Prediction of Continental-Scale Evapotranspiration by Combining MODIS and AmeriFlux Data Through Support Vector Machine

Feihua Yang, Michael A. White, Andrew R. Michaelis, Kazuhito Ichii, Hirofumi Hashimoto, Petr Votava, A-Xing Zhu, and Ramakrishna R. Nemani

***Abstract*—Application of remote sensing data to extrapolate evapotranspiration (ET) measured at eddy covariance flux towers is a potentially powerful method to estimate continental-scale ET. In support of this concept, we used meteorological and flux data from the AmeriFlux network and an inductive machine learning technique called support vector machine (SVM) to develop a predictive ET model. The model was then applied to the conterminous U.S. In this process, we first trained the SVM to predict 2000–2002 ET measurements from 25 AmeriFlux sites using three remotely sensed variables [land surface temperature, enhanced vegetation index (EVI), and land cover] and one ground-measured variable (surface shortwave radiation). Second, we evaluated the model performance by predicting ET for 19 flux sites in 2003. In this independent evaluation, the SVM predicted ET with a root-mean-square error (rmse) of 0.62 mm/day (approximately 23% of the mean observed values) and an R^2 of 0.75. The rmse from SVM was significantly smaller than that from neural network and multiple-regression approaches in a cross-validation experiment. Among the explanatory variables, EVI was the most important factor. Indeed, removing this variable induced an rmse increase from 0.54 to 0.77 mm/day. Third, with forcings from remote sensing data alone, we used the SVM model to predict the spatial and temporal distributions of ET for the conterminous U.S. for 2004. The SVM model captured the spatial and temporal variations of ET at a continental scale.**

***Index Terms*—AmeriFlux, evapotranspiration (ET), Moderate Resolution Imaging Spectroradiometer (MODIS), support vector machines (SVMs).**

Manuscript received September 10, 2005; revised April 10, 2006. The work of M. A. White was supported by the NASA New Investigator Program. The work of R. R. Nemani was supported by NASA's Science Mission Directorate through EOS and REASon grants. The work of A-X. Zhu was supported by the Chinese Academy of Sciences through the One-Hundred Talents Program and the Chinese Academy of Sciences International Partnership Project "Human Activities and Ecosystem Changes" (Project No. CXTD-Z2005-1).

F. Yang is with the Department of Geography, University of Wisconsin–Madison, Madison, WI 53706 USA, and also with NASA Ames Research Center, Moffett Field, CA 94035 USA (e-mail: feihuayang@wisc.edu).

M. A. White is with Utah State University, Logan, UT 84321 USA.

A. R. Michaelis, H. Hashimoto, and P. Votava are with NASA Ames Research Center, Moffett Field, CA 94035 USA, and also with California State University, Monterey Bay, Seaside, CA 93955 USA.

K. Ichii is with NASA Ames Research Center, Moffett Field, CA 94035 USA, and also with San Jose State University, San José, CA 95192 USA.

A-X. Zhu is with the Department of Geography, University of Wisconsin–Madison, Madison, WI 53706 USA, and also with the State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China.

R. R. Nemani is with NASA Ames Research Center, Moffett Field, CA 94035 USA.

Digital Object Identifier 10.1109/TGRS.2006.876297

I. Introduction

EVAPOTRANSPIRATION (ET) is the sum of water evaporated and transpired from the land surface to the atmosphere. Its energetic equivalent is latent heat flux, which is the energy required to evaporate water. ET is a crucial component for energy, hydrologic, carbon, and nutrient cycles and is a key mediator of ecosystem water status along the soil–vegetation–atmosphere continuum. Consequently, ET strongly controls processes occurring at leaf scales, such as stomatal conductance and photosynthesis, to hemispheric scales, such as midlatitude net primary production (NPP) [1].

Many semiempirical models have been developed to predict ET from meteorological data at individual points. Examples include the crop coefficient method [2], the Priestley–Taylor equation [3], and the Penman–Monteith equation [4]. Although these models are useful for calculating potential ET driven by meteorological data at regional scales [5], accurate calculation of regional actual ET is limited by requirements for extensive parameterizations of highly variable factors such as maximum stomatal conductance, soil water content, and roughness length. Process-based models [6] can dynamically simulate some of these parameters, especially stomatal conductance and soil water stress, leading to potentially useful application at the watershed scale [7]. However, process-based models are difficult to extend to large regions due to their complex structures and requirements for complete coverage of frequently poorly known land surface state variables, especially rooting depth and soil texture.

Models based on or ingesting remote sensing data have two central advantages over purely process-based models: 1) Satellite remote sensing offers broad spatial coverage and regular temporal sampling and 2) requirements for spatial and temporal parameterizations of water-constraining variables are reduced or eliminated. Remote sensing models are thus theoretically capable of accurately predicting actual ET at regional to continental scales.

Currently, methods using remote sensing data for ET calculation can be divided into two categories, namely: 1) residual methods and 2) vegetation index–surface temperature $(\mathrm{VI}{-}T_s)$ methods [8]. Residual methods [9] calculate ET by subtracting sensible heat flux from net radiation. Although conceptually attractive, the technique requires sensible heat flux estimated from land surface temperature (LST), which in turn is highly sensitive to and limited by estimates of canopy aerodynamic resistance. Although efforts to address this issue have been

made [10], estimating canopy aerodynamic resistance remains a major issue due to required knowledge of wind field, planetary boundary information, and roughness length. On the other hand, VI$-T_s$ methods utilize the scatterplot between VI and T_s [11]–[14], following the idea of Nemani and Running [15] that the slope of the scatterplot can be used as an approximation of surface resistance. However, deriving surface resistance from the scatterplot requires a continuum of soil moisture (from dry bare soil to saturated bare soil) and vegetation status (from water-stressed full-cover vegetation to well-watered full-cover vegetation) to provide a range of surface conditions.

The difficulties associated with the residual and VI$-T_s$ methods have led researchers to investigate the pertinence of models incorporating ground-based flux data collected by networks such as the AmeriFlux eddy covariance flux [16] network, which provides near-real-time observations of water and carbon exchanges. However, direct interpolation of eddy covariance flux to a regional scale is problematic due to the sparse distribution of flux towers. Researchers consequently have begun to explore statistical models using remote sensing to extrapolate eddy covariance water and carbon flux data to regional scales. For example, Wylie *et al.* [17] related coarse-resolution normalized difference vegetation index (NDVI) to carbon fluxes in a sage-brush-steppe ecosystem, and Nagler *et al.* [8] developed an empirical relationship for ET prediction over large reaches of western U.S. rivers by combining remote sensing with flux site measurements with a relative root-mean-square error (rmse) of 25%. These studies established the potential of using machine learning techniques to extrapolate ET measured at eddy covariance flux towers to a regional to continental scale.

The goal of this paper is to explore the application of support vector machine (SVM) learning techniques for ET prediction at the continental scale. To do so, we tuned and trained an SVM to predict ET measured by the AmeriFlux eddy covariance network using ground-measured and remotely sensed environmental variables, tested the SVM using a withheld portion of the flux data, and applied the final model for ET prediction over the conterminous U.S. For the purposes of this research, tuning and training are considered to occur simultaneously, i.e., parameters are tuned to different values and the SVM is retrained, producing a new SVM weighting scheme for predictions (see below and the Appendix). Testing refers to the production of error statistics using a withheld data set. In the following sections, we present: 1) a brief description of the SVM technique; 2) SVM tuning and training, including a description of the AmeriFlux [16] ET observations and the explanatory variables LST, enhanced vegetation index (EVI), shortwave radiation (SWR), and land cover; 3) results from independent testing of the SVM; 4) comparison of SVM to neural network and multiple regression; and 5) extrapolation of the SVM to the conterminous U.S.

II. Methods

A. SVM for Regression

Regression methods aim at constructing an approximate function that maps an input domain to a real-valued output domain based on a set of data examples [18]. Commonly used regression methods include conventional statistical method such as multiple regressions and machine learning methods such as neural network and SVM.

Multiple regression is a standard statistical method designed to predict the values of a target concept from two or more explanatory variables. It is conceptually simple but less suited for highly nonlinear problems, especially those outside a prescribed range of nonlinear approaches.

Machine learning is a subfield of artificial intelligence related to statistical learning and concerned with constructing computer programs capable of developing rules and behaviors without explicit guidance from human operators [19]. In the machine learning process, a set of training examples that are representative in the domain of interest are fed into a machine learning program designed to learn the connection between features of the examples and a specified target concept. Tuning is often performed in the training process to find the optimal model that not only matches the training examples but also would have good generalization over test examples. The test examples are a set of data examples independent of the training examples for model evaluation. The test result is used as an indicator of the model performance for new input features where the values of the target concept are unknown.

Commonly used machine learning techniques for nonlinear regressions are neural networks and SVM. A neural network is a computing system motivated by the function of a human brain [20]. It is widely used for regression approximation due to its ability to approximate any nonlinear functions. Although different algorithms have been proposed, multilayer perceptrons with backpropagation (MLP-BP) remain the most popular algorithm for neural networks [20] despite suffering from challenges in selecting proper network structure and finding optimal solutions. For example, the performance of an MLP-BP is related to the number of hidden layers, the number of neurons in each hidden layer, activation functions, weight initialization method, learning rate, momentum, epoch size, complexity penalty function, and regularization parameters. Furthermore, MLP-BP learning is a steepest descent method that has the risk of being trapped in a locally optimized solution [20].

The problems inherent to MLP-BP led researchers to look for alternatives such as SVM for nonlinear regressions. We fully describe SVM, the machine learning technique used in this research, in the Appendix, and provide a brief explanation as follows: SVMs were first developed by Vapnik [21], [22] for solving pattern classification problems, but they have been extended to the domain of regression approximation. For example, Zhan *et al.* [23] used SVM for the nonlinear approximation of the relationships between oceanic chlorophyll concentration and remotely sensed marine reflectance. SVMs transform nonlinear regression into linear regression by mapping the original low-dimensional input space to a higher dimensional feature space using kernel functions satisfying Mercer's condition [22]. A linear model is then constructed in the new feature space, leading to a convex quadratic programming (QP) problem guaranteed to have a global optimal solution.

The configuration of SVMs requires three types of parameters, namely: 1) C for the cost of errors; 2) ε for the width

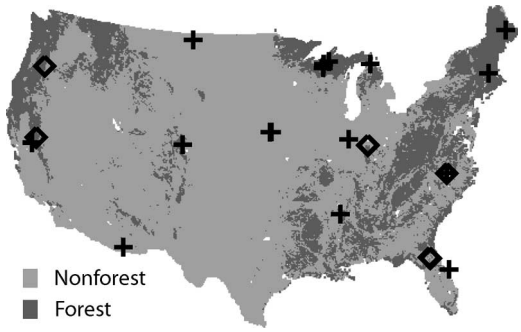


Fig. 1. Land cover and the distribution of AmeriFlux sites. Land cover was derived from MODIS land cover products (MOD12Q1) and regrouped into forest and nonforest. The 19 sites used for both training (2000–2002) and testing (2003) are shown as a plus symbol, whereas the six sites used for training alone (based on data availability) are shown as diamonds.

of an insensitive error band (ε-insensitive band); and 3) kernel parameters. The parameter C determines the tradeoff between model complexity and the training error, with higher values of C decreasing the impact of the model complexity on the optimization formulation. The parameter ε controls the tolerance for training errors. Data examples that have training errors smaller than ε are ignored in the optimization formulation. Finally, the kernel parameters vary with the selection of the kernel function. See the Appendix for details.

B. Data

Our SVM analysis required two types of data, namely: 1) ground-measured ET from AmeriFlux sites [16] for training and testing and 2) explanatory environmental data sets for both the AmeriFlux sites and the continental application. Temporal coverage was 2000–2003 for AmeriFlux sites and 2004 for conterminous U.S. application. For the continental application, all inputs were resampled and/or reprojected to an 8-km resolution.

We acquired hourly or half-hourly ground observations of ET (relative error of 25% [24]) from 25 AmeriFlux [16] sites (Fig. 1; Table I) and processed eight-day averages to correspond with satellite compositing intervals. Based on [25], we treated missing values as follows: 1) If more than 70% of data were missing in an eight-day period, we marked the period as missing; 2) if a particular time of day was missing in all eight days, i.e., all eight 2 A.M. values were missing, we marked the period as missing; 3) if neither condition 1 nor 2 was met, we filled missing values with the mean from the nonmissing days, i.e., if a single 2 A.M. value was missing from the eight-day period, we filled it with the mean of the remaining seven 2 A.M. values.

ET is a complex process influenced by a large suite of edaphic, atmospheric, and physiological variables. In particular, ET is related to LST, vapor pressure deficit (VPD), SWR, wind velocity, vegetation pattern, and soil properties. VPD and SWR are especially critical [26]. Based on the availability of remote sensing data for real-time SVM implementation, we selected LST, EVI, SWR, and land cover as explanatory variables.

We used LST as a surrogate for VPD because satellites do not provide VPD directly and VPD has a linear relationship with saturated vapor pressure derived from LST [27], [28]. For each AmeriFlux site, we used the MODIS 1-km daytime American Standard Code for Information Interchange (ASCII) subsets [29] consisting of 7×7 km regions centered on the flux towers. At each time step, we averaged the LST values using the pixels with good quality (mandatory quality assurance (QA) flag being zero in the QA data) [30] within the 7×7 km regions to represent the values at the flux site. If none of the 49 values was of good quality, we treated the period as missing. For the 2004 conterminous U.S. extrapolation, we obtained the MODIS eight-day average 1-km daytime LST product (MOD11A2; [30]), which has a deviation of ± 1 °C compared with ground measurements [30], [31].

EVI was used to indicate vegetation structure and phenological status. As for LST, we used the ASCII subsets for the AmeriFlux sites and the standard MODIS product (MOD13A2; [32]) for continental application. The EVI accuracy is about 0.03, corresponding to the rmse observed when validating against ground observations [33]. EVI is composited on a 16-day basis; for both the AmeriFlux and continental applications, we therefore assigned each 16-day composite EVI to the corresponding two eight-day periods.

We used satellite- and ground-based inputs for SWR. For continental application, we obtained daily 0.5° resolution SWR from the Surface Radiation Budget (SRB) project (derived from the Geostationary Operational Environmental Satellite [34]) and processed eight-day averages. The daily SWR has an rmse of 1.4 $\mathrm{MJ/m^2/day}$ (about 9% of the mean observed value) [35]. Due to the coarse resolution of remotely sensed SWR, we opted to use ground-measured SWR for SVM tuning, training, and testing over the AmeriFlux sites.

We obtained AmeriFlux land cover from the site descriptions. Due to data availability, we regrouped the land cover classes into two categories (Fig. 1), namely: 1) forest (evergreen needleleaf forest, evergreen broadleaf forest, deciduous broadleaf forest, and mixed forest) and 2) nonforest (savanna, shrubland, grassland, and cropland). For the conterminous U.S., we obtained land cover from the MODIS land cover product (MOD12Q1) [36] and again regrouped classes to forest and nonforest (Fig. 1). The accuracy of MODIS land cover is 70%–85% by continental regions compared with ground observations [37].

C. SVM Implementation

1) Tuning, Training, and Testing: Using AmeriFlux ET observations, we tuned and trained the SVM with 2000–2002 data and tested the SVM with 2003 data (see below). Our input variables were LST, EVI, SWR, and land cover, and our target concept was ET. After removing eight-day periods in which one or more of these variables were missing, we had a total of 1591 data examples in the training set and a total of 552 data examples in the test set. We scaled all the input variables to the range of -1 to 1, as per standard SVM techniques to eliminate the influence of variables with different absolute magnitudes.

We tuned the model as follows. First, we selected the radial basis function (RBF) kernel, as opposed to linear, polynomial, or sigmoid kernels, because it is highly flexible and requires only one parameter, σ [38]. Second, we tuned C (cost of errors), ε (width of insensitive error band), and

TABLE I
NAME, LONGITUDE, LATITUDE, AND LAND COVER OF EACH FLUX SITE IN THIS STUDY. ALL THE 25 SITES LISTED ARE IN THE TRAINING SET. THE 19 SITES THAT ARE IN THE TEST SET ARE IN BOLD

Name	Longitude (°)	Latitude (°)	land cover	Abbreviation
Audubon Research Ranch, AZ	**-110.510**	**31.600**	**shrublands**	**AR**
Blodgett, CA	-120.633	38.895	needleleaf and mixed forest	BL
Vaira Ranch, CA	**-120.951**	**38.407**	**grassland savanna**	**VR**
Niwot Ridge Forest, CO	**-105.546**	**40.033**	**needleleaf forest**	**NR**
Donaldson, FL (slash pine)	-82.163	29.755	cropland	DO
KSC Scrub Oak, FL	**-80.672**	**28.609**	**croplands**	**KSC**
Mize, FL	-82.245	29.832	needleleaf forest	MZ
Bondville, IL	**-88.292**	**40.006**	**croplands**	**BO**
Walnut River, KS	**-96.855**	**51.521**	**grasslands**	**WR**
Howland Forest, ME	**-68.727**	**45.207**	**mixed forest**	**HOF**
Sylvania Wilderness Area, MI	**-89.348**	**46.242**	**mixed forest**	**SW**
Goodwin Creek, MS	**-89.970**	**34.250**	**broadleaf forest**	**GW**
Fort Peck, MT	**-105.101**	**48.308**	**grasslands**	**FP**
Duke Forest Pine, NC	**-79.094**	**35.978**	**mixed forest**	**DF**
Lost Creek, WI	**-89.979**	**46.083**	**mixed forest**	**LC**
Willow Creek, WI	**-90.080**	**45.806**	**mixed forest**	**WC**
Metolius Young Ponderosa Pine, OR	-121.568	44.437	needleleaf forest	MY
Harvard Forest, MA	**-72.172**	**42.536**	**broadleaf forest**	**HF**
University of MI (Michigan), MI	**-84.7140**	**45.560**	**mixed forest**	**UM**
Indiana MMSF, IN	-86.413	39.321	broadleaf forest	IM
Duke Forest Hardwoods, NC	-79.100	35.974	mixed forest	DFH
Tonzi Ranch, CA	**-120.950**	**38.419**	**grass savanna**	**TR**
Mead Irrigated Nebraska, NE	**-96.286**	**41.099**	**croplands**	**MI**
Mead Rotation Nebraska, NE	**-96.281**	**41.099**	**croplands**	**MR**
Mead Rainfed Nebraska, NE	**-96.440**	**41.099**	**croplands**	**MRA**

σ (kernel parameter) using a grid search with a three-fold cross-validation training process [38]. In this approach, the training examples are randomly divided into three nonoverlapping subsets; training is performed three times on two of the subsets, with the remaining subset reserved for testing; parameters yielding the lowest cross-validation errors are selected. We initially conducted a coarse grid search for C $(2^{-1}, 2^0, 2^1, \ldots, 2^4)$, ε $(2^{-5}, 2^{-4.5}, 2^{-4}, \ldots, 2^{-2})$, and σ $(2^{-3}, 2^{-2.5}, 2^{-2}, \ldots, 2^4)$ and identified the C, ε, and σ combination producing the lowest mean cross-validation rmse. We then used a progressively finer grid search until the variance of the rmse was smaller than 0.01. Third, using the selected C, ε, and σ, we conducted a final training of the SVM with the 2000–2002 AmeriFlux data (see the Appendix for details).

Finally, we tested the trained model on the test set for three groups, namely: 1) forest; 2) nonforest; and 3) forest and nonforest combined. The testing data set (2003) was from 19 flux sites (nine forest and ten nonforest sites) where measurements were available. We evaluated SVM performance using rmse, R^2, scatterplots of predicted versus observed ET, seasonal variations between the predicted and observed ET, and residual analysis.

2) Contribution of Each Input Variable on ET Variations: We examined the contribution of each input variable on SVM ET predictions by sequentially removing one of the input variables (LST, EVI, SWR, and land cover) and replicating the cross-validation training process. We assessed the contribution of each input variable with the mean cross-validation rmse and R^2 from the cross-validation training process.

3) Comparison With Other Methods: We compared the performance of SVMs to that of: 1) MLP-BP and 2) multiple regression using all the input variables (LST, EVI, SWR, and land cover) on the training set. To conduct a statistical comparison, we employed a ten-fold cross validation on the training set, rather than the three-fold cross validation employed in the tuning/training process described in Section II-C1. We randomly divided the 1591 training examples into ten nonoverlapping subsets and trained the SVM, MLP-BP, or multiple regression ten times on nine of the subsets with the remaining subset reserved for testing. We configured MLP-BP with one hidden layer, sigmoid activation function, random weight initialization, and weight decay regularization. We set the regularization parameters to 1 and the epoch size to 64. Other parameters (the number of neurons in the hidden layer, learning rate, and momentum) were selected with the training process used for developing the SVM model. Based on [19], we calculated the rmse of the ten test sets for each of the three algorithms. Paired Student's t-tests (95% confidence level), a statistical test used to determine whether there is statistical difference between the mean of two samples, was then conducted between: 1) SVMs and neural network and 2) SVMs and multiple regression.

4) Generalization From AmeriFlux Sites to the Conterminous U.S.: Based on research showing that the AmeriFlux network is representative of conterminous U.S. ecoregions [39] and that the 25 flux sites in this study included most of the active flux sites in the AmeriFlux network (Fig. 1), we reasoned that the knowledge learned at flux sites can be extrapolated to the conterminous U.S.. To generalize the model learned from flux sites to the conterminous U.S., we first conducted a new training of the SVM with the entire 2000–2003 data set using the selected C, ε, and σ. The trained model was then used to

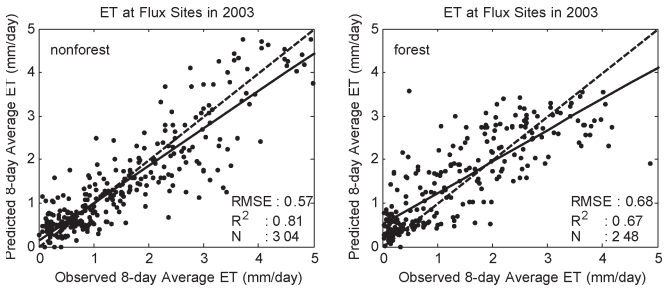


Fig. 2. Scatterplots of AmeriFlux-observed versus SVM-predicted ET in 2003 for forest and nonforest sites. The SVM was trained with forest and nonforest combined, but tested with forest and nonforest separately. Dashed lines show a 1 : 1 relationship; solid lines show a least squares regression line.

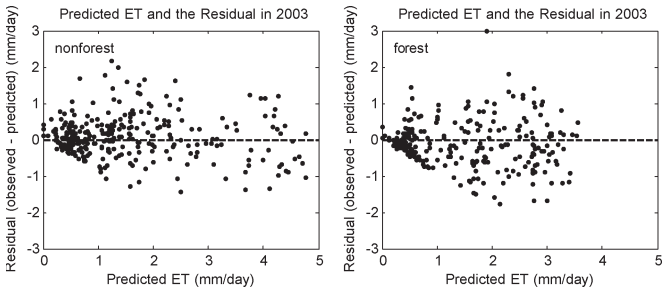


Fig. 3. Scatterplots of SVM-predicted ET versus residuals (observed − predicted) in 2003 for forest and nonforest sites.

investigate the spatial and temporal distributions of ET over the conterminous U.S. for 2004.

III. Results and Discussion

A. AmeriFlux Sites

Using the full input training set of LST, EVI, SWR, and land cover, the parameter combination of $C = 8.0$, $\varepsilon = 0.18$, and $\sigma = 1.41$ produced the smallest mean cross-validation rmse of 0.54 mm/day and an R^2 of 0.78. Using these parameter values and the trained SVM, the testing on the 2003 AmeriFlux data produced an rmse for the forest and nonforest combined sites of 0.62 mm/day with an R^2 of 0.75. Nonforest sites had 0.57 mm/day rmse and 0.81 R^2, whereas forest sites had 0.68 mm/day rmse and 0.67 R^2 (Fig. 2), indicating that the SVM performed better in nonforest than in forest sites. Furthermore, the residuals (observed − predicted ET) were significantly lower in nonforest than in forest sites (t-test, $p < 0.05$). We speculate that the low performance over forest sites might come from the saturation of optical remote sensing data over dense canopies such as forest. Analysis of the relationships between predicted ET and the residuals (Fig. 3) revealed that the residuals were not randomly distributed. Low prediction errors were associated with low ET values, whereas high prediction errors were associated with high ET values. This indicated that the lack of variables not included in the SVM model affected the ET retrieval performances.

The SVM represented most features of measured ET seasonality in the 2003 AmeriFlux test data (Fig. 4). For some specific sites, episodes of under- or overprediction occurred. For example, the SVM overpredicted ET at Willow Creek in June 2003 when the observed eight-day average ET at this site

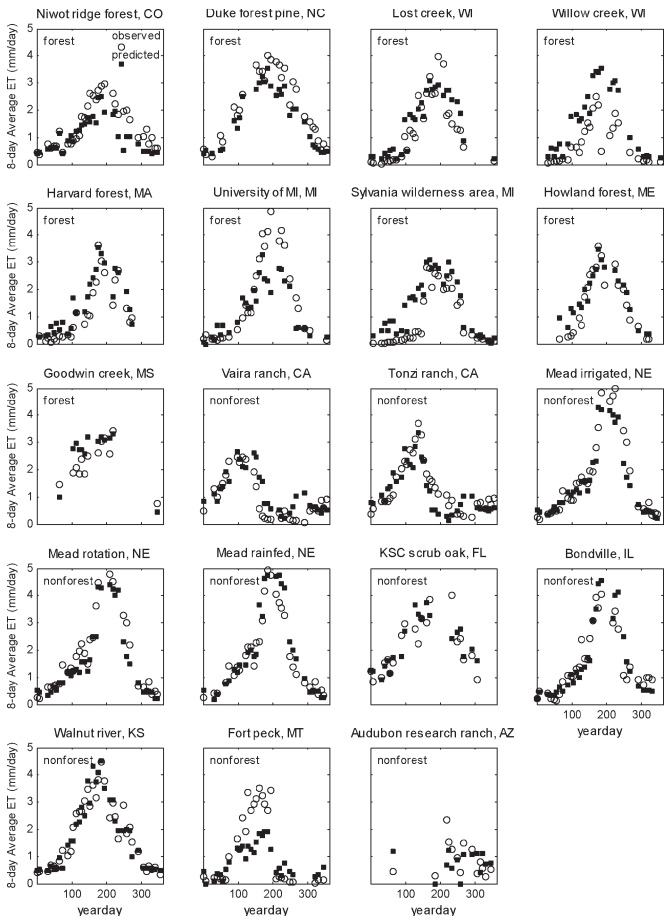


Fig. 4. Comparison of the seasonal variations of the observed eight-day average ET (open circle) and the predicted eight-day average ET (closed square) at flux sites for 2003.

dropped from 2.50 to 0.49 mm/day. Similar declines did not occur in measured Willow Creek ET during 2000–2002, yet the seasonal variations of LST, EVI, and SWR during 2003 were similar to those during 2000–2002. Therefore, we concluded that ET at Willow Creek in June 2003 was likely to be driven by other factors such as unusual wind and/or soil moisture patterns. Snow cover has the potential to produce aberrantly high EVI values, which in turn would lead to ET overprediction. This phenomenon may account for prediction errors at Sylvania Wilderness Area in the spring and early summer of 2003.

Underprediction errors occurred in the summer of 2003 at the University of Michigan and at Fort Peck. Although factors not explicitly included in the SVM, such as soil moisture, may account for some underprediction errors, we speculate that use of eight-day mean conditions may inadequately represent the effects of nonlinear ET processes, i.e., the SVM drivers of LST, EVI, and SWR may be incapable of producing extremely high ET values occurring during short periods of superoptimal physiological and boundary layer conditions.

Analysis of the mean observed and predicted ET showed that SVM performance varied by flux sites and land cover (Fig. 5). The predicted and observed mean ET was within 30% of the 1 : 1 line for all sites except Willow Creek, Fort Peck, and Sylvania Wilderness Area. High overprediction occurred at Sylvania Wilderness Area (62%) and Willow Creek (100%),

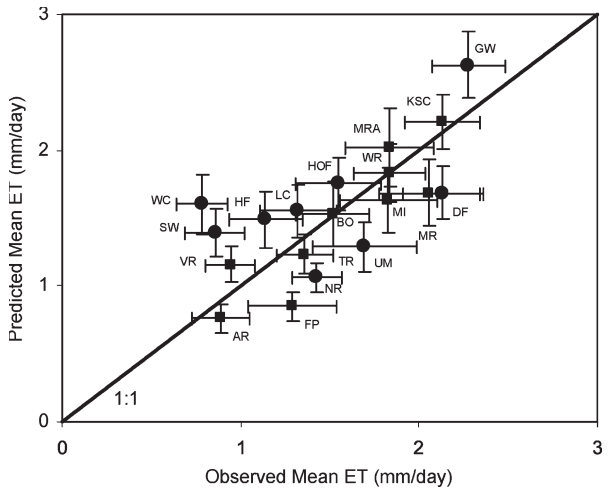


Fig. 5. Observed and predicted mean ET across flux sites. Forest sites are marked as circles and nonforest as squares. Error bars are standard deviations of the observed and predicted eight-day average ET. Abbreviations of flux sites refer to Table I.

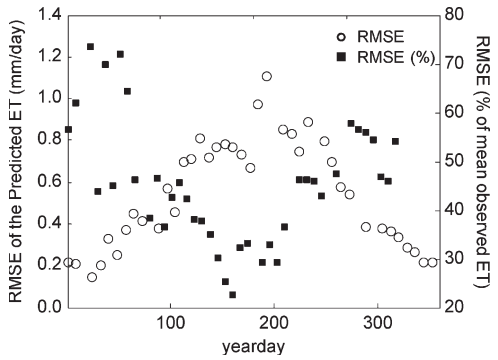


Fig. 6. Seasonal variations of ET prediction errors averaged across all flux sites. Left axis shows rmse error expressed as millimeters per day; right axis shows rmse error expressed as a percent of average observed ET.

whereas high underprediction occurred at Fort Peck (34%). In terms of land cover, the predicted mean ET was within 20% of the 1 : 1 line for all nonforest sites except Vaira Ranch and Fort Peck, but of the forest sites, only Lost Creek, Howland Forest, and Goodwin Creek were within 20% of the 1 : 1 line. The overall prediction error was 23%, which was close to the 25% flux observation errors [24] and near the 20% error limit for ET prediction using remote sensing methods [24]. Given the heterogeneity of the AmeriFlux data and the simplicity of the model inputs, model performance was promising.

Residual analysis showed that rmse averaged across all AmeriFlux sites showed a strong seasonality (Fig. 6). In absolute magnitudes, winter had low prediction errors ($\sim$0.2 mm/day), whereas warm season errors often exceeded 0.8 mm/day. Yet when expressed as a percentage of rmse to the mean observed ET, the pattern was reversed: Winter errors were often above 50%, but summer errors rarely exceeded 40%. The results were consistent with Fig. 3, where high/low rmse was correspondent to high/low ET. We suspected that the measurement error from the explanatory variables had additive or multiplicative influence on ET prediction. Thus, potential users should carefully consider the sensitivity of the system under study to the seasonality of absolute and relative ET prediction errors.

TABLE II
IMPACT OF REMOVING ONE OF THE FOUR INPUT VARIABLES ON THE PREDICTING PERFORMANCE OF SVM ON ET. THE RESULTS SHOWN ARE THE AVERAGE FROM THREE-WAY CROSS VALIDATION ON THE TRAINING SET

Variable removed	RMSE (mm/day)	R^2	C	ε	σ
None	0.54	0.78	8.00	0.18	1.41
LST	0.64	0.72	16.00	0.25	2.00
EVI	0.77	0.59	4.00	0.18	5.66
SWR	0.75	0.62	8.00	0.25	2.00
LC	0.60	0.75	16.00	0.25	2.83

B. Contribution of Each Input Variable on ET Variations

Removal of EVI caused the largest performance reduction in SVM cross-validation error statistics (Table II): rmse increased from 0.54 to 0.77 mm/day and R^2 decreased from 0.78 to 0.59. Removal of SWR was nearly as important, leading to an increased rmse of 0.75 mm/day and a decreased R^2 of 0.62. Removal of LST produced comparatively minor changes: rmse rose to 0.64 mm/day and R^2 fell to 0.72. The rmse only increased by 0.06 mm/day with removal of the land cover. Therefore, we concluded that EVI and SWR alone captured most of the ET variations. However, our input variable ranking was based on the eight-day averages within 7×7 km regions. Thus, potential users should be cautious on the relative importance of the input variables reported in this study because the importance may change with different spatial and temporal resolutions.

C. Comparison With Other Methods

SVMs outperformed other techniques (Table III). SVM rmse was smaller than neural network and multiple regression rmse in the ten trials. The 95% confidence interval of the difference in rmse was -0.043 ± 0.036 (-0.079 to -0.007) between SVMs and neural networks and -0.141 ± 0.033 (-0.174 to -0.108) between SVMs and multiple regression. We thus conclude that for ET prediction, SVMs perform significantly better than neural networks and multiple regressions.

D. Generalization From AmeriFlux Sites to the Conterminous U.S.

The purpose of our conterminous U.S. application was to assess qualitatively whether or not the SVM technique produced spatiotemporal ET estimates consistent with expected patterns. In the four eight-day periods in 2004 representing spring, summer, fall, and winter over the conterminous U.S., the SVM model trained at the AmeriFlux sites generally captured the expected ET features (Fig. 7). Temporally, March ET was low because of low temperature and low radiation; July ET was high because of high precipitation, peak vegetation, and intensive radiation; September ET dropped as vegetation senesced; and December ET was lowest with coldest temperatures and lowest radiation. Spatially, March ET was high in California, Florida, and Texas due to high temperature; July ET was highest in

TABLE III
RMSE (IN MILLIMETERS PER DAY) OF SVM, NEURAL NETWORK, AND MULTIPLE REGRESSION FROM TEN-FOLD CROSS VALIDATION ON THE TRAINING SET

Trials	1	2	3	4	5	6	7	8	9	10	mean
SVM	0.56	0.58	0.46	0.58	0.51	0.57	0.56	0.46	0.55	0.55	0.54
Neural network	0.64	0.64	0.52	0.60	0.55	0.65	0.57	0.55	0.62	0.57	0.58
Multiple regression	0.71	0.77	0.62	0.63	0.64	0.73	0.68	0.67	0.70	0.64	0.68

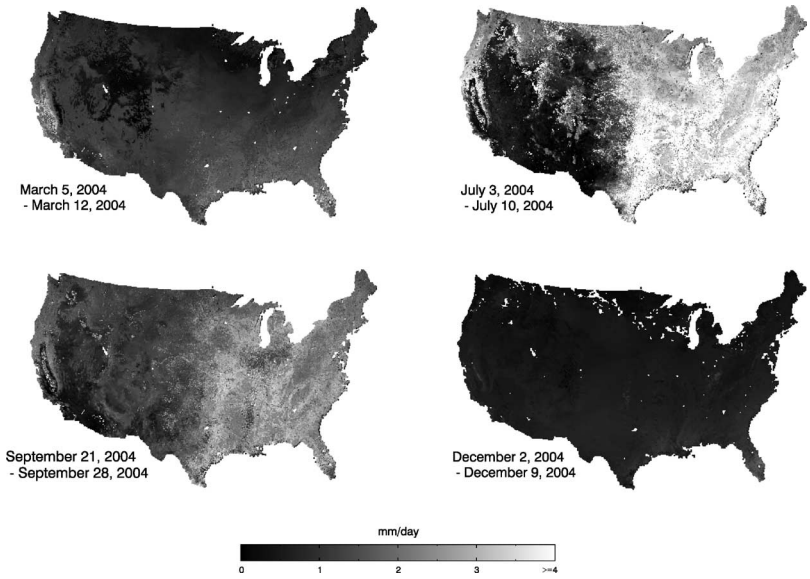


Fig. 7. SVM-predicted eight-day average ET for the continental U.S. in March 5–March 12, July 3–July 10, September 21–September 28, and December 2–December 9 of 2004. ET greater than 4.0 mm/day was cut off to 4.0 mm/day for display purposes.

the eastern U.S., due to high moisture and energy availability; and September ET showed patterns consistent with early phenological decline of agricultural regions in the Mississippi River valley and Midwest. During all periods, predicted ET followed expected elevational patterns, e.g., mountainous areas had higher ET than surrounding low elevation regions in July.

At least four factors may have limited the continental generalization: 1) the use of ground-observed SWR at flux sites for model development but the use of 0.5° resolution SRB-derived SWR for the conterminous U.S. tended to smooth ET variations; 2) the use of 16-day composite EVI for eight-day periods also tended to smooth ET; 3) the 25 flux sites included in this study may not fully represent the spatiotemporal variation of actual ET; and 4) the observed ET may have measurement errors jeopardizing the model generalization ability. Additional flux data would mitigate some of these issues, as would finer resolution EVI (temporally) and SWR (spatially). In summary, although there are potentially confounding factors and it is difficult to validate ET distribution over the conterminous U.S., our results show that the SVM trained at AmeriFlux sites generally captured the expected spatiotemporal variations of ET over the conterminous U.S. This result, although not quantitatively conclusive, strongly suggests that SVM models trained at flux sites can be generalized to larger regions.

IV. CONCLUSION

Using a combination of ground-measured ET, machine learning techniques, and remotely sensed inputs, we developed a technique to predict ET over the conterminous U.S. with an average test error of 0.62 mm/day. We found that EVI and SWR were more important than LST or land cover for ET prediction, suggesting that finer temporal resolution for EVI (possibly by including Terra and Aqua satellites) and finer spatial resolution for SWR could substantially enhance regional and/or continental ET estimates. The method can also be improved by increasing the number of ET ground observations. A central limitation of the SVM technique is that the knowledge learned is encoded as weights that are not directly comprehensible to humans. However, methods exist with which to convert the structure learned by SVMs to a more understandable format, such as rules, thus enhancing our understanding of ET processes at different scales. With these improvements, the combination of satellite data with ground observation of ET through machine learning should be able to provide prediction of ET with sufficient accuracy and timeliness for application in regional to continental natural resource management services and serve as a supplemental tool for existing ET retrieval methods. Furthermore, based on the product accuracy and the spatiotemporal scale considered in this study, the SVM-based ET prediction can also be useful for documenting hydroecological models at regional to continental scales.

APPENDIX
SVM FOR REGRESSION

SVMs are based on the theory that a multidimensional input space is more likely to be linearly separable in a new feature space if the transformation is nonlinear and the new feature space has higher dimensions than the original one [40].

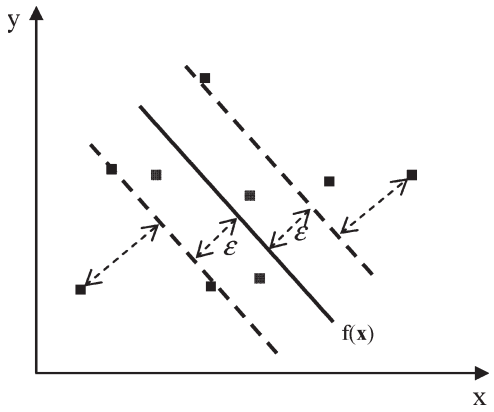


Fig. 8. One-dimensional linear regression with ε-insensitive band.

Therefore, the original nonlinear problem can be solved using linear models in the new feature space. Given a set of training examples $\{(\mathbf{x}_i, \mathbf{y}_i), 1 \le i \le n\}$, where $\mathbf{x}_i \in R^\ell$ $(\ell \ge 1)$ is the input, $\mathbf{y}_i \in R$ is the target concept, and n is the number of train examples, SVMs map the input space $\mathbf{x}$ to a higher dimensional feature space $\phi(\mathbf{x})$ using a nonlinear transformation function ϕ and construct a linear model in this new feature space as follows:

$$\mathbf{y} = f(\mathbf{x}) = \langle \mathbf{w} \cdot \phi(\mathbf{x}) \rangle + \mathbf{b} \tag{1}$$

where $\mathbf{w}$ is the weight vector, $\mathbf{b}$ is the noise, and $\langle \mathbf{w} \cdot \phi(\mathbf{x}) \rangle$ is the dot product between $\mathbf{w}$ and $\phi(\mathbf{x})$. The task is then to find a functional form for f, which can predict new cases that the SVM has not seen before known as generalization. This can be achieved by training the SVM on a sample set (training set) through a sequential optimization of a loss function (error function). A ε-insensitive loss function $L^\varepsilon(\mathbf{y}, \mathbf{x}, f(\mathbf{x}))$ proposed by Vapnik [22] is defined as follows:

$$L^\varepsilon(\mathbf{y}, \mathbf{x}, f(\mathbf{x})) = \begin{cases} 0, & \text{if } |\mathbf{y} - f(\mathbf{x})| \le \varepsilon \\ |\mathbf{y} - f(\mathbf{x})| - \varepsilon, & \text{otherwise.} \end{cases} \tag{2}$$

This loss function ignores errors when the difference between the predicted value and the true value is smaller than a threshold ε. Fig. 8 shows one-dimensional linear regression function with ε-insensitive band. The errors of data points within the band are ignored. Data points out of the ε-insensitive band are called support vectors, and only support vectors contribute to the optimization solution.

Unlike neural networks that have the risk of being trapped in a local minimum, the generalization of an SVM regression model is optimized by minimizing the generalization error bound—the combination of the training error (empirical risk) and model complexity (structural risk) as follows [18]:

$$\text{Minimize } \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n} L^\varepsilon(\mathbf{y}_i, \mathbf{x}_i, f(\mathbf{x}_i)) \tag{3}$$

where $\|\mathbf{w}\|$ is the Euclidean norm of the weight vector. Minimizing $\|\mathbf{w}\|$ is equivalent to minimizing the model complexity. The parameter C is the cost of errors. It determines the tradeoff between the model complexity and the training errors. If C is too large, then the model complexity part in the optimization formulation is ignored. On the other hand, if C is small, then the structural risk has more influence in the optimization formulation.

By introducing two nonnegative slack variables for each data point to measure the deviation of each point outside the ε-insensitive band, the optimization problem of (3) can be equivalently represented as follows:

$$\begin{aligned} \underset{w,b,\xi,\xi^*}{\text{Minimize}} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \\ \text{Subject to} \quad & (\langle \mathbf{w} \cdot \phi(\mathbf{x}_i) \rangle + \mathbf{b}) - \mathbf{y}_i \le \varepsilon + \xi_i \\ & \mathbf{y}_i - (\langle \mathbf{w} \cdot \phi(\mathbf{x}_i) \rangle + \mathbf{b}) \le \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \ge 0, \qquad i = 1, \ldots, n \end{aligned} \tag{4}$$

where $\|\mathbf{w}\|$ and C are defined in (3), and ξ_i denotes the predicted value to be above the true value by more than ε and ξ_i^* to be below the true value by more than ε.

The optimization problem presented in (4) can be solved using the technique of Lagrange multipliers, and the functional form of (1) can be represented as follows:

$$\mathbf{y} = f(\mathbf{x}) = \sum_{i=1}^{n}(\alpha_i^* - \alpha_i) K(\mathbf{x}_i, \mathbf{x}) + \mathbf{b} \tag{5}$$

where $K(\mathbf{x}_i, \mathbf{x}) = \langle \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}) \rangle$ is the kernel function satisfying Mercer's condition [22], and α_i and α_i^* are Lagrange multipliers obtained by solving the following QP problem:

$$\begin{aligned} \underset{\alpha,\alpha^*}{\text{Maximize}} \quad & \sum_{i=1}^{n}\mathbf{y}_i(\alpha_i^* - \alpha_i) - \frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i^* - \alpha_i) \\ & \times K(\mathbf{x}_i, \mathbf{x}_j)(\alpha_j^* - \alpha_i) - \varepsilon\sum_{i=1}^{n}(\alpha_i^* + \alpha_i) \\ \text{Subject to} \quad & \sum_{i=1}^{n}(\alpha_i^* - \alpha_i) = 0, \qquad 0 \le \alpha_i,\ \alpha_i^* \le C,\ i = 1, \ldots, n \end{aligned} \tag{6}$$

whereas $\mathbf{b}$ is chosen such that

$$\langle \mathbf{w} \cdot \phi(\mathbf{x}_i) \rangle + \mathbf{b} - \mathbf{y}_i = -\varepsilon \qquad \text{for any } i \text{ with } 0 < (\alpha_i^* - \alpha_i) < C. \tag{7}$$

The QP problem represented in (6) is a strict convex quadratic optimization problem that has a global optimal solution. In the functional form represented in (5), only support vectors have nonzero $(\alpha_i - \alpha_i^*)$ [22]. This is due to the use of ε-insensitive loss function. The parameter ε controls the width of the ε-insensitive band. It affects the number of support vectors used to construct the regression function. If the ε is too large, then fewer support vectors are selected, resulting in a sparse representation of the solution. On the other hand, if ε is too small, then more support vectors are selected, resulting in a complex model.

It is difficult to find an analytical form of the nonlinear transformation function ϕ. However, there is no need to know

ϕ explicitly. In the representation (6), only the dot product $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j) \rangle$ is necessary for the optimization process, and we can generalize the dot product to other kernel functions. For example, a commonly used kernel function is RBF, which has the following form:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}) \rangle = \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x}_i - \mathbf{x}_j\|^2\right)$$

where σ is *a priori*. Therefore, the dot product is computed in the new feature space without explicitly knowing the features in the new feature space.

Other kernel functions include linear, polynomial, and sigmoid. However, RBF is often preferred because linear kernel is only appropriate for linear problems and polynomial kernel has computational difficulties. Sigmoid kernel function is not widely used because it is not well studied and it behaves similar to RBF to some extent [38].

As a summary, SVM regression use kernel function to map the input space to a higher dimensional feature space implicitly. A linear model is then constructed in this new feature space using ε-insensitive loss while trying to reduce model complexity.

ACKNOWLEDGMENT

The SVM software from LIBSVM [38] facilitated this study. The authors would like to thank J. Dungan for providing constructive suggestions on earlier versions of the manuscript.

REFERENCES

[1] R. R. Nemani, C. D. Keeling, H. Hashimoto, W. M. Jolly, S. C. Piper, C. J. Tucker, R. B. Myneni, and S. W. Running, "Climate-driven increases in global terrestrial net primary production from 1982 to 1999," *Science*, vol. 300, no. 5625, pp. 1560–1563, Jun. 2003.

[2] R. G. Allen, L. S. Pereira, D. Raes, and M. Smith, *Crop Evapotranspiration—Guidelines for Computing Crop Water Requirements*, vol. 56. Rome, Italy: FAO, 1998.

[3] C. H. B. Priestley and R. J. Taylor, "On the assessment of surface heat flux and evaporation using large-scale parameters," *Mon. Weather Rev.*, vol. 100, no. 2, pp. 81–92, 1972.

[4] J. L. Monteith, "Evaporation and environment," in *Proc. 19th Symp. Soc. Exp. Biologists*, 1965, pp. 205–234.

[5] B. J. Choudhury, "Global pattern of potential evaporation calculated from the Penman-Monteith equation using satellite and assimilated data," *Remote Sens. Environ.*, vol. 61, no. 1, pp. 64–81, Jul. 1997.

[6] P. E. Thornton, "Regional ecosystem simulation: Combining surface- and satellite-based observations to study linkages between terrestrial energy and mass budgets," in *School of Forestry*. Missoula, MT: Univ. of Montana, 1998.

[7] A. X. Zhu and D. Scott Mackay, "Effects of spatial detail of soil information on watershed modeling," *J. Hydrol.*, vol. 248, no. 1, pp. 54–77, Jul. 2001.

[8] P. L. Nagler, R. L. Scott, C. Westenburg, J. R. Cleverly, E. P. Glenn, and A. R. Huete, "Evapotranspiration on western U.S. rivers estimated using the enhanced vegetation index from MODIS and data from eddy covariance and Bowen ratio flux towers," *Remote Sens. Environ.*, vol. 97, no. 3, pp. 337–351, Aug. 2005.

[9] M. Moran, T. Clarke, U. Inoue, and A. Vidal, "Estimating crop water deficit using the relation between surface air temperature and spectral vegetation index," *Remote Sens. Environ.*, vol. 49, no. 3, pp. 246–263, 1994.

[10] W. G. M. Bastiaanssen, M. Menenti, R. A. Feddes, and A. A. M. Holtslag, "The surface energy balance algorithm for land (SEBAL). Part 1. Formulation," *J. Hydrol.*, vol. 212/213, pp. 198–212, Dec. 1998.

[11] K. Nishida, R. R. Nemani, J. M. Glassy, and S. W. Running, "Development of an evapotranspiration index from Aqua/MODIS for monitoring surface moisture status," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 2, pp. 493–501, Feb. 2003.

[12] L. Jiang and S. Islam, "Estimation of surface evaporation map over Southern Great Plains using remote sensing data," *Water Resour. Res.*, vol. 37, no. 2, pp. 329–340, 2001.

[13] J. C. Price, "Using spatial context in satellite data to infer regional scale evapotranspiration," *IEEE Trans. Geosci. Remote Sens.*, vol. 28, no. 5, pp. 940–948, Sep. 1990.

[14] X. Yang, Q. Zhou, and M. Melville, "Estimating local sugarcane evapotranspiration using Landsat TM image and a VITT concept," *Int. J. Remote Sens.*, vol. 18, no. 2, pp. 453–459, Jan. 1997.

[15] R. R. Nemani and S. W. Running, "Estimation of regional surface resistance to evapotranspiration from NDVI and thermal-IR AVHRR data," *J. Appl. Meteorol.*, vol. 28, no. 4, pp. 276–284, 1989.

[16] D. Baldocchi, E. Falge, L. H. Gu, R. Olson, D. Hollinger, S. Running, P. Anthoni, C. Bernhofer, K. Davis, R. Evans, J. Fuentes, A. Goldstein, G. Katul, B. Law, X. H. Lee, Y. Malhi, T. Meyers, W. Munger, W. Oechel, K. T. Paw U, K. Pilegaard, H. P. Schmid, R. Valentini, S. Verma, T. Vesala, K. Wilson, and S. Wofsy, "FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities," *Bull. Amer. Meteorol. Soc.*, vol. 82, no. 11, pp. 2415–2434, Nov. 2001.

[17] B. Wylie, D. Johnson, E. Laca, N. Saliendra, T. Gilmanov, and B. Reed, "Calibration of remotely sensed, coarse resolution NDVI to CO_2 fluxes in a sage-brush-steppe ecosystem," *Remote Sens. Environ.*, vol. 85, no. 2, pp. 243–255, May 2003.

[18] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[19] T. M. Mitchell, *Machine Learning*. New York: McGraw Hill, 1997.

[20] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Englewood Cliffs, NJ: Prentice Hall, 1998.

[21] V. Vapnik and A. Chervonenkis, "The necessary and sufficient conditions for consistency in the empirical risk minimization method," *Pattern Recognit. Image Anal.*, vol. 1, no. 3, pp. 283–305, 1991.

[22] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.

[23] H. Zhan, P. Shi, and C. Chen, "Retrieval of oceanic chlorophyll concentration using support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 12, pp. 2947–2951, Dec. 2003.

[24] H. Jiang, S. Liu, P. Sun, S. An, G. Zhou, and C. Li, "The influence of vegetation type on the hydrological process at the landscape scale," *Can. J. Remote Sens.*, vol. 30, no. 5, pp. 743–763, 2004.

[25] E. Falge, D. Baldocchia, R. Olsonb, P. Anthonic, M. Aubinetd, C. Bernhofere, G. Burbaf, R. Ceulemansg, R. Clementh, H. Dolmani, A. Granierj, P. Grossj, T. Grünwalde, D. Hollingerk, N.-O. Jensenl, G. Katulm, P. Keronenn, A. Kowalskig, C. T. Laim, B. E. Lawc, T. Meyerso, J. Moncrieffh, E. Moorsi, J. W. Mungerp, K. Pilegaardl, Ü. Rannikn, C. Rebmannq, A. Suykerf, J. Tenhunenr, K. Tus, S. Vermaf, T. Vesalan, K. Wilsono, and S. Wofsyp, "Gap filling strategies for long term energy flux data sets," *Agricult. For. Meteorol.*, vol. 107, no. 1, pp. 71–77, Mar. 2001.

[26] G. S. Campbell and J. M. Norman, *An Introduction to Environmental Biophysics*, 2nd ed. New York: Springer-Verlag, 1998.

[27] R. J. Granger, "A feedback approach for the estimation of evapotranspiration using remotely-sensed data," in *Proc. Workshop Appl. Remote Sens. Hydrol.*, G. W. Kite, A. Pietroniro, and T. J. Pultz, Eds. Saskatoon, Canada: Nat. Hydrol. Res. Inst., 1995, vol. 14, pp. 211–222.

[28] ——, "Comparison of surface and satellite-derived estimates of evapotranspiration using a feedback algorithm," in *Proc. Workshop Appl. Remote Sens. Hydrol.*, G. W. Kite and T. J. Pultz, Eds. Saskatoon, Canada: Nat. Hydrol. Res. Inst., 1997, vol. 17, pp. 71–82.

[29] R. B. Cook, S. M. Margle, S. K. Holladay, F. A. Heinsch, and C. B. Schaaf, "Subsets of remote sensing products for AmeriFlux sites: MODIS ASCII subsets" presented at the Ameriflux Annual Meeting, Boulder, CO, Oct. 5–7, 2004. Poster available at http://www.modis.ornl.gov/modis/modis_presentations.cfm

[30] Z. Wan, Y. Zhang, Q. Zhang, and Z. Li, "Validation of the land-surface temperature products retrieved from Terra Moderate Resolution Imaging Spectroradiometer data," *Remote Sens. Environ.*, vol. 83, no. 1/2, pp. 163–180, Nov. 2002.

[31] ——, "Quality assessment and validation of the MODIS global land surface temperature," *Int. J. Remote Sens.*, vol. 25, no. 1, pp. 261–274, Jan. 2004.

[32] A. Huete, K. Didan, T. Miura, and E. Rodriguez, "Overview of the radiometric and biophysical performance of the MODIS vegetation indices," *Remote Sens. Environ.*, vol. 83, no. 1/2, pp. 195–213, 2002.

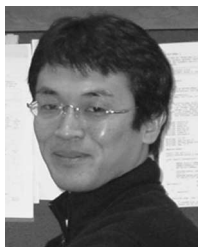
[33] X. Gao, A. R. Huete, and K. Didan, "Multisensor comparisons and validation of MODIS vegetation indices at the semiarid Jornada experimental range," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 10, pp. 2368–2381, Oct. 2003.
[34] R. T. Pinker, O. Laszlo, J. D. Tarpley, and K. Mitchell, "Geostaionary satellite parameters for surface energy balance," *Adv. Space Res.*, vol. 30, no. 11, pp. 2427–2432, Nov. 2002.
[35] J. A. Otkin, M. C. Anderson, J. R. Mecikalski, and G. R. Diak, "Validation of GOES-based isolation estimates using data from the U.S. climate reference network," *J. Hydrometeorol.*, vol. 6, no. 4, pp. 460–475, Aug. 2005.
[36] M. A. Friedl, D. K. McIver, J. C. F. Hodges, X. Y. Zhang, D. Muchoney, A. H. Strahler, C. E. Woodcook, S. Gopal, A. Schneider, A. Cooper, A. Baccini, F. Gao, and C. Schaaf, "Global land cover mapping from MODIS: Algorithms and early results," *Remote Sens. Environ.*, vol. 83, no. 1/2, pp. 287–302, Nov. 2002.
[37] M. A. Friedl, "Validation of the consistent-year VOO3 MODIS land cover product," in "Internal PI Document," NASA Goddard Space Flight Center, Greenbelt, MD, 2003.
[38] C.-C. Chang, and C.-J. Lin. (2005). *LIBSVM—A Library for Support Vector Machines,* Version 2.8. Software Available at http://www.csie.ntu.edu.tw/~cjlin/libsvm/
[39] W. W. Hargrove, F. M. Hoffman, and B. E. Law, "New analysis reveals representativeness of the AmeriFlux network," *EOS Trans. AGU*, vol. 84, no. 48, pp. 529–544, Dec. 2003.
[40] T. M. Cover, "Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition," *IEEE Trans. Electron. Comput.*, vol. EC-14, no. 3, pp. 326–334, Jun. 1965.

Feihua Yang received the B.S. degree in geophysics from China, the M.S. degree in geography from the University of Montana, Missoula, and the M.S. degree in computer science from the University of Wisconsin–Madison. She is currently working toward the Ph.D. degree in geography from the University of Wisconsin–Madison.

Michael A. White is an Associate Professor with Utah State University, Logan. His research interests include ecological impacts of climate change and variability on terrestrial ecosystems. His recent research interests include vegetation and land surface phenology, viticulture/climate interactions, science education, and parameterization of ecological models.

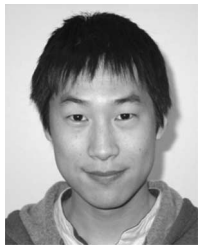
Andrew R. Michaelis received the B.S. degree in computer science from the University of Montana, Missoula.

He is currently a Software Developer with California State University, Monterey Bay, and the NASA Ames Research Center, Moffett Field, CA.



Kazuhito Ichii received the Ph.D. degree in Earth and planetary sciences from Nagoya University, Nagoya, Japan in 2002.

He is currently a Research Scientist with San Jose State University, San Jose, CA, and the NASA Ames Research Center, Moffett Field, CA.



Hirofumi Hashimoto received the Ph.D. degree in agriculture science from the University of Tokyo, Tokyo, Japan, in 2005.

He is currently a Postdoctoral Research Scientist with California State University, Monterey Bay, and the NASA Ames Research Center, Moffett Field, CA.

Petr Votava received the M.S. degree in computer science from the University of Montana, Missoula in 1997.

He is currently a Senior Software Engineer for the Terrestrial Observation and Prediction System with the NASA Ames Research Center and California State University, Monterey Bay. His research interests include distributed systems and software engineering.

A-Xing Zhu received the B.S. degree in geography from China, the M.S. degree in geography from the University of Calgary, Calgary, AB, Canada, and the Ph.D. degree in geography from the University of Toronto, Toronto, ON, Canada.

He is currently a Professor in geography with the University of Wisconsin–Madison.

Ramakrishna R. Nemani received the Ph.D. degree from the University of Montana, Missoula, in 1987.

He was with the University of Montana as a Research Faculty. Since 2003, he has been a Research Scientist with the NASA Ames Research Center, Moffett Field, CA, and is primarily involved with using MODIS products for ecosystem modeling and forecasting.