was trained and predicted under different land cover types with the following equation:

$$\text{ET}_i = M\left(R_{si}, \sum_{i-30}^{i} P_i, \text{LAI}_i, T_{ai}, \text{RH}_i\right) \tag{5}$$

where $M(\bullet)$ is the training model; ET_i is the target variable; R_s, P, LAI, T_a, and RH are the explanatory variables; and P is the cumulative precipitation for the 30-day period prior to ith day. The subscript i stands for ith day.

The global k-fold testing was adopted to examine the performance of each machine learning method. In this study, 36 sites (65 site years) of flux tower observations were divided into k parts (k = 10 herein). In each training, $k - 1$ parts of data were used as the training data, and the kth part of data was used as the validation data. The cross-validation process was repeated k times (all the data set should be trained and validated), and the results of the evaluation index were then averaged as the assessing results. Although k-fold testing is time consuming for operational application, the data are fully used and the model can lead to more accurate results than the residual testing method (Karimi et al., 2017; Marti et al., 2011; Shiri et al., 2014).

Five sensitivity tests were conducted to explore the ET prediction performances by using different explanatory variables (Figure 2). The coefficient of determination (R^2) and root-mean-square error (RMSE) were used to measure the different performances with five machining learning methods. The R_s was employed in test 1; the P, LAI, T_a, and RH were added successively from test 2 to test 5. As indicated in Figure 2, R^2 (RMSE) increased (decreased) with explanatory variables successively added. The R^2 (RMSE) reached their largest (lowest) values when all the variables were included in test 5, which indicate that the ET can be predicted accurately with the selected explanatory variables.

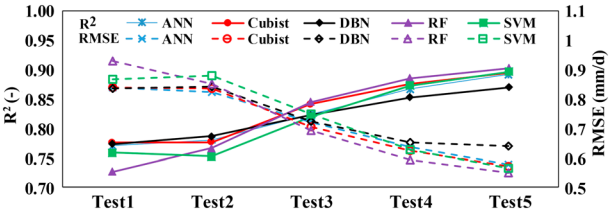

Figure 2. Sensitivity tests of ET predictions by using different explanatory variables. The R_s was employed in test 1; the P, LAI, T_a, and RH were added successively from test 2 to test 5.

4. Results

Figure 3 shows the k-fold testing results of the five machine learning methods including artificial neural network (ANN), Cubist, deep belief network (DBN), random forest (RF), and support vector machines (SVM) under all land cover types. Predicted daily ET from all five machine learning models agreed well with EC observations and fell around the 1:1 line. Additionally, the ANN, Cubist, RF, and SVM algorithms had almost identical performances. The DBN method had slightly higher RMSE (0.64 mm/day) and MAPE (12.47%), and slightly lower R^2 (0.87) than ANN, Cubist, RF, and SVM.

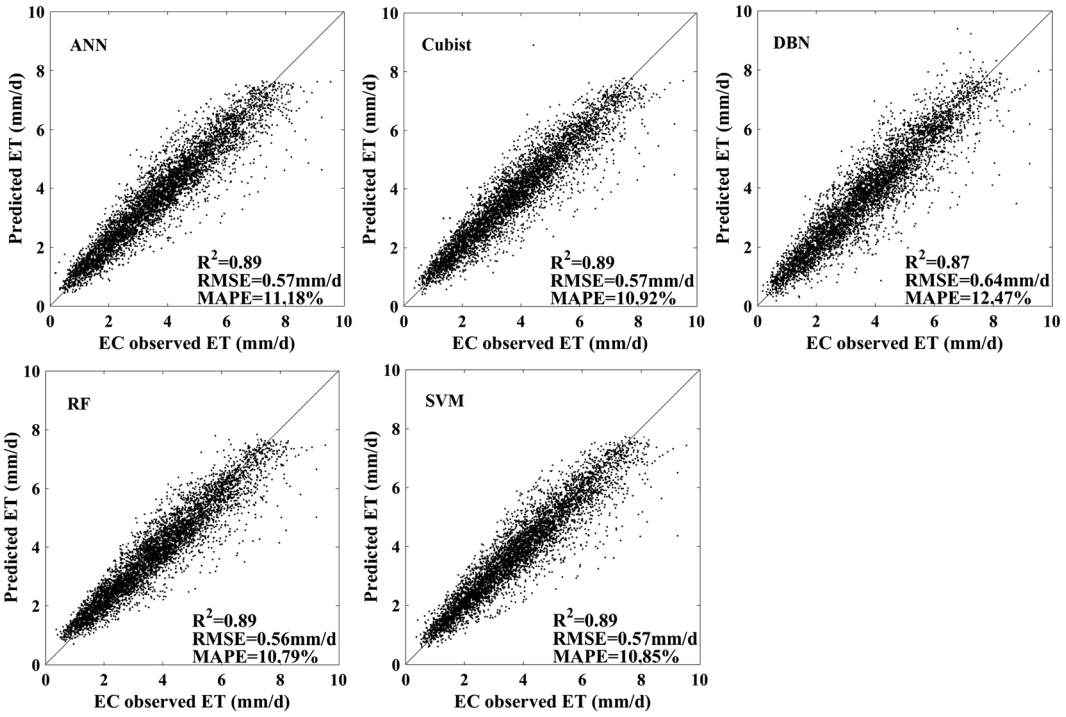

Figure 3. Performances of five machine learning algorithms for ET prediction at 36 flux tower sites.

The five trained models were applied over Heihe River Basin from May to September over the period 2012–2016. To assess the performance of five machine learning methods over the whole Heihe River Basin, the relative uncertainty of ET estimates was calculated with three-cornered hat (TCH) method. Figure 4 shows the relative uncertainties of ET predictions from five machine learning algorithms with TCH method. Generally, the RF algorithm produced slightly lower relative uncertainty than ANN, Cubist, and DBN and much lower relative uncertainty than SVM. The SVM method generated largest relative uncertainties in barren land areas. The relative uncertainty of ET was lower in the upper reaches and the oasis areas in the middle and lower reaches, while the largest relative uncertainty was found in the northern barren land areas.

Figure 5 summarizes the relative uncertainties of ET estimates from five machine learning methods with different FVC, land cover, and seasonal variations. As shown, the ET relative uncertainty increases as the FVC decreases. The relative uncertainty is the highest when the FVC is less than 10%. ET relative uncertainty is low for crop, grassland deciduous broadleaf forest, and evergreen needleleaf forest, while relative uncertainty is high for barren land. The relative uncertainty of deciduous broadleaf forest is higher than evergreen needleleaf forest. The deciduous broadleaf forest (*P. euphratica*) is sparsely distributed in downstream of the basin with low FVC compared with evergreen needleleaf forest. For seasonal variations, ET relative uncertainty is low from June to August (densely vegetated), while relative uncertainty is high for May and September (early and late vegetation growing season). The five machine learning methods perform well over densely vegetated conditions and degrade over barren land or lower FVC conditions.

As indicated, RF method slightly outperformed other machine learning methods for site training assessment with EC observations (Figure 3) and relative uncertainty evaluation with TCH method over the whole HRB (Figures 4 and 5). Thus, the RF method was used to produce daily ET over HRB from 2012 to 2016. The regional ET predicted from RF algorithm was referred to as "ETMap" here after. Figure 6 shows the spatial distribution and seasonal variation of ETMap during the growing season in the HRB over the period 2012–2016. ET increased from May to July and then decreased from July to September over the whole basin; this can be seen clearly in the upper reaches and oasis areas in the middle and lower reaches. In July and August, there was a sharp decrease of ET from south to north (upstream to downstream). In the upper reaches of HRB (Qilian Mountain areas), ET was large because of high precipitation and vegetation covers, while ET was low over barren land in middle and lower reaches due to rare

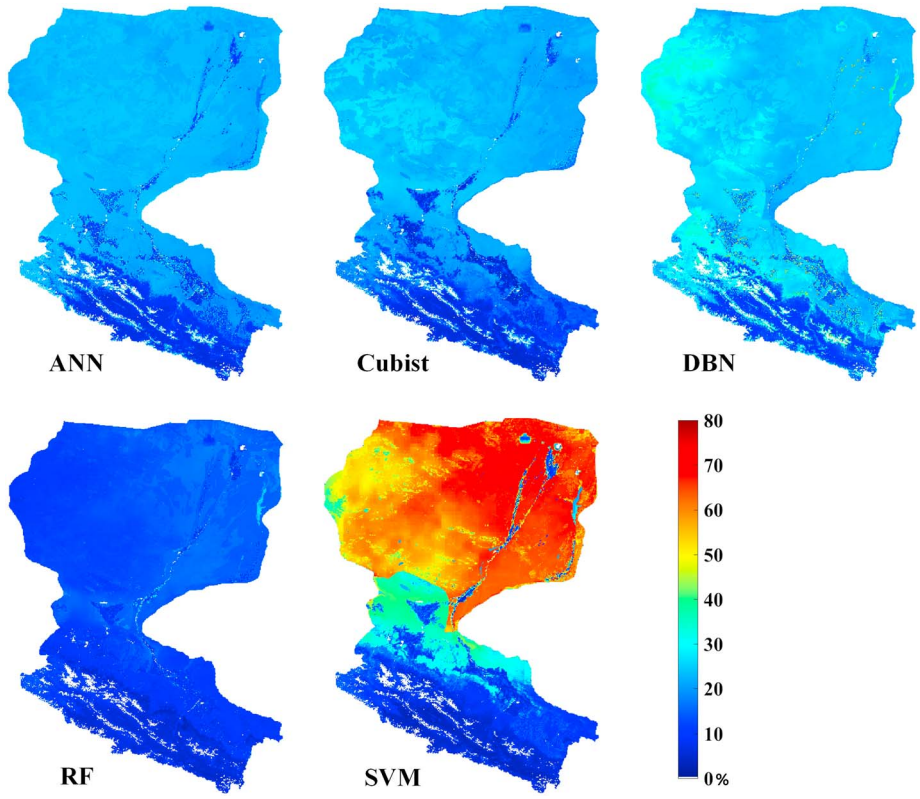

Figure 4. Relative uncertainties of ET predicted from five machine learning methods with three cornered hat (TCH) method over Heihe River Basin.

rainfalls and sparse vegetation. In the oasis areas of the middle reaches, ET was high because of heavy irrigation from the Heihe River. ET was also high along the Heihe River due to transpiration of the sparse riparian forest and evaporation. Similarly, high ET was also found around the terminal lake of HRB due to transpiration of shrub/forest (*P. euphratica* and *Tamarix*).

Figure 7 describes ET spatial distribution generated from ETMap over the whole HRB from south to north (left) and upstream HRB (Qilian Mountain areas) from low altitude to high altitude (right). The precipitation and vegetation cover ratio are also shown in Figure 7. ET was the highest in the upper reaches (south) and decreased from upstream to downstream areas (north). The key factors (precipitation and vegetation cover) that affect ET distribution also had the similar pattern as ET. Additionally, there are two ET peaks for latitude 39.5°N–40.3°N and 42.0°N–42.5°N. The high ET is observed in two big oasis areas that were irrigated with water from the Heihe River. In the upper reaches of HRB, ET increase with elevation due to the increasing

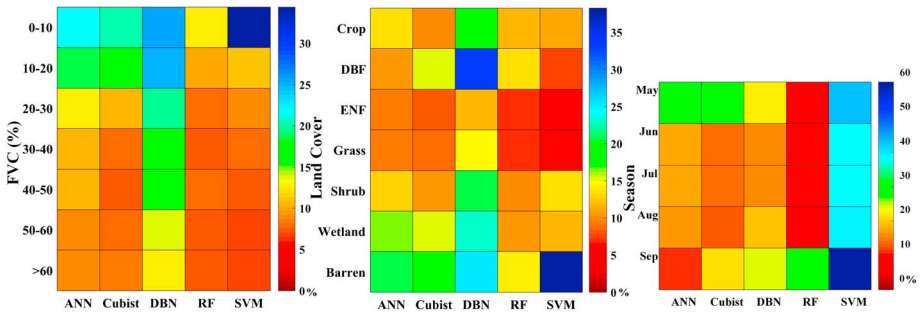

Figure 5. Relative uncertainties of ET from five machine learning methods with different factional vegetation cover (FVC), land cover, and seasonal variations over Heihe River Basin.

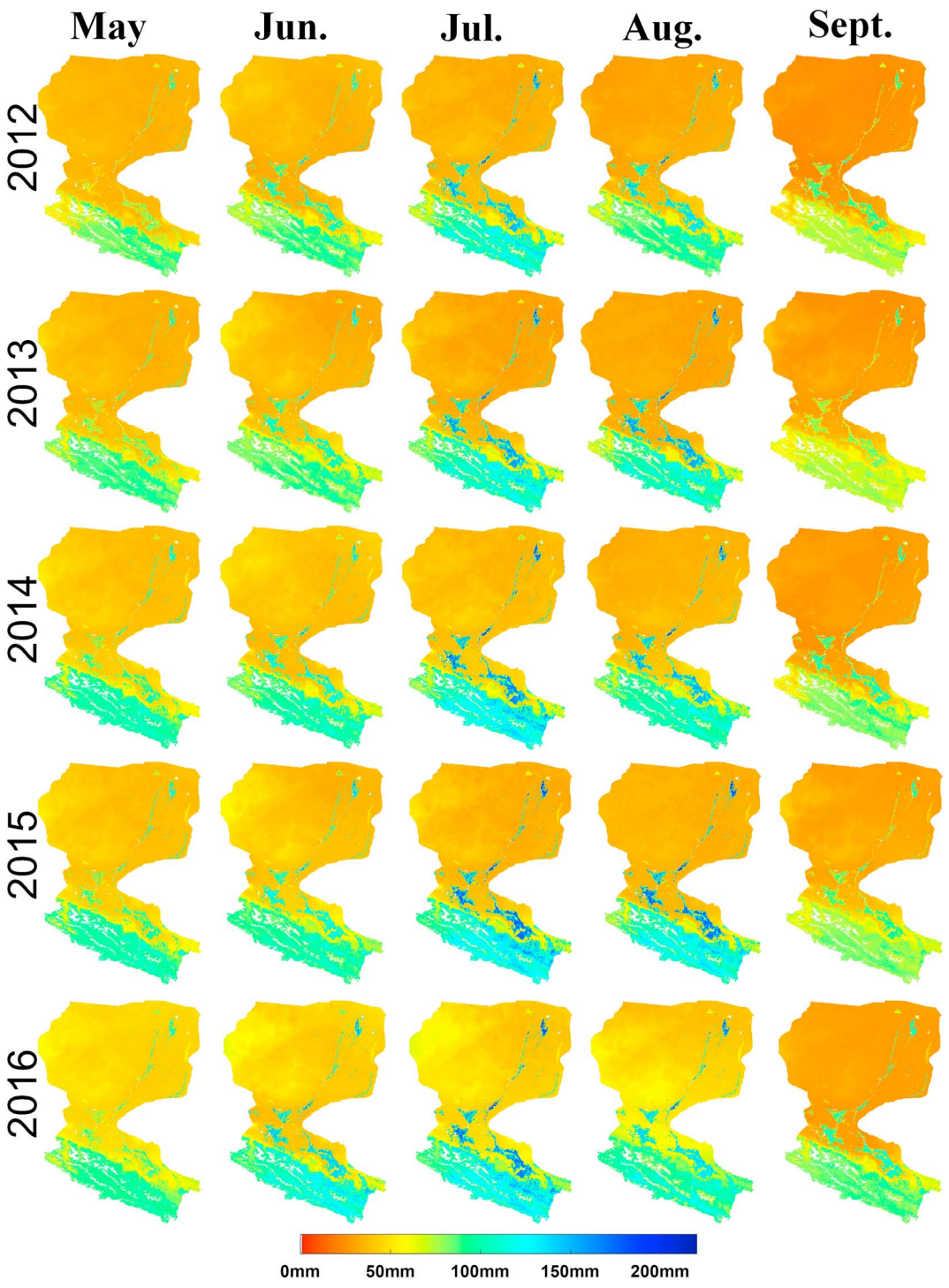

Figure 6. Spatial distribution of ET predicted by random forest method (ETMap) over the Heihe River Basin during the period 2012–2016.

precipitation and vegetation cover from 900 to 3,000 m and then decrease with elevation due to decreasing precipitation and vegetation cover for elevation above 3,000 m. The highest ET values occurred around the elevation of 2,800–3,200 m, where vegetation coverage and precipitation reached their peak values. All of these indicated that the spatial patterns of ET from ETMap agreed well with hydrological and vegetative conditions over HRB.

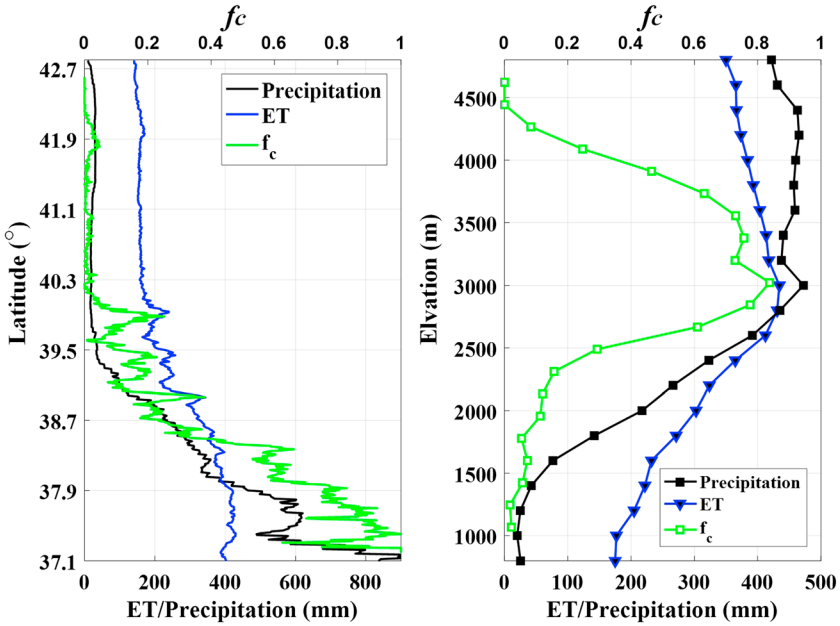

Figure 7. Changes of evapotranspiration (ET) (a) with latitude over whole Heihe River Basin and (b) with elevation for the upper reaches of the basin.

The monthly averaged ET showed seasonal variations across the main land cover types: cropland, grassland, populus forest, Qinghai spruce, and desert (Figure 8). ET increased from May to July and then decreased to September. The cropland showed obvious seasonal variations, while desert land ET did not change much from May to September due to low precipitation. The cropland had the largest magnitude of ET among all land cover types because of heavy irrigation. Moreover, the ET variation (e.g., interquartile range in the box plot) was larger over cropland areas than that over other land cover types (e.g., grassland). The rotation irrigation in the midstream oasis areas led to heterogeneous water and heat conditions over cropland, thus leading to large ET variations over cropland areas.

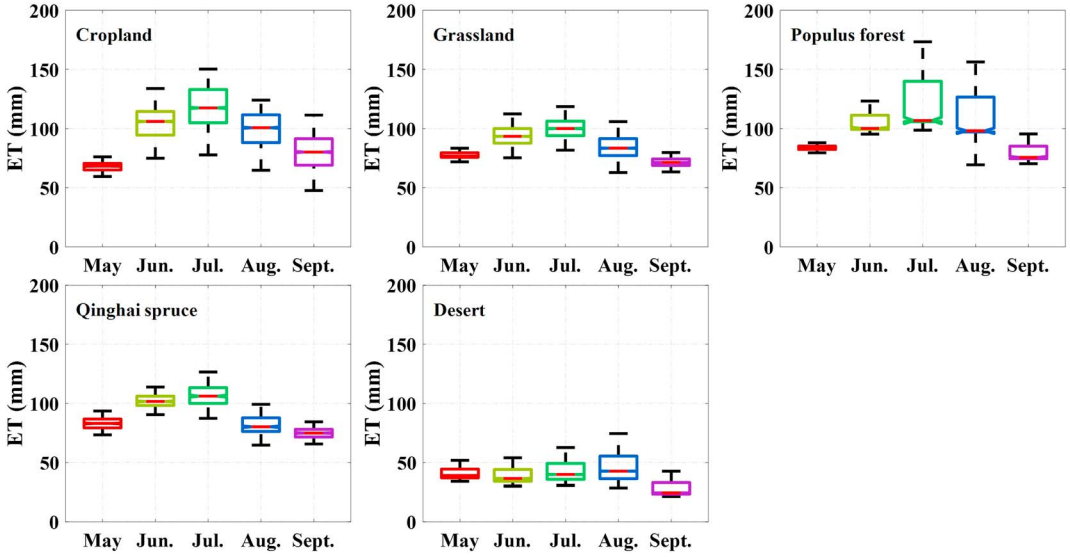

Figure 8. Seasonal variations of ET from ETMap over the main land covers in the Heihe River Basin.

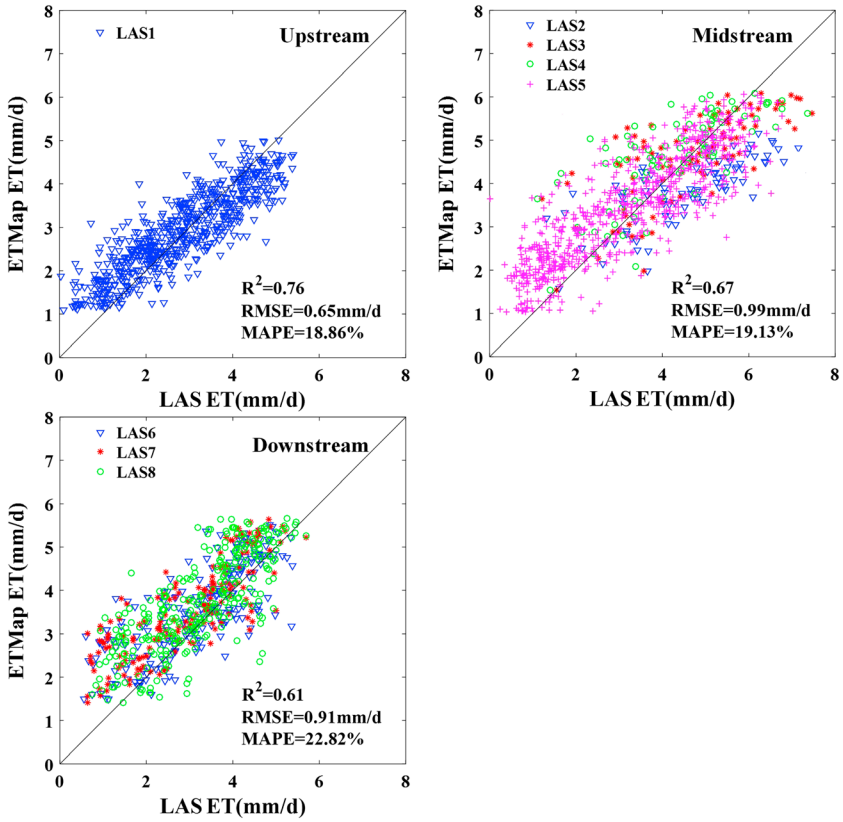

Figure 9. Comparison of daily ET derived from ETMap and eight groups of LAS observations.

The daily ET derived from ETMap was compared with independent large-scale ground measurements (eight groups of LAS observations) over Heihe River Basin (HRB; Figure 9). The LAS instrument can measure spatial-averaged surface flux along the optical path. As indicated in Figure 1, LAS1 covers 2 × 2 grids (Figure 1b); LAS2-LAS4 cover 3 × 1 grids, while LAS5 and LAS8 cover 2 × 1 grids (Figures 1c and 1d); and LAS6 covers the right 2 × 2 grids, while LAS7 covers the left 2 × 2 grids, respectively (Figure 1d). The LAS observations were compared to spatially averaged ET values from model predictions with same spatial representativeness. Figure 9 shows ET comparison scatterplots between ETMap and LAS observations. As indicated, the ET from ETMap agreed well with LAS observations and mainly fell around the 1:1 line. The RMSE (MAPE) over the upstream (LAS1), midstream (LAS2–LAS5), and downstream (LAS6–LAS8) areas were 0.65 mm/day (18.86%), 0.99 mm/day (19.13%), and 0.91 mm/day (22.82%), respectively. The discrepancies between ETMap and LAS observations were mainly caused by uncertainties in training data, LAS observations, and the heterogeneous land surfaces. Moreover, the RMSE and MAPE over the upstream area (LAS1) were lower than those values over the midstream (LAS2–LAS5) and downstream (LAS6–LAS8) areas. The land surface at LAS1 is dominated by grassland and is relatively homogeneous. However, land surface is heterogeneous around LAS2–LAS5 (land cover is composed by cropland and buildup) and LAS6–LAS8 (land cover is mainly composed by cropland, barren land, forest, shrub land, etc.). The heterogeneous land surface condition is a key factor for the discrepancies between ETMap and LAS observations.

Figure 10 shows the time series of daily ET derived from ETMap over the LAS1 and LAS5–LAS7 stations during the growing season in 2014. Daily ET from the LAS observations and precipitation observations were also shown in Figure 10. Daily ET derived from ETMap was consistent with LAS observations in terms of both magnitude and day-to-day dynamics, implying that the trained model with the RF method performed well at the regional scale. Daily ET from ETMap decreased at rainy days (with precipitation). The good agreement between the ET prediction and LAS observation illustrates that the ETMap can effectively estimate ET at the regional scale by using EC ground measurement over different hydrology and vegetative conditions.

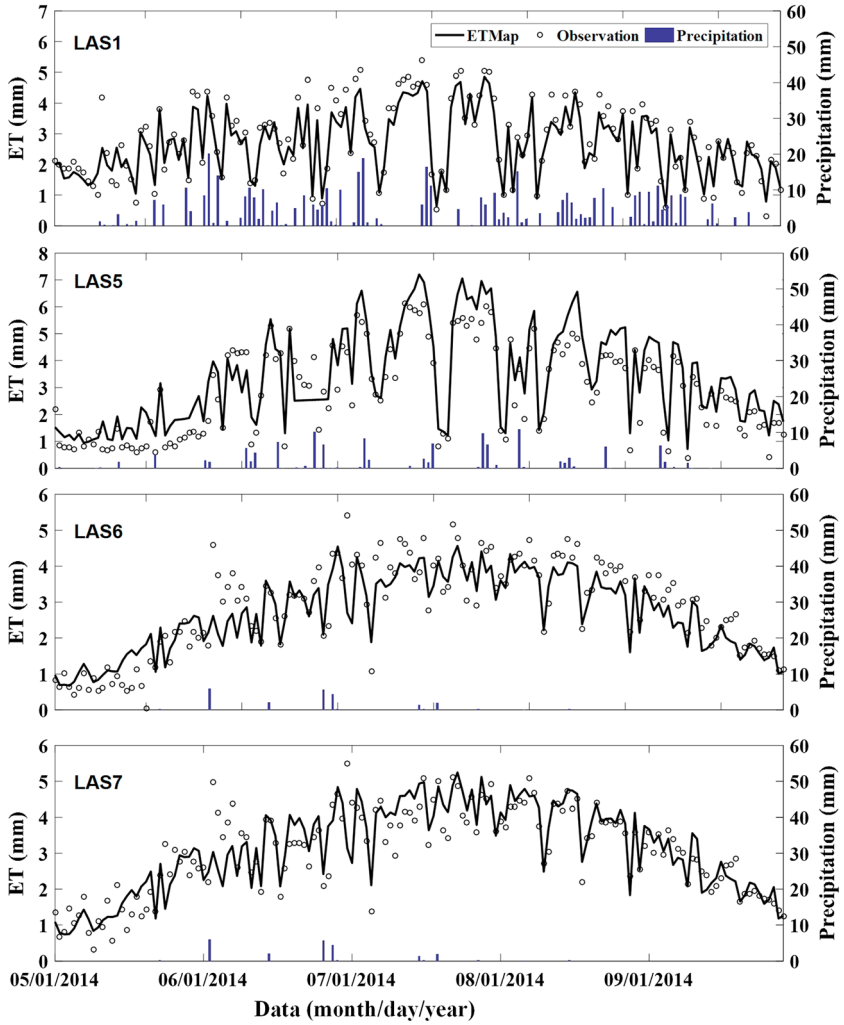

Figure 10. Time series of daily ET from ETMap and LAS observations during 2014.

5. Discussion

Five machine learning methods (ANN, Cubist, DBN, RF, and SVM) were compared for ET upscaling over Heihe River Basin. Because the five selected methods were commonly applied and have different learning strategy (introduced in section 2.1). Some other machine learning methods (model tree ensembles, boosted regression trees, etc.) have been reported good performances for ET and other variable prediction (Jung et al., 2011; Metzger et al., 2013; Xu et al., 2017a, 2017b). However, these methods are similar to Cubist and RF methods that we have already used in this study. Our results showed that the five machine learning approaches have almost identical performances in upscaling ET observations from EC flux towers to the regional scale. These methods have been widely used to upscale ET observations to broad spatial scales (Bodesheim et al., 2018; Lu & Zhuang, 2010; Wang et al., 2017; Xiao, Ollinger, et al., 2014; Yang et al., 2006). Our results show that overall, it does not matter much which method is used for regional upscaling of ET observations. Meanwhile, our results also identified some relatively minor differences in performance among these methods. These minor differences are mainly caused by different machine learning strategy as same training data were used for different machine learning algorithms.

The RF algorithm can generate independent regression trees through randomly selecting training samples automatically (Breiman, 2001). Thus, the RF algorithm can handle the high-dimensional data and strong nonlinear problems. In Heihe River Basin, the land cover is complex and diverse (grassland, evergreen needleleaf forest, deciduous broadleaf forest, cropland, wetland, barren land, etc.; Li et al., 2013). Thus, the RF produced

regional ET with lower uncertainty than other algorithms. Since the DBN algorithm is a deep learning method, it needs massive training data for model training to "optimal" (Hinton et al., 2006). In Heihe River Basin, the 36 site observations (65 site years; sample number is 65 site year × 365 day/year = 23,725) is not large enough for DBN model training. Thus, the DBN algorithm underperformed other four algorithms in this study.

ET models were trained with land surface and meteorological variables including solar radiation (R_s), precipitation (P), leaf area index (LAI), air temperature (T_a), and relative humidity (RH). Results showed that ET models were sensitive to these variables and could generate accurate results by incorporating all these variables. ET is also a function of other variables, such as land surface temperature, wind speed, and soil type (or water storage capacity). However, the remotely sensed regional LST is often contaminated by cloud; it is hard, if not impossible, to get regional gap-free LST for every day. Except for the selected variables, wind speed is also added as an additional explanatory variable for ET prediction. However, the RMSE for ET prediction was increased by including wind speed (not shown in Figure 2). The soil type (or water storage capacity) is one of the key limiting factors for soil moisture and ET prediction over humid regions. The soil type can represent soil wetness distribution over regional scale. However, the precipitation is always low over dry regions where soil type cannot represent soil wetness well. The soil type is not suitable for ET prediction as there is a big desert area in Heihe River Basin (Figure 1). Therefore, LST, wind speed, and soil type (or water storage capacity) were not included for ET prediction.

The discrepancies between model predictions and observations were mainly caused by machine learning methods, observation data uncertainty, and land surface heterogeneity (Figures 3 and 9). The machine learning methods train the prediction model without physical meanings, and neglect the real interactions between explanatory variables and the target. Thus, the machine learning method generates uncertainty in daily ET prediction. Moreover, the uncertainty in latent heat flux observations derived from EC instruments are 16% (Wang et al., 2015). The flux observations from LAS instruments are in good agreement with that from EC instruments (Liu et al., 2016; Xu et al., 2013), which represent similar uncertainty in LAS and EC observations. In addition, the larger misfits between LAS observation and ET estimation were found over downstream compared those from upstream (Figure 9). The land surface heterogeneity over downstream may lead to this large misfit (Figure 1b).

The relative uncertainties of ET estimates from five machine learning methods are assessed by TCH method over Heihe River Basin (Figure 4). The five machine learning methods perform well over densely vegetated conditions and degrade over barren land or lower FVC conditions (Figure 5), because ET is mainly explained by meteorological variables (e.g., radiation, air temperature) via equation (5) over sparsely vegetated conditions (low LAI or FVC). Since LAI is an important variable for model training and regional application, the machine learning methods are degraded with low LAI values. Hence, ET relative uncertainties are high over sparsely vegetated conditions (low LAI/FVC, barren land, early and late vegetation growing season).

Our results can inform future upscaling efforts based on machine learning approaches. First, our analysis demonstrated that the five machine learning methods (ANN, Cubist, DBN, RF, and SVM) have almost identical performances in predicting ET. Any of these methods can be used to upscale ET measurements to the regional scale and thereby to estimate regional ET fairly well. Second, all these methods have higher relative uncertainty for sparsely vegetated areas than for densely vegetated areas. It should be noted that, however, the lower performance in sparsely vegetated areas is not due to an inherent shortcoming of machine learning methods but due to the weaker signal of vegetation and larger influences of soil background in the remotely sensed data used for these areas and the higher land surface heterogeneity. The use of finer spatial resolution soil moisture products are expected to improve the accuracy of ET prediction for sparsely vegetated areas. Third, RF has slightly lower relative uncertainty at the regional scale than other methods in our study region, but this is not conclusive and could be region specific. Finally, to train the model to optimal DBN requires massive training data. We recommend that a method other than DBN be used if there are only a very limited number of measurements available for training.

6. Conclusions

Evapotranspiration (ET) is upscaled from flux towers to the regional scale with five machine learning algorithms. The five machine learning methods include artificial neural network (ANN), Cubist, deep belief network (DBN), random forest (RF), and support vector machines (SVM). The daily ET derived from eddy

covariance (EC) instruments over 36 flux tower sites (65 site years) are taken as the training target. The solar radiation (R_s), precipitation (P), leaf area index (LAI), air temperature (T_a), and relative humidity (RH) are taken as the explanatory variables that related to ET. The machine learning algorithms are trained over 36 flux tower sites (65 site years), and then applied over whole Heihe River Basin (HRB) from 2012 to 2016.

Five machine learning algorithms are assessed with *k*-fold testing method with EC observations. The RF, Cubist, ANN, and SVM methods had almost identical performances in estimating ET and had slightly lower RMSE and MAPE than DBN. The trained models are applied over whole HRB from 2012 to 2016. The three-cornered hat (TCH) method is employed to quantify relatively uncertainties of daily ET from different machine learning algorithms over whole watershed. Our results indicate that RF algorithm generates slightly lower relative uncertainty than other machine learning approaches. Additionally, the machine learning methods perform better over densely vegetated conditions than barren land or sparsely vegetated conditions. For future studies, the spatiotemporal continuous land surface variables (e.g., land surface temperature, soil moisture) at fine resolution might be added as explanatory variables, if available, to enhance the modeling skill over sparsely vegetated conditions.

Regional ET (ETMap) was obtained through upscaling observations from 36 flux tower sites. The upscaled ET had high accuracy based on the validation with LAS-derived ET data. ETMap is a useful product for ET-related research (e.g., ET spatial distribution, carbon-water interaction) and for validation of ET products from physical models over the Heihe River Basin.

Acknowledgments
This work is supported by the National Natural Science Foundation of China (41531174) and National Basic Research Program of China (2015CB953702). J. Xiao was supported by the National Aeronautics and Space Administration (NASA; Climate Indicators and Data Products for Future National Climate Assessments, grant NNX16AG61G). We would like to thank all the scientists, engineers, and students who participated in WATER and HiWATER field campaigns. The ground-measured turbulent heat fluxes, meteorological variables, and land cover data are downloaded freely via Heihe Data Center (http://www.heihedata.org/). Leaf area index (LAI) data are available on the Beijing Normal University data center (http://glass-product.bnu.edu.cn/). We thank the three anonymous reviewers for their helpful comments on the manuscript.

References

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 Sensing of Environment*, *60*(2), 195–216. https://doi.org/10.1016/S0034-4257(96)00215-5

Baldocchi, D., & Ryu, Y. (2001). In D. C. M. D. E. Levia & T. Tanaka (Eds.), *A synthesis of forest evaporation fluxes-from days to years-as measured with eddy covariance, forest hydrology and biogeochemistry* (pp. 101–116). Berlin: Springer.

Bateni, S. M., & Entekhabi, D. (2012). Relative efficiency of land surface energy balance components. *Water Resources Research*, *48*, W04510. https://doi.org/10.1029/2011WR011357

Bateni, S. M., Entekhabi, D., & Jeng, D. S. (2013). Variational assimilation of land surface temperature and the estimation of surface energy balance components. *Journal of Hydrology*, *481*, 143–156. https://doi.org/10.1016/j.jhydrol.2012.12.039

Bateni, S. M., Entekhabi, D., Margulis, S., Castelli, F., & Kergoat, L. (2014). Coupled estimation of surface heat fluxes and vegetation dynamics from remotely sensed land surface temperature and fraction of photosynthetically active radiation. *Water Resources Research*, *50*, 8420–8440. https://doi.org/10.1002/2013WR014573

Bodesheim, P., Jung, M., Gans, F., Mahecha, M. D., & Reichstein, M. (2018). Upscaled diurnal cycles of land-atmosphere fluxes: A new global half-hourly data product. *Earth System Science Data Discussions*, 1–47. https://doi.org/10.5194/essd-2017-130

Bogena, H., Schulz, K., & Vereecken, H. (2006). Towards a network of observatories in terrestrial environmental research. *Advances in Geosciences*, *9*, 1–6.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chen, R., Song, Y., Kang, E., Han, C., Liu, J., Yang, Y., et al. (2014). A cryosphere-hydrology observation system in a small alpine watershed in the Qilian Mountains of China and its meteorological gradient. *Arctic, Antarctic, and Alpine Research*, *46*(2), 505–523. https://doi.org/10.1657/1938-4246-46.2.505

Dai, Y. J., & Coauthors (2003). The common land model. *Bulletin of the American Meteorological Society*, *84*(8), 1013–1024. https://doi.org/10.1175/BAMS-84-8-1013

Fang, Y., Sun, G., Caldwell, P., McNulty, S. G., Noormets, A., Domec, J. C., et al. (2016). Monthly land cover-specific evapotranspiration models derived from global eddy flux measurements and remote sensing data. *Ecohydrology*, *9*(2), 248–266. https://doi.org/10.1002/eco.1629

Fisher, J. B., Melton, F., Middleton, E., Hain, C., Anderson, M., Allen, R., et al. (2017). The future of evapotranspiration: Global requirements for ecosystem functioning, carbon and climate feedbacks, agricultural management, and water resources. *Water Resources Research*, *53*, 2618–2626. https://doi.org/10.1002/2016WR020175

Galindo, F. J., & Palacio, J. (1999). Estimating the instabilities of N correlated clocks, in Proc. of the 31st annual precise time and time interval meeting, pp. 285–296.

Gao, S. G., Zhu, Z. L., Liu, S. M., Jin, R., Yang, G. C., & Tan, L. (2014). Estimating the spatial distribution of soil moisture based on Bayesian maximum entropy method with auxiliary data from remote sensing. *International Journal of Applied Earth Observation*, *32*(1), 54–66. https://doi.org/10.1016/j.jag.2014.03.003

Ge, Y., Liang, Y. Z., Wang, J. H., Zhao, Q. Y., & Liu, S. M. (2015). Upscaling sensible heat fluxes with area-to-area regression kriging. *IEEE Transactions on Geoscience and Remote Sensing*, *12*(3), 656–660.

Gray, J. E., & Allan, D. W. (1974). A method for estimating the frequency stability of an individual oscillator. Symposium on frequency control IEEE, 243-246.

Heinemann, G., & Kerschgens, M. (2005). Comparison of methods for area-averaging surface energy fluxes over heterogeneous land surfaces using high-resolution non-hydrostatic simulations. *International Journal of Climatology*, *25*(3), 379–403. https://doi.org/10.1002/joc.1123

Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, *18*(7), 1527–1554. https://doi.org/10.1162/neco.2006.18.7.1527

Hu, M. G., Wang, J. H., Ge, Y., Liu, M. X., Liu, S. M., Xu, Z. W., & Xu, T. R. (2015). Scaling flux tower observations of sensible heat flux using weighted area-to-area regression Kriging. *Atmosphere*, *6*(8), 1032–1044. https://doi.org/10.3390/atmos6081032

Ji, X. B., Zhao, W. Z., & Kang, E. S. (2011). Carbon dioxide exchange in an irrigated agricultural field within an oasis, northwest China. *Journal of Applied Meteorology and Climatology*, *50*(11), 2298–2308. https://doi.org/10.1175/2011JAMC2614.1

Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., et al. (2011). Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *Journal of Geophysical Research*, *16*, G00J07. https://doi.org/10.1029/2010JG001566

Kalma, J. D., Mcvicar, T. R., & Mccabe, M. F. (2008). Estimating land surface evaporation: A review of methods using remotely sensed surface temperature data. *Surveys in Geophysics*, *29*(4-5), 421–469. https://doi.org/10.1007/s10712-008-9037-z

Karimi, S., Shiri, J., Kisi, O., & Xu, T. (2017). Forecasting daily streamflow values: Assessing heuristic models. *Hydrology Research*, *49*(3), 658–669. https://doi.org/10.2166/nh.2017.111

Kumar, M., Raghuwanshi, N. S., Singh, R., Wallender, W. W., & Pruitt, W. O. (2002). Estimating evapotranspiration using artificial neural network. *Journal of Irrigation and Drainage Engineering*, *128*(4), 224–233. https://doi.org/10.1061/(ASCE)0733-9437(2002)128:4(224)

Li, X., Cheng, G. D., Liu, S. M., Xiao, Q., Ma, M. G., Jin, R., et al. (2013). Heihe watershed allied telemetry experimental research (HiWATER): Scientific objectives and experimental design. *Bulletin of the American Meteorological Society*, *94*(8), 1145–1160. https://doi.org/10.1175/BAMS-D-12-00154.1

Li, X., Li, X., Li, Z., Ma, M., Wang, J., Xiao, Q., et al. (2009). Watershed allied telemetry experimental research. *Journal of Geophysical Research*, *114*, D22103. https://doi.org/10.1029/2008JD011590

Li, X., Liu, S. M., Xiao, Q., Ma, M. G., Jin, R., Che, T., et al. (2017). A multiscale dataset for understanding complex eco-hydrological processes in a heterogeneous oasis system. *Scientific Data*, *4*, 170083. https://doi.org/10.1038/sdata.2017.83

Liu, S. M., Xu, Z. W., Song, L. S., Zhao, Q. Y., Ge, Y., Xu, T. R., et al. (2016). Upscaling evapotranspiration measurements from multi-site to the satellite pixel scale over heterogeneous land surfaces. *Agricultural and Forest Meteorology*, *230*, 97–113.

Liu, S. M., Xu, Z. W., Wang, W. Z., Jia, Z. Z., Zhu, M. J., Bai, J., & Wang, J. M. (2011). A comparison of eddy-covariance and large aperture scintillometer measurements with respect to the energy balance problem. *Hydrology and Earth System Sciences*, *15*(4), 1291–1306. https://doi.org/10.5194/hess-15-1291-2011

Liu, S. M., Xu, Z. W., Zhu, Z. L., Jia, Z. Z., & Zhu, M. J. (2013). Measurements of evapotranspiration from eddy-covariance systems and large aperture scintillometers in the Hai River basin, China. *Journal of Hydrology*, *487*, 24–38. https://doi.org/10.1016/j.jhydrol.2013.02.025

Long, D., Longuevergne, L., & Scanlon, B. R. (2014). Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites. *Water Resources Research*, *50*, 1131–1151. https://doi.org/10.1002/2013WR014581

Lu, X. L., & Zhuang, Q. L. (2010). Evaluating evapotranspiration and water-use efficiency of terrestrial ecosystems in the conterminous United States using MODIS and AmeriFlux data. *Remote Sensing of Environment*, *114*(9), 1924–1939. https://doi.org/10.1016/j.rse.2010.04.001

Ma, Y., Liu, S., Song, L., Xu, Z., Liu, Y., Xu, T., & Zhu, Z. (2018). Estimation of daily evapotranspiration and irrigation water efficiency at a Landsat-like scale for an arid irrigation area using multi-source remote sensing data. *Remote Sensing of Environment*, *216*, 715–734. https://doi.org/10.1016/j.rse.2018.07.019

Marti, P., Manzano, J., & Royuela, A. (2011). Assessment of 4 input artificial neural network for ET0 estimation through data set scanning procedures. *Irrigation Science*, *29*, 181–195.

Metzger, S., Junkermann, W., Mauder, M., Butterbach-Bahl, K., Trancón, Y., Widemann, B., et al. (2013). Spatially explicit regionalization of airborne flux measurements using environmental response functions. *Biogeosciences*, *10*(4), 2193–2217. https://doi.org/10.5194/bg-10-2193-2013

Niu, G. Y., Yang, Z. L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research*, *116*, D12109. https://doi.org/10.1029/2010JD015139

Norman, J. M., Kustas, W. P., & Humes, K. (1995). Source approach for estimating soil and vegetation energy fluxes in observations of directional radiometric surface temperature. *Agricultural and Forest Meteorology*, *77*(3-4), 263–293. https://doi.org/10.1016/0168-1923(95)02265-Y

Pan, X. D., Li, X., Shi, X. K., Han, X. J., Luo, L. H., & Wang, X. (2012). Dynamic downscaling of near-surface air temperature at the basin scale using WRF—A case study in the Heihe river basin, China. *Frontiers in Earth Science*, *6*(3), 314–323. https://doi.org/10.1007/s11707-012-0306-2

Qin, J., Yang, K., Lu, N., Chen, Y., Zhao, L., & Han, M. (2013). Spatial upscaling of in-situ soil moisture measurements based on MODIS-derived apparent thermal inertia. *Remote Sensing of Environment*, *138*, 1–9. https://doi.org/10.1016/j.rse.2013.07.003

RuleQuest. (2008). http://www.rulequest.com. Visited on 10/18/2007.

Shiri, J., Marti, P., & Singh, V. P. (2014). Evaluation of gene expression programming approaches for estimating daily evaporation through spatial and temporal data scanning. *Hydrological Processes*, *28*(3), 1215–1225. https://doi.org/10.1002/hyp.9669

Song, L. S., Kustas, W. P., Liu, S., Colaizzi, P. D., Nieto, H., Xu, Z. W., et al. (2016). Applications of a thermal-based two-source energy balance model using Priestley-Taylor approach for surface temperature partitioning under advective conditions. *Journal of Hydrology*, *540*, 574–587. https://doi.org/10.1016/j.jhydrol.2016.06.034

Sun, G., Alstad, K., Chen, J., Chen, S., Ford, C. R., Lin, G., & Zhang, Z. (2011). A general predictive model for estimating monthly ecosystem evapotranspiration. *Ecohydrology*, *4*(2), 245–255. https://doi.org/10.1002/eco.194

Tavella, P., & Premoli, A. (1994). Estimating the instabilities of N clocks by measuring differences of their readings. *Metrologia*, *30*(5), 479–486. https://doi.org/10.1088/0026-1394/30/5/003

Twine, T. E., Kustas, W. P., Norman, J. M., Cook, D. R., Houser, P. R., Meyers, T. P., et al. (2000). Correcting eddy-covariance flux underestimates over grassland. *Agricultural and Forest Meteorology*, *103*(3), 279–300. https://doi.org/10.1016/S0168-1923(00)00123-4

Vapnik, V. N. (1998). *Statistical learning theory*. New York: Wiley-Interscience.

Velpuri, N. M., Senay, G. B., Singh, R. K., Bohms, S., & Verdin, J. P. (2013). A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: Using point and gridded FLUXNET and water balance ET. *Remote Sensing of Environment*, *139*(4), 35–49. https://doi.org/10.1016/j.rse.2013.07.013

Wang, J. M., Zhuang, J. X., Wang, W. Z., Liu, S. M., & Xu, Z. W. (2015). Assessment of uncertainties in eddy covariance flux measurement based on intensive flux matrix of HiWATER-MUSOEXE. *IEEE Transactions on Geoscience and Remote Sensing*, *12*(2), 259–263. https://doi.org/10.1109/LGRS.2014.2334703

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. *Journal of Hydrometeorology*, *9*(4), 712–727. https://doi.org/10.1175/2007JHM911.1

Wang, K., Wang, P., Li, Z., Cribb, M., & Sparrow, M. (2007). A simple method to estimate actual evapotranspiration from a combination of net radiation, vegetation index, and temperature. *Journal of Geophysical Research*, *112*, D15107. https://doi.org/10.1029/2006JD008351

Wang, X. Y., Yao, Y. J., Zhao, S. H., Jia, K., Zhang, X. T., Zhang, Y. H., et al. (2017). MODIS-based estimation of terrestrial latent heat flux over North America using three machine learning algorithms. *Remote Sensing*, *9*(12), 1326. https://doi.org/10.3390/rs9121326

Wen, X., Lu, S., & Jin, J. (2012). Integrating remote sensing data with WRF for improved simulations of oasis effects on local weather processes over an arid region in northwestern China. *Journal of Hydrometeorology*, *13*(2), 573–587. https://doi.org/10.1175/JHM-D-10-05001.1

Xia, Y. L., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., et al. (2012). Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2). 1: Intercomparison and application of model products. *Journal of Geophysical Research*, *117*, D03109. https://doi.org/10.1029/2011JD016048

Xiao, J. F., Ollinger, S. V., Frolking, S., Hurtt, G. C., Hollinger, D. Y., Davis, K. J., et al. (2014). Data-driven diagnostics of terrestrial carbon dynamics over North America. *Agricultural and Forest Meteorology*, *197*, 142–157. https://doi.org/10.1016/j.agrformet.2014.06.013

Xiao, J. F., Zhuang, Q., Baldocchi, D. D., Law, B. E., Richardson, A. D., Chen, J., et al. (2008). Estimation of net ecosystem carbon exchange for the conterminous United States by combining MODIS and AmeriFlux data. *Agricultural and Forest Meteorology*, *148*(11), 1827–1847. https://doi.org/10.1016/j.agrformet.2008.06.015

Xiao, Z., Liang, S., Wang, J., Chen, P., Yin, X., Zhang, L., & Song, J. (2014). Use of general regression neural networks for generating the GLASS leaf area index product from time-series MODIS surface reflectance. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(1), 209–223. https://doi.org/10.1109/TGRS.2013.2237780

Xu, K., Metzger, S., & Desai, A. R. (2017a). Upscaling tower-observed turbulent exchange at fine spatio-temporal resolution using environmental response functions. *Agricultural and Forest Meteorology*, *232*, 10–22. https://doi.org/10.1016/j.agrformet.2016.07.019

Xu, K., Metzger, S., & Desai, A. R. (2017b). Surface-atmosphere exchange in a box: Space-time resolved storage and net vertical fluxes from tower-based eddy covariance. *Agricultural and Forest Meteorology*, *255*, 81–91. https://doi.org/10.1016/j.agrformet.2017.10.011

Xu, T., Bateni, S. M., Liang, S., Entekhabi, D., & Mao, K. (2014). Estimation of surface turbulent heat fluxes via variational assimilation of sequences of land surface temperatures from geostationary operational environmental satellites. *Journal of Geophysical Research*, *119*, 10,780–10,798. https://doi.org/10.1002/2014JD021814

Xu, T., Bateni, S. M., Neale, C. M. U., Auligne, T., & Liu, S. (2018). Estimation of turbulent heat fluxes by assimilation of land surface temperature observations from GOES satellites into an ensemble Kalman smoother framework. *Journal of Geophysical Research: Atmospheres*, *123*, 2409–2423. https://doi.org/10.1002/2017JD027732

Xu, T., Liang, S., & Liu, S. (2011). Estimating turbulent fluxes through assimilation of geostationary operational environmental satellites data using ensemble Kalman filter. *Journal of Geophysical Research*, *116*, D09109. https://doi.org/10.1029/2010JD015150

Xu, Z., Liu, S., Li, X., Shi, S. G., Wang, J. M., Zhu, Z. L., et al. (2013). Intercomparison of surface energy flux measurement systems used during the HiWATER-MUSOEXE. *Journal of Geophysical Research: Atmospheres*, *118*, 13,140–13,157. https://doi.org/10.1002/2013JD020260

Yang, F., White, M. A., Michaelis, A. R., Ichii, K., Hashimoto, H., Votava, P., & Nemani, R. R. (2006). Prediction of continental-scale evapotranspiration by combining MODIS and AmeriFlux data through support vector machine. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(11), 3452–3461. https://doi.org/10.1109/TGRS.2006.876297

Yao, Y., Liang, S., Cheng, J., Liu, S., Fisher, J. B., Zhang, X., et al. (2013). MODIS-driven estimation of terrestrial latent heat flux in China based on a modified Priestley–Taylor algorithm. *Agricultural and Forest Meteorology*, *171*, 187–202.

Yao, Y., Liang, S., Li, X., Hong, Y., Fisher, J. B., Zhang, N., et al. (2014). Bayesian multimodel estimation of global terrestrial latent heat flux from eddy covariance, meteorological, and satellite observations. *Journal of Geophysical Research: Atmospheres*, *119*, 4521–4545. https://doi.org/10.1002/2013JD020864

Zhong, B., Ma, P., Nie, A., Yang, A., Yao, Y., Lv, W., et al. (2014). Land cover mapping using time series HJ-1/CCD data. *Science China Earth Sciences*, *57*(8), 1790–1799. https://doi.org/10.1007/s11430-014-4877-5